

UNIVERSIDADE ESTADUAL DO OESTE DO PARANÁ
CAMPUS DE CASCAVEL
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA AGRÍCOLA

MODELOS DE PREDIÇÃO DO COEFICIENTE DE SORÇÃO NO SOLO DE PESTICIDAS
NÃO IÔNICOS: DIFERENTES ALGORITMOS DE LOGP E UMA ABORDAGEM
ALTERNATIVA DE LOGS

RALPHO RINALDO DOS REIS

CASCAVEL – PARANÁ – BRASIL

MAIO 2013

RALPHO RINALDO DOS REIS

**MODELOS DE PREDIÇÃO DO COEFICIENTE DE SORÇÃO NO SOLO DE PESTICIDAS
NÃO IÔNICOS: DIFERENTES ALGORITMOS DE LOGP E UMA ABORDAGEM
ALTERNATIVA DE LOGS**

Tese apresentada ao Programa de Pós-Graduação *Stricto Sensu* em Engenharia Agrícola em cumprimento parcial aos requisitos para obtenção do título de Doutor em Engenharia Agrícola, área de concentração Recursos Hídricos e Saneamento Ambiental

Orientador: Dr. Silvio César Sampaio
Co-orientador: Dr. Eduardo Borges de Melo

CASCADEL – PARANÁ – BRASIL

MAIO 2013

Dados Internacionais de Catalogação-na-Publicação (CIP)
Biblioteca Central do Campus de Cascavel – Unioeste
Ficha catalográfica elaborada por Jeanine da Silva Barros CRB-9/1362

R312m Reis, Ralpho Rinaldo dos
 Modelos de predição do coeficiente de sorção no solo de pesticidas não iônicos: diferentes algoritmos de logP e uma abordagem alternativa de logS. / Ralpho Rinaldo dos Reis — Cascavel, PR: UNIOESTE, 2013. 71 f. ; 30 cm.

 Orientador: Prof. Dr. Silvio César Sampaio
 Co-orientador: Prof. Dr. Eduardo Borges de Melo
 Tese (Doutorado) – Universidade Estadual do Oeste do Paraná.
 Programa de Pós-Graduação *Stricto Sensu* em Engenharia Agrícola, Centro de Ciências Exatas e Tecnológicas.
 Bibliografia.

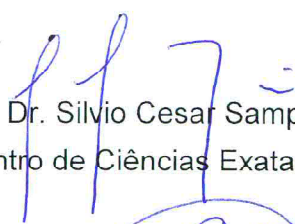
 1. Riscos ambientais. 2. Modelagem. 3. Parâmetro hidrofóbico. 4. Solo – Contaminação. 5. Solubilidade em água. I. Universidade Estadual do Oeste do Paraná. II. Título.

CDD 21. ed. 631.4

RALPHO RINALDO DOS REIS

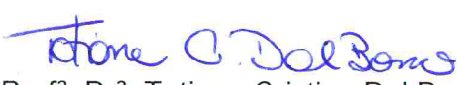
"Modelos de predição do coeficiente de sorção no solo de pesticidas não iônicos:
diferentes algoritmos de logP e uma abordagem alternativa de logS"

Tese apresentada ao Programa de Pós-Graduação "*Stricto Sensu*" em Engenharia Agrícola em cumprimento parcial aos requisitos para obtenção do título de doutor em Engenharia Agrícola, área de concentração Recursos Hídricos e Saneamento Ambiental, **aprovada** pela seguinte banca examinadora:


Orientador: Prof. Dr. Silvio Cesar Sampaio

Centro de Ciências Exatas e Tecnológicas, UNIOESTE


Prof. Dr. Raimundo Clécio Dantas Muniz Filho
Departamento de Química, UFPI


Profª. Drª. Tatiane Cristina Dal Bosco
Coordenação de Engenharia Ambiental, UTFPR


Prof. Dr. Eduardo Borges de Melo
Centro de Ciências Médicas e Farmacêuticas, UNIOESTE


Prof. Dr. Miguel Angel Uribe Opazo
Centro de Ciências Exatas e Tecnológicas, UNIOESTE

Cascavel, 17 de maio 2013.

BIOGRAFIA

Ralpho Rinaldo dos Reis, nascido em 28 de junho de 1965, possui bacharelado e licenciatura em Química pela Universidade Estadual de Campinas - UNICAMP (1987) e mestrado em Química, na área de Química Inorgânica, também pela Universidade Estadual de Campinas (1994). Atuou como professor no ensino médio e técnico entre 1991 e 1998. Atua, desde 1999, como Professor Assistente na área de Química no curso de graduação em Farmácia da Universidade Estadual do Oeste do Paraná – UNIOESTE. Foi diretor do Centro de Ciências Médicas e Farmacêuticas no quadriênio 2004/2007. Participa do Colegiado do Curso de Farmácia e do Grupo de Pesquisa em Química Aplicada. Possui trabalhos de pesquisa nas áreas de Química de Materiais, Biomateriais, Fármacos e Meio Ambiente. Atualmente ocupa o cargo de diretor de pesquisa na Pró-Reitoria de Pesquisa e Pós-Graduação da UNIOESTE.

As ciências têm as raízes amargas, porém os frutos são doces.
- Aristóteles

Dedico esta tese com todo meu amor para minha esposa Alessandra e minha filha Sophia. Também a dedico para meus pais, Lúcia e Antonio, por tudo que fizeram por mim.

AGRADECIMENTOS

A Deus, que por sua infinita misericórdia e graça, permitir a conclusão desta etapa tão importante em minha vida;

Ao Prof. Dr. Silvio César Sampaio, pela orientação, apoio, confiança depositada e principalmente pela grande amizade;

Ao Prof. Dr. Eduardo Borges de Melo, pelos valiosos conhecimentos transmitidos, incentivos, amizade e convivência;

Ao Prof. Dr. Miguel Angel Uribe Opazo, por tudo que aprendi na área de estatística nos últimos anos;

À UNIOESTE, à Fundação Araucária e ao Governo do Estado do Paraná, pelos recursos financeiros que permitiram a execução deste trabalho;

Ao corpo docente e servidores do PGEAGRI, pelo suporte teórico, técnico e administrativo que permitiu a conclusão deste trabalho;

Ao Colegiado do Curso de Farmácia, em especial aos colegas da área de Química, pelo apoio durante meu afastamento para o doutorado;

A todas as pessoas, que de forma direta ou indireta, colaboraram para a realização deste trabalho de tese.

RESUMO

A coleta de dados relativos aos danos causados pelos pesticidas sobre o meio ambiente e seus ecossistemas é lenta e onerosa. Desta maneira, grandes incentivos têm sido destinados às pesquisas que visam à construção de modelos matemáticos para predição de propriedades físicas, químicas ou biológicas de interesse ambiental. O coeficiente de sorção no solo normalizado para o conteúdo de carbono orgânico (K_{oc}) é um importante parâmetro físico-químico utilizado nas avaliações de riscos ambientais das substâncias lançadas no meio ambiente. Assim, vários modelos para predição de $\log K_{oc}$, utilizando o parâmetro hidrofóbico ($\log P$) ou o logaritmo da solubilidade em água ($\log S$) como descritores, têm sido publicados nas últimas décadas. Muitas vezes, em virtude da ausência de valores experimentais confiáveis de $\log P$ ou $\log S$, são usados algoritmos para o cálculo dessas propriedades. Apesar da disponibilidade e facilidade de acesso a diversos algoritmos para tal finalidade, os artigos científicos não descrevem o procedimento adotado para escolha do algoritmo usado nos estudos QSPR. Além disto, a forte correlação entre $\log P$ e $\log S$ impede que sejam usados em uma mesma equação obtida por regressão linear múltipla. Como o processo de sorção de um composto químico no solo está relacionado tanto com sua solubilidade em água como com sua partição água/matéria orgânica, espera-se que modelos que sejam capazes de combinar essas duas informações possam gerar resultados mais realistas. Este trabalho de tese é constituído de dois artigos. No primeiro artigo, foi feito um estudo para verificar a influência da escolha do algoritmo de $\log P$ na modelagem de $\log K_{oc}$. Foram construídos modelos que relacionam $\log K_{oc}$ com $\log P$ a partir de diferentes algoritmos livres disponíveis. Todos os modelos foram avaliados quanto às suas qualidades estatísticas e poder de predição. Os resultados obtidos mostraram claramente que uma escolha arbitrária deste algoritmo pode não levar ao melhor modelo de predição. Por outro lado, uma boa escolha pode conduzir à obtenção de modelos simples com qualidades estatísticas e poder de predição comparáveis a de modelos mais complexos. No segundo artigo, o objetivo foi a proposição de uma abordagem alternativa para a modelagem de $\log K_{oc}$, utilizando um descritor simples de solubilidade, aqui designado como logaritmo da solubilidade corrigida pela partição octanol/água ($\log S_p$). Assim, foram construídos modelos com tal descritor e também com os descritores convencionais $\log P$ e $\log S$, isolados ou associados com outras variáveis explicativas de fácil interpretação físico-química. Os modelos obtidos foram validados e comparados com outros modelos publicados anteriormente. Os resultados mostraram que o uso do descritor $\log S_p$ em substituição aos descritores convencionais conduziu à obtenção de modelos simples com qualidades estatísticas e poder de predição superiores a de outros modelos mais complexos encontrados na literatura.

PALAVRAS – CHAVE: riscos ambientais; modelagem; parâmetro hidrofóbico; contaminação do solo; solubilidade em água; QSPR.

ABSTRACT

Collecting data on pesticide effects on the environment and several ecosystems is a slow and costly process. Therefore, significant research efforts have been focused on developing mathematical models to predict physical, chemical or biological properties of environmental interest. The soil sorption coefficient normalized to organic carbon content (K_{oc}) is a physicochemical key parameter used in environmental risk assessments of substances released into the environment. Thus, several $\log K_{oc}$ prediction models that use hydrophobic parameter ($\log P$) or the logarithm of water solubility ($\log S$) as descriptor have been reported in the last decades. Mostly, due to the lack of reliable experimental values of $\log P$ or $\log S$, algorithms are used to calculate such properties. Despite the availability and easiness to access several algorithms for this purpose, scientific studies do not describe the procedure adopted to choose the algorithm used in quantitative structure-property relationship (QSPR) studies. Furthermore, the strong correlation between $\log P$ and $\log S$ prevents their application in the same mathematical equation obtained by multiple linear regression method. Since the sorption process of a chemical compound in soil is related both to its water solubility and its water/organic matter partition, it is expected models that are able to combine these two properties will can record more realistic results. This doctoral dissertation consists of two scientific papers. In the first one, a study was carried out to check the influence of choosing $\log P$ algorithm on $\log K_{oc}$ modeling. Models were constructed to relate $\log K_{oc}$ with $\log P$ according to different freeware algorithms. All models were assessed based on their statistic qualities and predictive power. The obtained results clearly showed that an arbitrary choice of the algorithm may not result in the best prediction model. On the other hand, a good choice can lead to obtaining simple models with statistic qualities and predictive power comparable to more complex models. The second paper aims at proposing an alternative approach for $\log K_{oc}$ modeling, using simple descriptor of solubility, here referred as logarithm of corrected solubility by octanol/water partition ($\log S_p$). Thus, models were built with this descriptor and also with $\log P$ and $\log S$ conventional descriptors, which are isolated or associated with other explicative variables of easy physicochemical interpretation. The obtained models were validated and compared to other models previously published. The results showed that the use of $\log S_p$ descriptor to replace the conventional ones led to obtaining simple models with statistic qualities and predictive power that are higher than other more complex models already found in literature.

KEYWORDS: environmental risks; modeling; hydrophobic parameter; soil contamination; water solubility; QSPR.

SUMÁRIO

LISTA DE FIGURAS	Ix
LISTA DE TABELAS	X
LISTA DE ABREVIATURAS E SÍMBOLOS	Xi
ARTIGO 1	1
INFLUÊNCIA DE DIFERENTES ALGORITMOS DE $\log P$ NA MODELAGEM DO COEFICIENTE DE SORÇÃO NO SOLO DE PESTICIDAS NÃO IÔNICOS	
ARTIGO 2	22
UMA ABORDAGEM ALTERNATIVA NO USO DA SOLUBILIDADE EM ÁGUA DE PESTICIDAS NÃO IÔNICOS NA MODELAGEM DO COEFICIENTE DE SORÇÃO	
ANEXO 1	43
Fórmulas dos parâmetros estatísticos	
ANEXO 2	45
Figuras extras do artigo 1	
ANEXO 3	51
Figuras extras do artigo 2	
Anexo 4	57
Tabela dos valores de $\log K_{oc}$ e $\log P$ de todos os pesticidas usados no artigo 1	
Anexo 5	65
Tabela dos valores de $\log K_{oc}$ e dos descritores de todos os pesticidas usados no artigo 2	

LISTA DE FIGURAS

FIGURAS DO ARTIGO 1

Figura 1	Gráfico da validação cruzada LNO do modelo 2 (ALOGPS)	10
Figura 2	Gráfico dos valores observados em função dos valores previstos de $\log K_{oc}$ para o conjunto de validação B do modelo 2	12
Figura 3	Gráfico dos valores observados em função dos valores previstos de $\log K_{oc}$ para o conjunto de validação B do modelo 5	13

FIGURAS DO ARTIGO 2

Figura 1	Gráfico da validação cruzada LNO do modelo A3	31
Figura 2	Gráfico da validação cruzada LNO do modelo B3	32
Figura 3	Gráfico dos valores observados em função dos valores previstos de $\log K_{oc}$ para o conjunto de validação externa do modelo A3	33
Figura 4	Gráfico dos valores observados em função dos valores previstos de $\log K_{oc}$ para o conjunto de validação externa do modelo B3	33
Figura 5	Histograma dos resíduos do modelo A3	34
Figura 6	Histograma dos resíduos do modelo B3	34

LISTA DE TABELAS

TABELAS DO ARTIGO 1

Tabela 1	Algoritmos usados para predição de $\log P$.	5
Tabela 2	Coeficiente de correlação linear de Pearson entre os dados de $\log P$ obtidos pelos diferentes algoritmos de predição e os dados experimentais (n=137)	8
Tabela 3	Modelos para predição de $\log K_{oc}$ de pesticidas não iônicos (n=143)	8
Tabela 4	Parâmetros estatísticos dos modelos para predição de $\log K_{oc}$	9
Tabela 5	Dados da validação cruzada LNO e teste de correlação por acaso	10
Tabela 6	Dados estatísticos da 1ª validação externa (conjunto de teste A, n=20)	11
Tabela 7	Dados estatísticos da 2ª validação externa (conjunto de teste B, n=38)	13
Tabela 8	Comparação entre modelos de $\log K_{oc}$	15

TABELAS DO ARTIGO 2

Tabela 1	Classes químicas dos pesticidas usados no conjunto de treinamento	25
Tabela 2	Modelos de regressão para predição de $\log K_{oc}$ de pesticidas não iônicos (n=143)	29
Tabela 3	Parâmetros estatísticos dos modelos para predição de $\log K_{oc}$	30
Tabela 4	Dados da validação cruzada LNO e teste de correlação por acaso	31
Tabela 5	Dados estatísticos da validação externa (conjunto de teste, n=20)	32
Tabela 6	Coeficiente de correlação linear de Pearson entre as variáveis usadas nos modelos	35
Tabela 7	Comparação entre modelos de $\log K_{oc}$	37

LISTA DE ABREVIATURAS E SÍMBOLOS

ARE	Erro relativo médio
Cs	Concentração no solo (mg.kg^{-1})
Cw	Concentração na fase aaquosa (mg.L^{-1})
CCC	Coeficiente de correlação de concordância
F	Estatística F
k	Inclinação da linha de regressão sem intercepto
K_d	Coeficiente de sorção no solo linear
K_{oc}	Coeficiente de sorção no solo normalizado para o conteúdo de carbono orgânico (L.kg^{-1})
LNO	Validação leave-one-out
LOO	Validação leave-N-out
Mv	Volume médio de van der Waals dos átomos de uma molécula
MAXDP	Máxima diferença positiva entre estados intrínsecos dos átomos
OC	Conteúdo de carbono orgânico
P	Coeficiente de partição octanol/água
PRESS_{cv}	Soma dos quadrados dos resíduos da validação cruzada
PRESS_{ev}	Soma dos quadrados dos resíduos da validação externa
Q	Quantidade total de soluto (mol)
Q^2_{LNO}	Coeficiente de determinação da validação leave-N-out
Q^2_{LOO}	Coeficiente de determinação da validação leave-one-out
QSPR	Relações quantitativas estrutura-propriedade
R^2	Coeficiente de determinação
R^2_0	Coeficiente de determinação da linha de regressão sem intercepto
r^2_m	Coeficiente de determinação modificado
R^2_{PRED}	Coeficiente de determinação da validação externa
RSS	Soma dos quadrados dos resíduos
S	Solubilidade em água (mol.L^{-1})
SEC	Desvio padrão da calibração
SECV	Desvio padrão da validação cruzada
SEP	Desvio padrão da predição
S_o	Solubilidade na fase orgânica (mol.L^{-1})
S_p	Solubilidade em água corrigida pela partição octanol/água (mol.L^{-1})
S_w	Solubilidade na fse aquosa (mol.L^{-1})
VAR	Variação topológico variação

ARTIGO 1 - INFLUÊNCIA DE DIFERENTES ALGORITMOS DE LOGP NA MODELAGEM DO COEFICIENTE DE SORÇÃO NO SOLO DE PESTICIDAS NÃO IÔNICOS

Versão adaptada do artigo aceito para publicação na revista *Water Research*

1. Introdução

Os produtos químicos apresentam um papel essencial na qualidade de vida da sociedade moderna. Compostos como pesticidas, fármacos, solventes, dentre outros, estão presentes em praticamente todos os setores da atividade econômica. Assim, a visão idealizada de um mundo livre de produtos químicos é incompatível com o atual modo de vida. Como consequência desse uso disseminado, a preocupação com o destino final e os riscos ambientais e à saúde dessas substâncias vêm aumentando por parte de alguns segmentos da sociedade (Mackay e Webster, 2003). Desta forma, pesquisas que levantem informações acerca das propriedades dos produtos químicos são necessárias para o planejamento de ações que objetivem a recuperação de áreas contaminadas ou a preservação de locais ainda não degradados.

Os pesticidas são substâncias amplamente usadas na agricultura, com a finalidade de se obter maior produtividade, melhor qualidade das culturas e menores custos de mão de obra e energia. Entretanto, apenas uma pequena parcela (cerca de 0,1%) do que é aplicado nas lavouras atinge seus objetivos específicos. O restante tende a mover-se nos diferentes compartimentos ambientais, podendo contaminar tanto as águas superficiais como as subterrâneas (Pimentel e Levitan, 1986; Arias-Estévez et al., 2008).

O coeficiente de sorção no solo normalizado para o conteúdo de carbono orgânico (K_{oc}) é uma importante propriedade físico-química, que pode ser usado para avaliar o risco ambiental das substâncias químicas lançadas no meio ambiente. Tal coeficiente é definido pela seguinte equação:

$$K_{oc} = K_d / OC = C_s / (C_w \times OC) \quad \text{Eq. 1}$$

Onde C_s e C_w são, respectivamente, a concentração do componente químico sorvido pelo solo (mg.kg^{-1}) e a concentração da solução aquosa circundante (mg.L^{-1}); OC é o conteúdo de carbono orgânico em unidades de g de OC/g de solo seco; e K_d é o coeficiente de sorção linear (Doucette, 2003). Quanto menor for o valor do coeficiente de sorção de um dado composto, maior será sua mobilidade no solo e maior será seu potencial para contaminar as águas subterrâneas (Dearden, 2002; Arias-Estévez et al., 2008).

O valor do coeficiente K_{oc} de um composto químico pode ser mensurado experimentalmente pelo método de equilíbrio em batelada (OECD, 2000; ASTM, 2001) ou pelo método da coluna de solo (Lee et al., 1991). A extensão da sorção de um composto

químico é determinada pela sua estrutura molecular, mas depende também de fatores relacionados com as características intrínsecas do solo, tais como: tamanho de partículas, porosidade, pH e conteúdo de carbono orgânico presente (Dearden, 2002).

O conteúdo de carbono orgânico é fator determinante no processo de sorção de compostos não iônicos. Desta maneira, a hidrofobicidade desses compostos é a força motriz no processo de sorção no solo (Wen et al., 2012). É importante notar que o coeficiente K_d varia de solo para solo e, geralmente, é normalizado a fim de se obter o coeficiente K_{oc} , pois para muitos compostos orgânicos, tais como os hidrofóbicos, a sorção no solo é diretamente proporcional ao teor de carbono orgânico. Assim, quando $OC > 0,001$ e o composto químico em estudo for não iônico ou neutro, K_{oc} pode ser considerado uma constante para esta substância. Desta forma, K_{oc} reflete o fato do carbono orgânico ser o principal domínio de sorção de compostos hidrofóbicos. No entanto, a abordagem de K_{oc} pode não ser adequada quando os solos possuírem teores baixos de carbono orgânico ou elevados de argilas, e também pode não ser aplicável para compostos que apresentem grupos funcionais altamente polarizáveis que interajam com sítios significativamente polares ou carregados presentes na superfície das partículas de solo (Doucette, 2003; Wen et al., 2012). Uma descrição mais detalhada das interações intermoleculares envolvidas no processo de sorção de compostos orgânicos no solo pode ser obtida nos trabalhos de Hassett e Banwart (1989), von Oepen et al. (1991) e Allen-King et al. (2002).

A necessidade de um banco completo, confiável e atualizado dos valores de K_{oc} , tem levado ao desenvolvimento de uma grande quantidade de modelos para predição desse parâmetro. Vários estudos de relação quantitativa estrutura-propriedade (QSPR), com base na estrutura molecular dos compostos químicos, foram disponibilizados nas últimas décadas. Assim, na modelagem de K_{oc} , diversos tipos de descritores têm sido usados, tais como propriedades físico-químicas, fatores cromatográficos, fatores estéricos, relações de energia de solvatação lineares (LSER), índices topológicos e parâmetros quânticos (Gawlik et al., 1997; Müller, 1997; Andersson et al., 2002; Doucette, 2003; Huuskonen, 2003; Kahn et al., 2005; Nguyen e Ball, 2005; Gramatica et al., 2007; Razzaque et al., 2008; Goudarzi et al., 2009; Wen et al., 2012). Porém, a grande maioria dos modelos QSPR propostos foi construída a partir de conjuntos pequenos de dados, ou são de classes específicas ou não foram validados adequadamente.

A não validação de um modelo de predição inviabiliza a sua aplicação. Desta maneira, para um modelo ser considerado confiável, é necessário que se conheça seu poder de predição a partir de procedimentos rigorosos de validação (Kiralj e Ferreira, 2009; Chirico e Gramatica, 2011; Roy et al., 2012).

Muitos modelos foram construídos a partir da relação entre $\log K_{oc}$ e os logaritmos da solubilidade em água ($\log S$) ou do coeficiente de partição octanol/água ($\log P$) (Gawlik et al.,

1997; Doucette, 2003; Razzaque et al., 2008). Este importante coeficiente de partição é dado pela equação:

$$P = S_o / S_w \quad \text{Eq. 2}$$

Onde S_o e S_w são as concentrações (mol.L^{-1}) do composto nas fases orgânica e aquosa de um sistema octanol/água em equilíbrio.

A forte correlação entre estes dois descritores ($\log P$ e $\log S$) indica que tais modelos podem ser considerados idênticos ou paralelos. O parâmetro $\log P$ é um bom descritor para modelagem do coeficiente de sorção no solo e tem sido utilizado em estudos QSPR para diversas classes de compostos orgânicos, inclusive pesticidas (Lohnimger, 1994; Sablhc et al., 1995). Seu uso na modelagem da sorção no solo é comum em modelos com uma ou múltiplas variáveis. Baker et al. (2000) demonstraram que as relações entre $\log K_{oc}$ com $\log P$ desviam-se da linearidade para compostos com $\log P > 5$. Wen et al. (2012) propuseram um modelo não linear entre $\log K_{oc}$ e $\log P$, para um conjunto heterogêneo de 594 compostos orgânicos, mas observaram a existência de linearidade na faixa de $\log P$ entre 0,5-7,5.

O uso de valores experimentais de $\log P$ proporciona a obtenção de modelos QSPR mais realistas. Todavia, em muitos casos, os dados experimentais não são disponíveis e a determinação de tais valores pode ser muito trabalhosa e demandar muito tempo e recursos. Tal situação, juntamente com a existência de uma grande quantidade de novas substâncias, naturais ou sintetizadas, dificulta a obtenção de um banco de dados atualizado. Desta maneira, abordagens computacionais para cálculo de $\log P$ são ferramentas valiosas em estudos QSPR (Melo e Ferreira, 2009).

Atualmente, existem diversos algoritmos, comerciais ou livres, disponíveis para cálculo de $\log P$ (Tetko, 2003). Basicamente, as duas principais metodologias usadas para este cálculo são: (a) o método da subestrutura com base nos fragmentos ou átomos, e (b) o método molecular por inteiro, baseado em propriedades moleculares (Mannhold e van de Waterbeend, 2001).

Os estudos QSPR do coeficiente de sorção no solo simplesmente omitem como ou porque determinado algoritmo foi escolhido para cálculo de $\log P$, ou se há algum critério objetivo nesta escolha de modo a garantir que se gere o melhor modelo de regressão. Assim, o objetivo deste trabalho foi avaliar a influência de diferentes algoritmos livres disponíveis para cálculo de $\log P$ na modelagem de $\log K_{oc}$, de modo a determinar qual ou quais são os mais adequados a serem usados nestes modelos.

A opção pelo uso de algoritmos livres visou à obtenção de valores de $\log P$ de maneira rápida e simples, a fim de facilitar a reprodução dos resultados. Por fim, os modelos obtidos foram testados e validados, conforme as recomendações da literatura específica da

área (Kiralj e Ferreira, 2009; Chirico e Gramatica, 2011; Roy et al., 2012), para garantir que sejam confiáveis e possam ser utilizados com propósitos de predição.

2. Material e Método

2.1. Dados experimentais de K_{oc}

Os valores experimentais dos coeficientes de sorção no solo ($\log K_{oc}$) de 163 pesticidas orgânicos não iônicos foram obtidos na literatura (Sabljić et al., 1995; Gramatica et al., 2000). Os dados foram divididos em um conjunto de treinamento com 143 compostos e um conjunto de validação externa com 20 compostos (conjunto de teste A). Os valores experimentais de $\log K_{oc}$ estão disponíveis no Anexo 4. Os mesmos conjuntos foram usados nos trabalhos de Gramatica et al. (2000), Huuskonen (2003) e Duchowicz et al. (2007).

Os compostos que constituem o conjunto de treinamento são heterogêneos e pertencem a diversas classes (6 acetonitrilas; 29 carbamatos; 8 dinitroanilinas; 8 organoclorados; 28 organofosfatos; 44 feniluréias; 13 triazinas; 7 di- e triazolas). Seus valores de $\log K_{oc}$ variam entre 0,42 e 5,31, enquanto para o conjunto de teste A, os valores variam entre 0,56 e 4,50.

Um segundo conjunto de validação externa (conjunto de teste B), contendo dados de 38 pesticidas, foi extraído do trabalho de Huuskonen (2003). Para o conjunto de teste B, os valores de $\log K_{oc}$ variam de 0,30 a 6,00. Portanto, um conjunto de treinamento foi usado para a construção dos modelos, e dois conjuntos de validação externa (A e B) foram utilizados para avaliação das capacidades de predição. O fato desses conjuntos de dados terem sido usados em outros trabalhos permitiu a comparação dos nossos resultados com os da literatura (Gramatica et al., 2000; Huuskonen, 2003; Duchowicz et al., 2007).

2.2. Determinação de $\log P$ e Modelos QSPR

Os algoritmos livres ALOGP, ALOGPS, ACLOGP, MLOGP, KOWWIN, XLOGP2 e XLOGP3 foram usados para a determinação dos valores de $\log P$ de cada composto. Na Tabela 1 estão os detalhes dos algoritmos usados (Mannhold e van de Waterbeemd, 2001; Tetko et al., 2001; Cheng et al., 2007; Todeschini e Consomi, 2009).

Os valores de $\log P$ experimentais e calculados por cada algoritmo foram obtidos a partir da introdução das *Simplified Molecular Input Line Entry Specification* (SMILES) de cada composto no programa ALOGPS 2.1 (<http://www.vcclab.org/lab/alogps>). Os valores de $\log P$ de todos compostos do conjunto de treinamento e dos conjuntos de teste A e B estão apresentados no Anexo 4.

Os modelos para estimação de $\log K_{oc}$ em função de $\log P$ foram obtidos por regressão linear simples com o uso dos programas Minitab 15 e QSAR Modeling (software livre desenvolvido e cedido pelo Laboratório de Quimiometria Teórica e Aplicada da UNICAMP). Assim, como foram usados sete algoritmos diferentes para cálculo de $\log P$, foram obtidos sete modelos para predição de $\log K_{oc}$. Em seguida, os modelos foram avaliados em relação às suas qualidades estatísticas e comparados entre si.

Tabela 1

Algoritmos usados para predição de $\log P$.

Algoritmos	Abordagem de cálculo	Número de compostos do conjunto de calibração
ALOGP	115 contribuições de átomos	8364
ALOGPS	75 índices eletrotopológicos	12908
ACLOGP	369 contribuições de átomos	5000
MLOGP	13 parâmetros de regressão linear múltipla	1230
KOWWIN	150 contribuições de grupos e 250 fatores de correção	2351
XLOGP2	90 contribuições de átomos e 10 fatores de correção	1863
XLOGP3	87 contribuições de átomos e 2 fatores de correção	8199

2.3. Qualidade estatística e poder de predição

A qualidade estatística de modelos de predição tem sido verificada a partir da avaliação do grau de ajuste e de significância, das validações interna e externa, da avaliação da robustez, do teste da correlação por acaso e da análise dos resíduos (Gaudio e Zandonade, 2001; Kiralj e Ferreira, 2009).

O grau de ajuste é medido em termos de o quanto a especificação do modelo se adequa aos dados experimentais. Assim, um bom modelo é aquele no qual o valor calculado pela equação é o mais próximo possível do valor observado. Esta avaliação é feita a partir do cálculo do coeficiente de determinação (R^2), da soma dos quadrados dos resíduos (RSS) e do desvio padrão da calibração (SEC). Para que os modelos QSPR sejam considerados ajustados, recomenda-se que apresentem valores de $R^2 > 0,6$ e o mais próximo possível de

1 (100%), que os valores de RSS e de SEC sejam os mais próximos possíveis de zero. Isto significa que o modelo deve ser capaz de explicar pelo menos 60% da variabilidade da propriedade de interesse (Gaudio e Zandonade, 2001).

A significância dos modelos foi avaliada pelo teste estatístico F . Este teste permitiu verificar se a variabilidade de Y em função de X foi significativamente explicada pela regressão. O valor da estatística F obtida na regressão deve ser maior que um valor de referência tabelado (considera-se $F_{p,n-p-1}$, onde p é o número de variáveis do modelo e n é o número de compostos). Quanto maior for o valor da estatística F obtida na regressão em comparação com o valor tabelado, mais significativo será o modelo, sendo comumente utilizado o nível de significância de 5% ($\alpha = 0,05$) (Gaudio e Zandonade, 2001).

De acordo com a literatura, procedimentos rigorosos de validação devem ser executados a fim de garantir a confiabilidade estatística dos modelos de regressão (Kiralj e Ferreira, 2009; Chirico e Gramatica, 2011; Roy et al., 2012). A validação interna ou validação cruzada *leave-one-out* (LOO) é feita da seguinte forma: (a) exclui-se um a um cada objeto do modelo; (b) reconstrói-se o modelo sem esse objeto e calcula-se o valor do objeto excluído. Por fim, calculam-se a soma dos quadrados das diferenças entre os valores previstos e observados ($PRESS_{cv}$), o coeficiente de determinação da validação cruzada (Q^2_{LOO}) e o desvio padrão da validação cruzada ($SECV$). Um modelo pode ser considerado aprovado se o valor de $SECV$ for próximo de zero e se Q^2_{LOO} for maior que 0,5 (Todeschini e Consonni, 2009).

O teste de robustez ou validação cruzada *leave-N-out* (LNO) avalia se os modelos não são significativamente afetados por pequenas e deliberadas mudanças em seus parâmetros. Neste teste, o modelo é recalculado a partir da retirada sistemática de até N elementos do conjunto de treinamento. Geralmente, usa-se N máximo na faixa de 25-30% do conjunto original. Neste estudo, N variou de 1 a 40, com seis repetições para cada valor de N . Esta técnica produz uma perturbação maior que a validação LOO. Para um modelo ser considerado robusto, o valor de Q^2_{LNO} médio deve ser maior que 0,5 e o mais próximo possível do valor de Q^2_{LOO} (Melagraki et al., 2007).

A possibilidade da correlação por acaso pode ser feita a partir do teste *y-randomization*. Esse teste permite avaliar se a relação entre as variáveis explicativas e a variável resposta foi decorrente de um mero acaso. Os modelos são recalculados após a randomização do vetor Y (variável resposta) e assim espera-se uma piora significativa nos parâmetros estatísticos dessas novas regressões. O procedimento foi repetido 50 vezes e foram construídos dois gráficos, nos quais os parâmetros R^2 e Q^2_{LOO} (eixo y) foram colocados em função, no eixo x , de valores de um coeficiente de correlação $R(Y_o, Y_r)$, obtidos a partir dos valores de Y originais e randomizados. Para descartar a possibilidade de correlação por acaso, os interceptos destes gráficos devem ser menores que 0,3 para R^2 e 0,05 para Q^2_{LOO} (Eriksson et al., 2003).

O poder ou habilidade de predição de um modelo de regressão é medido a partir da validação externa. A validação externa dos modelos foi feita através da predição dos valores de $\log K_{oc}$ para os conjuntos de teste. O poder de predição de modelos de regressão pode ser avaliado pelos coeficientes de determinação da validação externa (R^2_{PRED}) e dos coeficientes de determinação modificados da validação externa (r^2_m). Recomenda-se que os valores de tais coeficientes sejam maiores que 0,5 em ambos os casos (Todeschini e Consoni, 2009; Mitra et al., 2011).

Os valores da soma dos quadrados dos resíduos das predições da validação externa ($PRESS_{ev}$), do desvio padrão da predição (SEP) e do erro relativo médio (ARE) também foram determinados, os quais devem ser os menores possíveis, bem como as inclinações das linhas de regressão que passam pela origem, entre os valores observados e previstos. Recomenda-se que os valores destas inclinações (k e k') estejam entre 0,85 e 1,15 e que o módulo da diferença entre os coeficientes R^2_o e R^2_o seja menor que 0,3 (Golbraik e Tropsha, 2002; Golbraik et al., 2003; Tropsha et al., 2003).

A análise dos gráficos de resíduos foi feita com a finalidade de se verificar se os resíduos obtidos apresentam distribuição normal e média zero. As fórmulas usadas para o cálculo de todos os parâmetros estatísticos citados acima são apresentadas no Anexo 1.

O uso destas ferramentas de validação garantiu que este estudo esteja de acordo com as recomendações da literatura para os estudos QSPR (Kiralj e Ferreira, 2009; Chirico e Gramatica, 2011; Roy et al., 2012). Após esses procedimentos de validação, os melhores modelos obtidos foram comparados com outros modelos já apresentados na literatura (Gramatica et al., 2000; Huuskonen, 2003; Duchowicz et al., 2007), os quais empregaram os mesmos conjuntos de pesticidas não iônicos em suas construções e validações.

3. Resultados e Discussão

3.1. Correlação entre os valores de $\log P$ experimental e calculados

Os dados experimentais de $\log P$ nem sempre são disponíveis para todos os compostos usados em um estudo QSPR. Assim, a primeira etapa deste trabalho consistiu em avaliar o coeficiente de correlação linear de Pearson (r) dos valores de $\log P$ gerados por cada um dos sete algoritmos testados e os respectivos dados experimentais. Tal informação permitiu uma avaliação preliminar de quais algoritmos estariam mais ajustados com os valores observados e, conseqüentemente, poderiam gerar os melhores e mais realistas modelos QSPR. Os resultados estão apresentados na Tabela 2 e observou-se que os algoritmos XLOGP3, KOWWIN e ALOGPS apresentaram as melhores correlações entre os dados calculados e experimentais.

Tabela 2

Coeficiente de correlação linear de Pearson entre os dados de $\log P$ obtidos pelos diferentes algoritmos de predição e os dados experimentais (n=137).

Algoritmos de $\log P$	r
ALOGP	0,933
ALOGPS	0,972
ACLOGP	0,921
MLOGP	0,815
KOWWIN	0,973
XLOGP2	0,892
XLOGP3	0,975

3.2. Modelos de predição de $\log K_{oc}$

Os modelos para predição de $\log K_{oc}$, considerando os diferentes algoritmos para cálculo de $\log P$, são apresentados na Tabela 3. Todos os modelos são lineares e possuem apenas uma variável explicativa, e os parâmetros necessários para a avaliação da qualidade estatística dos mesmos estão na Tabela 4.

Tabela 3

Modelos para predição de $\log K_{oc}$ de pesticidas não iônicos (n=143).

Modelos	Equação
1	$\log K_{oc} = 1,030 + 0,576 \text{ ALOGP}$
2	$\log K_{oc} = 0,945 + 0,560 \text{ ALOGPS}$
3	$\log K_{oc} = 0,999 + 0,548 \text{ ACLOGP}$
4	$\log K_{oc} = 1,200 + 0,574 \text{ MLOGP}$
5	$\log K_{oc} = 1,100 + 0,497 \text{ KOWWIN}$
6	$\log K_{oc} = 1,300 + 0,493 \text{ XLOGP2}$
7	$\log K_{oc} = 1,000 + 0,536 \text{ XLOGP3}$

Os resultados do teste F (Tabela 4) mostraram que todas as regressões são estatisticamente significativas a 5% de significância ($\alpha=0,05$), pois os valores do parâmetro F são maiores que o valor de F tabelado ($F_{1,141} = 3,9$) em todos os casos. Ademais, observou-se que os modelos mais significativos a 5% foram construídos a partir dos algoritmos ALOGPS, KOWWIN e XLOGP3 (modelos 2, 5 e 7 respectivamente).

Os dados estatísticos também mostraram que os modelos mais ajustados foram obtidos a partir dos algoritmos ALOGPS, KOWWIN e XLOGP3. Estes modelos apresentaram os maiores valores de R^2 e Q^2_{LOO} e os menores valores de SEC e $SECV$.

Notou-se claramente que a escolha do algoritmo afetou significativamente a qualidade estatística dos modelos de regressão. Por exemplo, o modelo 4, obtido a partir do algoritmo MLOGP, apresentou qualidade estatística muito abaixo dos demais, sendo inclusive rejeitado pelo critério de grau de ajuste ($R^2 < 0,6$).

Os valores de *SEC* devem ser os menores possíveis, porém dentro de limites impostos pelo erro experimental das medidas da variável resposta. Lohninger (1994) mostrou que há limites na precisão das medidas experimentais de $\log K_{oc}$. Em um estudo realizado em um conjunto de pesticidas, com pelo menos quatro valores diferentes de $\log K_{oc}$ para cada composto, o desvio padrão médio dessas medidas foi 0,44 unidades logarítmicas. A modelagem de uma variável resposta depende da qualidade dos dados experimentais disponíveis. Assim, todos os modelos aqui obtidos apresentaram valores de *SEC* compatíveis com este erro experimental.

Todos os modelos foram aprovados na validação cruzada LOO, pois apresentaram Q^2_{LOO} maior que 0,5. Desta forma, a partir dos dados apresentados na Tabela 4, em relação ao ajuste dos dados e à validação cruzada LOO, é possível colocar os modelos em ordem decrescente de qualidade estatística em função do algoritmo utilizado:

ALOGPS > KOWWIN > XLOGP3 > ACLOGP > ALOGP > XLOGP2 > MLOGP

Tabela 4

Parâmetros estatísticos dos modelos para predição de $\log K_{oc}$.

Modelos	R^2	<i>SEC</i>	<i>RSS</i>	<i>F</i>	Q^2_{LOO}	<i>SECV</i>	<i>PRESS_{cv}</i>
1	0,632	0,536	40,462	242,22	0,619	0,542	41,941
2	0,744	0,447	28,139	410,04	0,735	0,451	29,146
3	0,667	0,510	36,637	282,22	0,656	0,515	37,858
4	0,567	0,581	47,569	184,96	0,552	0,587	49,247
5	0,713	0,473	31,586	349,9	0,701	0,48	32,914
6	0,619	0,545	41,909	228,99	0,606	0,551	43,354
7	0,695	0,488	33,535	321,37	0,686	0,492	34,578

A validação cruzada LNO e o teste de correlação por acaso são importantes ferramentas para verificação da qualidade interna dos modelos. Os dados da validação cruzada LNO mostraram que todos os modelos apresentaram valores de Q^2_{LNO} médio próximos ao valor de seu respectivo Q^2_{LOO} (Tabela 5). O gráfico da validação cruzada para o modelo 2 (ALOGPS) apresentado na Figura 1 mostra que o modelo é estável e resiste a pequenas variações. Os gráficos dos demais modelos estão disponíveis no Anexo 2 e conduzem à mesma conclusão. Entretanto, apesar de todos os modelos poderem ser considerados robustos, mais uma vez ficou evidente que a qualidade estatística destes foi

influenciada pela escolha do algoritmo de $\log P$, pois os valores de Q^2_{LNO} médio foram bem diferentes entre si, da mesma forma como fora observado anteriormente para os valores de R^2 e Q^2_{LOO} .

A correlação por acaso foi avaliada através do teste *y-randomization* e os resultados apresentados (Tabela 5) mostraram a ausência da mesma em todos os modelos obtidos, uma vez que os valores dos interceptos dos gráficos de R^2 e de Q^2_{LOO} em função de $R(Y_o, Y_r)$ estão dentro dos limites (menores que 0,3 e 0,05 respectivamente), propostos por Ericksson et al. (2003).

Tabela 5

Dados da validação cruzada LNO e teste de correlação por acaso.

Modelos	Robustez	Correlação por acaso	
	Q^2_{LNO} médio	Intercepto R^2 vs $R(Y_o, Y_r)$	Intercepto Q^2_{LOO} vs $R(Y_o, Y_r)$
1	0,618	-0,033	-0,061
2	0,734	-0,033	-0,062
3	0,655	-0,043	-0,074
4	0,551	-0,030	-0,059
5	0,700	-0,034	-0,064
6	0,604	-0,029	-0,059
7	0,685	-0,036	-0,065

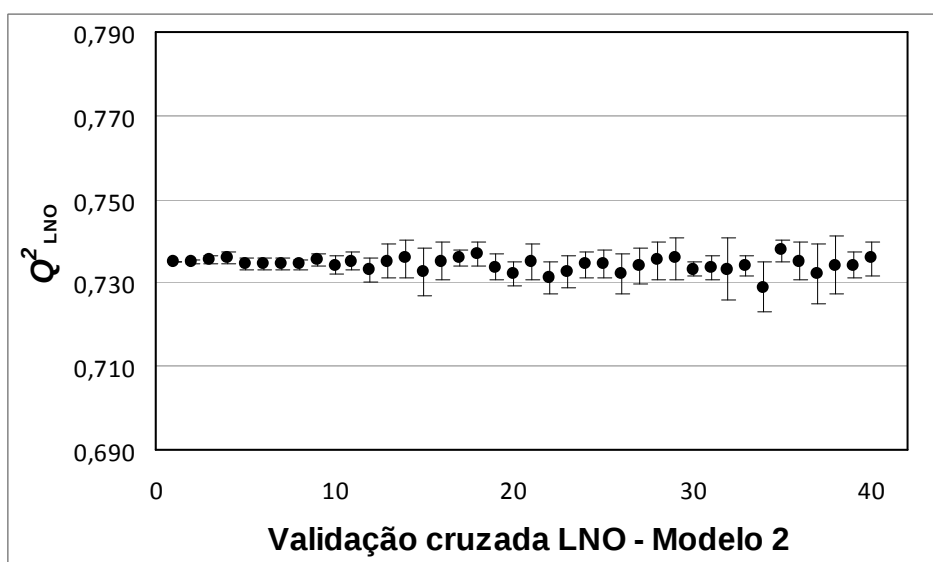


Figura 1. Gráfico da validação cruzada LNO do modelo 2 (ALOGPS).

A validação externa é uma etapa primordial nos estudos QSPR. Muitos autores argumentam que somente modelos validados interna e externamente podem ser

considerados estatisticamente confiáveis e aplicáveis em propósitos práticos (Gramatica, 2007; Melo e Ferreira, 2009; Chirico e Gramatica, 2011).

Os resultados obtidos para a validação externa com o conjunto de teste A (Tabela 6) mostraram claramente que o poder de predição dos modelos foi significativamente afetado pela escolha do algoritmo de logP. Observou-se que os melhores modelos foram os construídos a partir dos algoritmos KOWWIN e ALOGPS. Tais modelos apresentaram os maiores valores de R^2_{PRED} e os menores valores de SEP , $PRESS_{\text{ev}}$ e ARE . O ARE é um parâmetro estatístico que mostra a percentagem média da variação entre os valores previstos e observados, ou seja, é uma medida relativa enquanto o $PRESS_{\text{ev}}$ é uma medida absoluta dos resíduos (diferença entre os valores previstos e observados). Desta maneira, para modelos que apresentam valores de $PRESS_{\text{ev}}$ próximos, os valores de ARE podem ser considerados importantes a fim de se escolher o melhor modelo.

O modelo construído a partir do algoritmo MLOGP foi rejeitado pela validação externa ($R^2_{\text{PRED}} < 0,5$). O poder de predição dos modelos em relação ao algoritmo utilizado também pode ser expresso em ordem decrescente:

$$\text{KOWWIN} > \text{ALOGPS} > \text{ACLOGP} > \text{XLOGP3} > \text{XLOGP2} > \text{ALOGP} > \text{MLOGP}$$

Em relação aos parâmetros k , k' e $|R^2_o - R^2_o|$, todos os modelos estão dentro dos limites propostos por Tropsha et al. (2003). Entretanto, se os valores do coeficiente de determinação modificado (r^2_m) forem analisados, os modelos gerados pelos algoritmos ALOGP, MLOGP e XLOGP3 serão rejeitados, pois apresentaram valores de r^2_m (médio) menores que 0,5, considerando os dois gráficos de regressão (valores observados versus preditos e valores preditos versus observados).

Tabela 6

Dados estatísticos da 1ª validação externa (conjunto de teste A, n=20).

Modelos	R^2_{PRED}	SEP	$PRESS_{\text{ev}}$	ARE (%)	r^2_m médio	$ \Delta r^2_m $	k	K'	$ R^2_o - R^2_o $
1	0,522	0,644	8,29	28,6	0.449	0,085	0,93	1,01	0,067
2	0,668	0,536	5,75	16,1	0.622	0,127	0,97	0,99	0,050
3	0,638	0,560	6,27	24,0	0.556	0,065	0,95	1,00	0,033
4	0,437	0,698	9,75	28,1	0.335	0,111	1,04	0,90	0,139
5	0,733	0,481	4,62	18,0	0.648	0,031	1,00	0,97	0,011
6	0,550	0,624	7,80	26,4	0.511	0,032	0,92	1,03	0,020
7	0,563	0,615	7,57	22,3	0.487	0,023	0,99	0,95	0,016

Para ajudar na seleção dos melhores modelos, os resultados também foram analisados considerando critérios mais rigorosos propostos por Chirico e Gramatica (2011): R^2 maior que 0,7, Q^2_{LOO} e R^2_{PRED} maiores que 0,6. Assim, apenas os modelos obtidos a partir dos algoritmos ALOGPS e KOWWIN puderam ser considerados ajustados e com efetiva habilidade de predição. Desta forma, dos sete modelos iniciais, apenas dois atenderam aos critérios estatísticos quanto ao ajuste dos dados, às validações internas e ao poder de predição. Em relação à análise dos resíduos destes dois modelos, os gráficos disponíveis no Anexo 2 mostraram que os resíduos podem ser considerados com distribuição normal e média zero.

Para os dois melhores modelos (2 e 5), foram realizados testes com um segundo conjunto de validação externa (conjunto de teste B), cujos resultados são apresentados na Tabela 7. Assim, observou-se que ambos os modelos apresentaram resultados muito próximos. Todavia, o modelo 2 (ALOGPS) apresentou ligeira superioridade em relação ao modelo 5 (KOWWIN) no que se refere ao *ARE*. Nas Figuras 2 e 3 são apresentados os gráficos dos valores observados em função dos valores preditos de $\log K_{oc}$ para tais modelos.

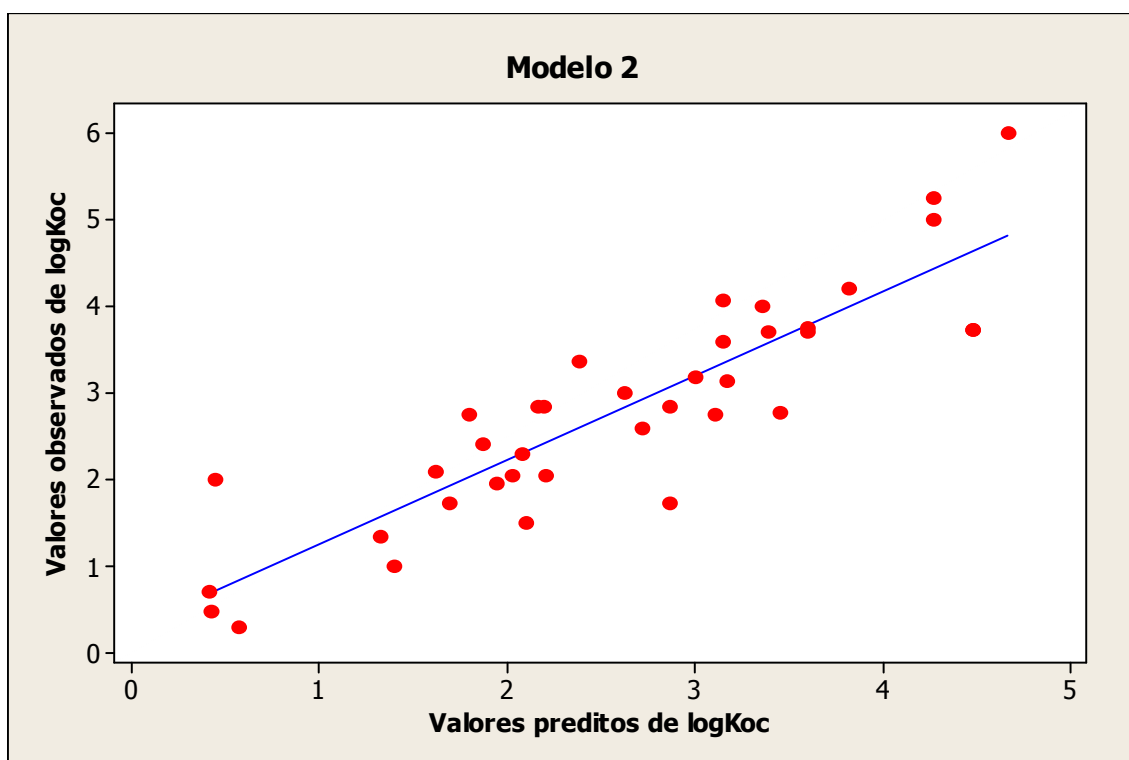


Figura 2. Gráfico dos valores observados em função dos valores preditos de $\log K_{oc}$ para o conjunto de validação B do modelo 2.

A concordância entre os valores previstos por estes dois modelos foi razoável. Para 34 dos 38 compostos testados, as diferenças entre os valores previstos no modelo 2 e os valores previstos no modelo 5 foram menores que 0,5. Oxidemeton-metil apresentou a maior

diferença entre os valores previstos (0,81), mas esta diferença está compatível com os valores da *SEC* dos modelos 2 e 5 (0,447 e 0,473 respectivamente).

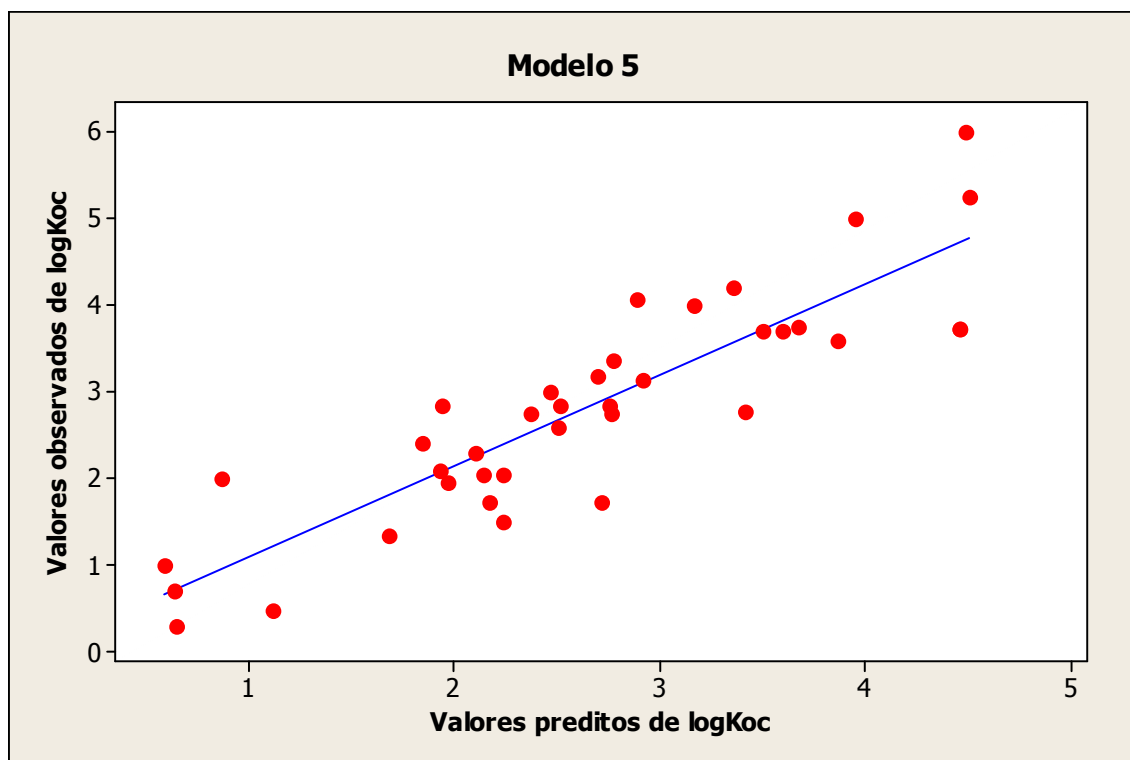


Figura 3. Gráfico dos valores observados em função dos valores preditos de $\log K_{oc}$ para o conjunto de validação B do modelo 5.

Por outro lado, os compostos amitrole e fluvalinato foram pontos discrepantes em ambos os modelos (as diferenças entre os valores previstos e experimentais $> 2,5 \times SEC$). Os valores experimentais de $\log K_{oc}$ dos compostos do conjunto de treinamento estão na faixa 0,42-5,31. O fluvalinato ($\log K_{oc}$ experimental = 6) está fora dessa faixa. Para o amitrole, Huuskonen (2003) relatou que poderia haver erro experimental no valor utilizado.

Tabela 7

Dados estatísticos da 2ª validação externa (conjunto de teste B, n=38).

Modelo	R^2_{PRED}	SEP	$PRESS_{ev}$	ARE (%)
Modelo 2 – ALOGPS	0,768	0,621	14,66	20,5
Modelo 5 – KOWWIN	0,769	0,620	14,62	23,8

De acordo com os resultados obtidos, ficou evidente que o uso de determinado algoritmo para cálculo de $\log P$, sem seleção ou comparação prévia entre eles, pode levar a resultados ruins em um estudo QSPR. Portanto, constatou-se que diferentes algoritmos podem contribuir com diferentes tipos e quantidades de informações codificadas de $\log P$,

levando assim a modelos com diferenças significativas em suas qualidades estatísticas e poder de predição (Melo e Ferreira, 2009).

A sorção de compostos não iônicos no solo ocorre a partir de um mecanismo que envolve interações hidrofóbicas, pois o principal domínio desta sorção é a matéria orgânica do solo. As forças de van der Waals são as principais forças atuantes nas interações entre as moléculas sorvidas e a superfície sorvente. Portanto, a hidrofobicidade de uma molécula é fundamental neste processo e pode ser estimada pelo valor de $\log P$. Assim, a correta medida do valor de $\log P$ é essencial para a correta determinação do coeficiente de sorção de determinado composto químico.

Em estudos QSPR, nos quais o parâmetro hidrofóbico afeta significativamente o mecanismo do fenômeno investigado, o uso como descritores de valores experimentais de $\log P$, quando disponíveis, deve sempre ser a primeira escolha. Assim, um modelo de predição deve ser utilizado somente diante da impossibilidade de obtenção dos valores experimentais, os quais são mais precisos e confiáveis do que os valores previstos.

No presente estudo, a modelagem de $\log K_{oc}$ foi executada a partir de dados estimados de $\log P$. Os algoritmos que são capazes de gerar as melhores estimativas de $\log P$ com respeito ao seu valor verdadeiro (valor experimental) certamente serão os mais adequados. O poder de predição dos algoritmos é afetado pela quantidade e diversidade estrutural dos compostos usados em suas calibrações. Desta maneira, o algoritmo a ser escolhido deve ser aquele que foi calibrado de forma a atender o espaço químico das substâncias relevantes em estudo. Um algoritmo que tenha sido calibrado para determinada classe de compostos não será adequado para estimar valores de $\log P$ para compostos estruturalmente diferentes.

Assim, na análise dos dados relacionados com as características de cada algoritmo (Tabela 1), é importante observar que ALOGPS foi calibrada a partir de um conjunto com 12.908 compostos (o maior conjunto de calibração utilizado). Portanto, devido à extensão de seu espaço químico, este algoritmo tem boas chances de produzir previsões aceitáveis de $\log P$. Desta maneira, quanto melhor for o conjunto de calibração, melhor será o espaço químico atendido pelo algoritmo.

A abordagem de cálculo de $\log P$ também é uma característica importante a ser considerada. Muitos estudos anteriores (Mannhold e van de Waterbeend, 2001) demonstraram que os algoritmos que somam as contribuições de grupos/fragmentos produzem melhores resultados do que aqueles que somam as contribuições de tipos de átomos. Além disso, os métodos aditivos que contêm fatores de correção são mais precisos do que aqueles que não utilizam tais fatores. Assim, o número de contribuições classificadas de fragmentos ou tipos de átomos e o número de fatores de correção podem afetar as previsões. Diante deste ponto de vista e dentre os algoritmos usados (Tabela 1), KOWWIN pode ser mais adequado do que os outros. Portanto, a escolha do algoritmo para predição

de $\log P$ não pode ser arbitrária, mas deve ser baseada na estrutura química dos compostos de interesse e nas características dos algoritmos disponíveis.

Embora a melhor correlação linear de Pearson (Tabela 2) entre $\log P$ experimental e calculado tenha sido obtida com o algoritmo XLOGP3, isso não garantiu que este algoritmo fosse o mais adequado para a modelagem de $\log K_{oc}$. Desta forma, todos os resultados obtidos devem ser considerados para a escolha do melhor algoritmo. Assim, a escolha do algoritmo mais apropriado deve ser feita a partir dos modelos completamente testados e validados. Além disso, como foi observado neste estudo, mais de um algoritmo pode vir a ser considerado adequado para a obtenção de modelos estatisticamente razoáveis.

3.3. Comparação com modelos QSPR da literatura

Os modelos 2 e 5 foram comparados com outros modelos apresentados na literatura obtidos a partir do mesmo conjunto de dados. Os principais parâmetros estatísticos destes modelos e respectivos poderes de predição externa (conjuntos A e B) são sumarizados na Tabela 8. Observou-se que os modelos 2 e 5, apesar de apresentarem R^2 e Q^2_{LOO} menores que dos modelos da literatura, possuem poderes de predição compatíveis.

É importante notar também que os valores de SEC dos modelos 2 e 5 são maiores quando comparados aos demais. No entanto, não é apropriado o desenvolvimento de modelos de predição com erros que sejam muito menores do que o erro experimental. Assim, os modelos 2 e 5 podem ser considerados confiáveis, pois apresentaram valores de SEC que são consistentes com o erro experimental (0,44) (Lohntnger, 1994).

Tabela 8

Comparação entre modelos de $\log K_{oc}$.

Modelos	Nº de variáveis	R^2	SEC	Q^2_{LOO}	R^2_{PRED} (A)	ARE (%)	R^2_{PRED} (B)	ARE (%)
Modelo 2 – ALOGPS	1	0,744	0,447	0,735	0,668	16,1	0,768	20,5
Modelo 5 – KOWWIN	1	0,713	0,473	0,701	0,733	18,0	0,769	23,8
Gramática et al. (2000)	6	0,843	0,35	0,824	0,67	27,4*	ND	ND
Huuskonen (2003)	12	0,82	0,37	0,79	0,79	20,0*	0,74	29,6*
Duchowicz et al. (2007)	6	0,90	0,29	0,89	0,71	21,7*	ND	ND

ND = não disponível no artigo original; *calculados pelos autores deste trabalho, com base nos valores experimentais e preditos disponibilizados na literatura.

O grau de ajuste de um modelo de regressão pode ser melhorado pelo aumento do número de variáveis explicativas. Todavia, a inclusão de uma quantidade excessiva dessas variáveis pode causar o sobreajuste (aumento artificial dos valores de R^2) do modelo. Em

estudos QSPR, convencionou-se a inclusão de uma variável explicativa para cada grupo de cinco compostos do conjunto de dados (Gaudio e Zandonade, 2001). Os modelos da literatura possuem 6 ou 12 variáveis explicativas, ou seja, estão dentro dos limites convencionados. Os modelos univariados 2 e 5, apesar de apresentarem valores inferiores para R^2 e Q^2_{LOO} , apresentaram melhores valores de ARE . O uso de grande número de descritores pode levar a uma falsa impressão de que o modelo está bem ajustado (Daniel, 1995). Assim, o melhor grau de ajuste dos modelos da literatura pode ser atribuído simplesmente à existência de um maior número de variáveis explicativas. Contudo, o maior grau de ajuste não refletiu em seus respectivos poderes de predição, pois apresentaram valores próximos aos obtidos pelos modelos univariados.

Outro ponto positivo dos modelos 2 e 5, em comparação aos demais, é que os modelos da literatura foram construídos utilizando descritores moleculares de interpretação mecanística mais complexa do ponto de vista químico do que o parâmetro $\log P$. Assim, com a escolha adequada do algoritmo de $\log P$, foi possível a obtenção de modelos matematicamente mais simples, de fácil interpretação físico-química e com poderes de predição compatíveis a de modelos mais complexos. Em tal tipo de situação, configura-se a viabilidade da aplicação do princípio da *navalha de Hoccan*: se existem várias soluções possíveis para o mesmo problema, que seja adotada a explicação mais simples (Gaudio e Zandonade, 2001). Portanto, os modelos propostos neste trabalho, além de serem confiáveis em relação às suas qualidades estatísticas e poder de predição, também preencheram este requisito de simplicidade.

4. Conclusões

Os resultados deste trabalho mostraram claramente que, quando os valores experimentais de $\log P$ não são disponíveis, a escolha do algoritmo a ser usado para seu cálculo pode influenciar o resultado final de um estudo QSPR. Desta forma, esta escolha não pode ser arbitrária sob o risco de conduzir à construção de modelos inadequados para a predição da propriedade em estudo. Assim, muitos modelos apresentados na literatura que não tiveram essa preocupação podem até vir a ser melhorados, se o algoritmo de $\log P$ mais adequado for considerado.

Neste trabalho, constatou-se que os melhores modelos QSPR para predição do coeficiente de sorção no solo de pesticidas não iônicos foram obtidos com os algoritmos ALOGPS e KOWWIN. Contudo, deve-se destacar que este resultado não pode ser generalizado para qualquer estudo QSPR. Este trabalho mostrou apenas que os algoritmos aqui selecionados foram os melhores para a modelagem de $\log K_{oc}$ deste conjunto de dados. Assim, a escolha do melhor algoritmo de $\log P$ a ser usado em um modelo QSPR deve ser

feita caso a caso, levando-se em conta a propriedade que está sendo modelada, o conjunto de treinamento empregado e as características dos algoritmos disponíveis.

Por fim, os estudos QSPR devem visar à obtenção de modelos que conjuguem, dentro do possível, simplicidade matemática, facilidade de interpretação físico-química, alta qualidade estatística e efetivo poder de predição.

5. Agradecimentos

Agradecemos à Fundação Araucária e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelos recursos financeiros disponibilizados na execução deste trabalho.

Agradecemos ao Laboratório de Quimiometria Teórica e Aplicada da Universidade Estadual de Campinas pela cessão do programa QSAR Modeling.

6. Referências Bibliográficas

- Allen-King, R.M. Grathwohl, P. Ball, W.P., 2002. New modeling paradigms for the sorption of hydrophobic organic chemicals to heterogeneous carbonaceous matter in soils, sediments, and rocks. *Advances in Water Resources* 25 (8-12), 985–1016.
- Andersson, P.L., Maran, U., Fara, D., Karelson, M., Hermens, J.L.M., 2002. General and class specific models for prediction of soil sorption using various physicochemical descriptors. *Journal of Chemical Information and Computer Sciences* 42 (6), 1450–1459.
- Arias-Estévez, M., Lopez-Periago, E., Martinez-Carballo, E., Simal-Gandara, J., Mejuto, J.C., Garcia-Rio, L., 2008. The mobility and degradation of pesticides in soils and the pollution of groundwater resources. *Agriculture, Ecosystems & Environment* 123 (4), 247–260.
- ASTM, 2001. Standard Test Method 24-Hour Batch-Type Measurement Containment Sorption by Soils and Sediments. American Society for Testing and Materials, Philadelphia, PA.
- Baker, J.R., Mihelcic, J.R., Shea, E., 2000. Estimating Koc for persistent organic pollutants: limitations of correlations with Kow. *Chemosphere* 41, 813–817.

- Cheng, T., Zhao, Y., Li, X., Lin, F., Xu, Y., Zhang, X., Li, Y., Wang, R., Lai, L., 2007. Computation of octanol-water partition coefficients by guiding an additive model with knowledge. *Journal of Chemical Information and Modeling* 7 (6), 2140-2148.
- Chirico N., Gramatica P., 2011. Real external predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. *Journal of Chemical Information and Modeling* 51 (9), 2320–2335.
- Daniel, W.W., 1995. *Biostatistics: A Foundation for Analysis in the Health Sciences*. John Wiley & Sons, New York.
- Dearden, J.C., 2002. Prediction of environmental toxicity and fate using quantitative structure-activity relationships (QSARs). *Journal of the Brazilian Chemical Society* 13 (6), 754–762.
- Doucette, W.J., 2003. Quantitative structure–activity relationships for predicting soil–sediment sorption coefficients for organic chemicals. *Environmental Toxicology and Chemistry* 22 (8), 1771–1788.
- Duchowicz, P.R., González, M.P., Helguera, A.M., Cordeiro, M.N.D.S., Castro, E.A., 2007. Application of the replacement method as novel variable selection in QSPR. 2. Soil sorption coefficients. *Chemometrics and Intelligent Laboratory Systems* 88 (2), 197–203.
- Eriksson, L., Jaworska, J., Worth, A.P., Cronin, M.T.D., McDowell, R.M., Gramatica, P., 2003. Methods for reliability and uncertainty assessment and for applicability evaluations of classification and regression-based QSARs. *Environmental Health Perspectives* 111 (10), 1361–1375.
- Gaudio, A.C., Zandonade, E., 2001. Proposition, validation and analysis of qsar models. *Química Nova* 24 (5), 658–671.
- Gawlik, B.M., Sotiriou, N., Feicht, E.A., Schulte-Hostede, S., Kettrup, A., 1997. Alternatives for the determination of the soil adsorption coefficient, KOC, of non-ionic-organic compounds - a review. *Chemosphere* 34 (12), 2525–2551.

- Golbraikh, A., Shen, M., Xiao, Z., Xiao, Y., Lee, K., Tropsha, A., 2003. Rational selection of training and test set for the development of validated QSAR models. *Journal of Computer-Aided Molecular Design* 17 (2-4), 241–253.
- Golbraikh, A., Tropsha, A., 2002. Beware of q²! *Journal of Molecular Graphics and Modelling* 20 (4), 269–276.
- Goudarzi N., Goodarzi M., Araujo M.C., Galvão R.K., 2009. QSPR modeling of soil sorption coefficients (KOC) of pesticides using SPA-ANN and SPA-MLR. *Journal of Agricultural and Food Chemistry* 57 (15), 7153–7158.
- Gramatica, P., Corradi, M., Consonni, V., 2000. Modelling and prediction of soil sorption coefficients of non-ionic organic pesticides by molecular descriptors. *Chemosphere* 41 (5), 763–777.
- Gramatica, P., Giani, E., Papa, E., 2007. Statistical external validation and consensus modeling: a QSPR case study for Koc prediction. *Journal of Molecular Graphics and Modelling* 25 (6), 755–766.
- Hassett, J.J., Banwart, W.L., 1989. The sorption of non-polar organics by soils and sediments. In: Sawhney, B.L., editor. *Reactions and Movement of Organic Chemicals in Soils*. Soil Science Society of America and American Society of Agronomy (SSSA) Special Publication no. 22, Madison, WI, pp. 31–44.
- Huuskonen, J., 2003. Prediction of soil sorption coefficient of organic pesticides from the atom-type electrotopological state indices. *Environmental Toxicology and Chemistry* 22 (4), 816–820.
- Kahn, I., Fara, D., Karelson, M., Maran, U., 2005. QSPR treatment of the soil sorption coefficients of organic pollutants. *Journal of Chemical Information and Modeling* 45 (1), 94–105.
- Kiralj, R., Ferreira, M.M.C., 2009. Basic validation procedures for regression models in QSAR and QSPR studies: theory and application. *Journal of the Brazilian Chemical Society* 20 (4), 770–787.

- Lee, L.S., Rao, P.S.C., Brusseau, M.L., 1991. Nonequilibrium sorption and transport of neutral and ionized chlorophenols. *Environmental Science & Technology* 25 (4), 722–729.
- Lohninger, H., 1994. Estimation of soil partition coefficients of pesticides from their chemical structure. *Chemosphere* 29 (8), 1611–1626.
- Mackay, D., Webster, E., 2003. A perspective on environmental models and QSARs. *SAR and QSAR in Environmental Research* 14 (1), 7–16.
- Mannhold, R., van de Waterbeemd, H., 2001. Substructure and whole molecule approaches for calculating logP. *Journal of Computer-Aided Molecular Design* 15 (4), 337–354.
- Melagraki, G., Afantitis, A., Sarimveis, H., Koutentis, P.A., Markopolus, J., Igglessi-Markopoulou, O., 2007. Optimization of biaryl piperidine and 4-amino-2- biarylurea MCH1 receptor antagonists using QSAR modeling, classification techniques and virtual screening. *Journal of Computer-Aided Molecular Design* 21 (5), 251–267.
- Melo, E.B., Ferreira, M.M.C., 2009. Nonequivalent effects of diverse LogP algorithms in three QSAR studies. *QSAR & Combinatorial Science* 28 (10), 1156–1165.
- Mitra, I., Saha, A., Roy, K., 2011. Chemometric QSAR modeling and in silico design of antioxidant NO donor phenols. *Scientia Pharmaceutica* 79 (1), 31–57.
- Müller, M., 1997. Quantum chemical modelling of soil sorption coefficients: multiple linear regression models. *Chemosphere* 35 (1-2), 365–377.
- Nguyen, T.H., Goss, K.U., Ball, P.W., 2005. Polyparameter linear free energy relationships for estimating the equilibrium partition of organic compounds between water and the natural organic matter in soils and sediments. *Environmental Science & Technology* 39 (4), 913–924.
- Pimentel, D., Levitan, L., 1986. Pesticides: amounts applied and amounts reaching pests. *BioScience* 36 (2), 86–91.
- Razzaque M.M., Grathwohl P., 2008. Predicting organic carbon-water partitioning of hydrophobic organic chemicals in soils and sediments based on water solubility. *Water Research* 42 (14), 3775–3780.

- Roy, K., Mitra, I., Kar, S., Ojha, P.K., Das, R.N., Kabir, H., 2012. Comparative studies on some metrics for external validation of QSPR models. *Journal of Chemical Information and Modeling* 52 (2), 396–408.
- Sabljić, A., Güsten, H., Verhaar, H., Hermens, J., 1995. QSAR modeling of soil sorption. Improvements and systematics of log KOC vs log KOW correlations. *Chemosphere* 31 (11-12), 4489-4514.
- Tetko, I.V., Tanchuk, V.Y., Villa, A.E., 2001. Prediction of n-octanol/water partition coefficients from PHYSPROP database using artificial neural networks and E-state indices. *Journal of Chemical Information Computer Science* 41 (5), 1407-1421.
- Tetko, I.V., 2003. The WWW as a tool to obtain molecular parameters. *Mini Reviews in Medicinal Chemistry* 3 (8), 809–820.
- Todeschini, R., Consonni, V., 2009. *Molecular Descriptors for Chemoinformatics*. Wiley-VCH, Weinheim.
- Tropsha, A., Gramatica, P., Gombar, V.K., 2003. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR & Combinatorial Science* 22 (1), 69–77.
- von Oepen, B., Kördel, W., Klein, W., 1991. Sorption of non-polar and polar compounds to soil processes measurements and experience with the applicability of the modified OECD-Guideline 106. *Chemosphere* 22 (3-4), 285–304.
- Wen, Y., Su, L.M., Qin, W.C., Fu, L., He, J., Zhao, Y.H., 2012. Linear and non-linear relationships between soil sorption and hydrophobicity: model, validation and influencing factors. *Chemosphere* 86 (6), 634–640.

ARTIGO 2 - UMA ABORDAGEM ALTERNATIVA NO USO DA SOLUBILIDADE EM ÁGUA DE PESTICIDAS NÃO IÔNICOS NA MODELAGEM DO COEFICIENTE DE SORÇÃO

Versão adaptada de artigo enviado à revista *Water Research*

1. Introdução

Os pesticidas ou agrotóxicos são produtos químicos usadas na agricultura, com o propósito de se obter maior produtividade, melhor qualidade das culturas e menores custos de mão de obra e energia. Tais compostos agem direta ou indiretamente em um organismo vivo alvo, de modo a matá-lo ou interferindo em sua reprodução. Geralmente, atuam como bloqueadores rápidos e eficientes de um processo metabólico vital dos organismos para os quais são tóxicos (Bard, 2002). Entretanto, apenas uma pequena parcela do que é aplicado nas lavouras atinge suas finalidades, enquanto o restante pode causar sérios danos ao meio ambiente pela contaminação do solo, águas superficiais e subterrâneas (Pimentel e Levitan, 1986; Arias-Estévez et al., 2008).

Os pesticidas podem ser classificados de acordo com sua ação tóxica: acaricidas, bactericidas, fungicidas, herbicidas, inseticidas, nematicidas, raticidas, vermífugos, entre outros, de acordo com as pragas que controlam. Podem também ser classificados em função de sua estrutura química: orgânicos (carbamatos, clorados, fosforados e clorofosforados); inorgânicos (compostos que contém arsênio, tálio, bário, nitrogênio, fósforo, cádmio, ferro, selênio, chumbo, cobre, mercúrio ou zinco); e botânicos: (compostos extraídos de plantas) (Bard, 2002).

Nos últimos anos, vem aumentando a preocupação com o destino final e os riscos ambientais e à saúde destas substâncias por parte de agências ambientais governamentais (Mackay e Webster, 2003). Em geral, os compostos químicos orgânicos dispostos no meio ambiente sofrem modificações estruturais, tanto no número de átomos de carbono como em seus grupos funcionais originalmente presentes. Desta maneira, o conhecimento da natureza química destas moléculas é de grande valia para o pleno entendimento dos processos de contaminação. Assim, o estudo das características físicas, químicas e biológicas de cada composto torna possível a obtenção de informações sobre a reatividade, toxicidade, degradabilidade e mobilidade dessas substâncias nos diferentes compartimentos ambientais (solo, águas superficiais e subterrâneas). Portanto, pesquisas que levarem o máximo de informações sobre essas propriedades são essenciais para o planejamento de ações que objetivem a recuperação de áreas contaminadas ou a preservação de locais ainda não degradados (Jardim et al., 2009).

O coeficiente de sorção no solo normalizado para o conteúdo de carbono orgânico (K_{oc}) é um importante parâmetro físico-químico que pode ser usado para determinar o destino final e a mobilidade das substâncias químicas lançadas na natureza. Tal coeficiente

é expresso pela razão entre a concentração do componente químico sorvido pelo solo (mg.kg^{-1}) e a concentração na fase aquosa circundante (mg.L^{-1}) dividida pelo conteúdo de carbono orgânico (g de carbono orgânico por g de solo seco). Sua determinação experimental pode ser feita pelo método de equilíbrio em batelada (ASTM, 2001) ou pelo método da coluna de solo (Lee et al., 1991). Quanto menor for o valor do K_{oc} de um dado composto, maior será a sua mobilidade no solo, e maior será seu potencial para contaminar as águas subterrâneas (Dearden, 2002; Arias-Estévez et al., 2008).

A extensão da sorção de um composto químico é determinada pela sua estrutura molecular, mas depende também de fatores relacionados com as características intrínsecas do solo (Dearden, 2002). O conteúdo de carbono orgânico é o fator determinante no processo de sorção de compostos não iônicos. Desta maneira, a hidrofobicidade destes compostos é a força motriz do processo de sorção no solo (Wen et al., 2012). Enquanto para compostos não iônicos, a interação hidrofóbica é o principal mecanismo no processo de sorção, para compostos mais polares, as interações hidrofílicas tornam-se mais relevantes. Nesta situação, as interações das argilas e da sílica com os grupos funcionais polares destas substâncias passam a ser importantes (Dearden, 2002; Doucette, 2003).

A necessidade de um banco atualizado e confiável dos valores de K_{oc} tem motivado a proposição e publicação de uma grande quantidade de modelos para predição deste parâmetro (Gramatica et al., 2000). Vários estudos de relação quantitativa estrutura-propriedade (QSPR), com base na estrutura molecular dos compostos químicos, foram disponibilizados nas últimas décadas. Assim, na modelagem de K_{oc} têm sido usados diversos tipos de descritores, tais como propriedades físico-químicas, fatores cromatográficos, fatores estéricos, relações de energia de solvatação lineares (LSER), índices topológicos e parâmetros quânticos (Gawlik et al., 1997; Müller, 1997; Andersson et al., 2002; Doucette, 2003; Huuskonen, 2003; Kahn et al., 2005; Nguyen e Ball, 2005; Gramatica et al., 2007; Razzaque et al., 2008; Goudarzi et al., 2009; Wen et al., 2012). Porém, a grande maioria dos modelos QSPR propostos foi construída a partir de conjuntos pequenos de dados, ou são de classes específicas ou não foram validados adequadamente. A não validação de um modelo de predição inviabiliza a sua aplicação em virtude do desconhecimento de seu poder de predição. Portanto, para um modelo ser considerado confiável, é necessário que seja devidamente validado (Kiralj e Ferreira, 2009; Chirico e Gramatica, 2011; Chirico e Gramatica, 2012; Roy et al., 2012).

Diversos modelos têm sido propostos a partir da relação entre $\log K_{oc}$ com os logaritmos do coeficiente de partição octanol/água ($\log P$) ou da solubilidade em água ($\log S$) (Gawlik et al., 1997; Doucette, 2003; Razzaque et al., 2008). A forte correlação entre estes dois descritores ($\log P$ e $\log S$) indica que tais modelos podem ser considerados equivalentes. Além disto, esta correlação inviabiliza que estes dois descritores sejam usados conjuntamente em uma equação obtida por regressão linear múltipla. Neste caso, os

métodos de projeção multivariada, tais como a regressão por componentes principais ou regressão por mínimos quadrados parciais, são opções que podem ser consideradas (Livingstone, 2003).

Os modelos que usam apenas $\log P$, como descritor, são incapazes de diferenciar compostos que possuam $\log P$ iguais e $\log S$ diferentes. Da mesma maneira, os modelos que usam apenas $\log S$, não distinguem os compostos que possuem o mesmo $\log S$ e diferentes $\log P$. O processo de sorção de um composto químico no solo é afetado pelo processo de partição água/matéria orgânica (quanto maior a afinidade da substância pela matéria orgânica, maior será sua sorção no solo) e pela sua solubilidade em água (quanto maior a solubilidade, maior será sua lixiviação). Desta maneira, espera-se que modelos, capazes de reunir estas informações, podem gerar resultados mais realistas.

Portanto, além do uso dos dois descritores clássicos, este trabalho propõe uma abordagem alternativa com a utilização de um parâmetro de solubilidade simples, gerado pela combinação matemática entre $\log P$ e $\log S$, designado como logaritmo da solubilidade corrigida pela partição octanol/água ($\log S_p$). Assim, foram construídos modelos para $\log K_{oc}$ com base nestes três descritores ($\log P$, $\log S$ e $\log S_p$). Os modelos obtidos foram testados e validados, conforme as recomendações da literatura específica da área (Kiralj e Ferreira, 2009; Chirico e Gramatica, 2011; Chirico e Gramatica, 2012; Roy et al., 2012), para garantir que sejam confiáveis e possam ser utilizados com propósitos de predição.

2. Material e Método

2.1. Dados experimentais de K_{oc}

Os valores experimentais dos coeficientes de sorção no solo ($\log K_{oc}$) de 163 pesticidas orgânicos não iônicos foram extraídos da literatura (Sabljic et al., 1995; Gramatica et al., 2000). Tais valores experimentais estão disponíveis no Anexo 5. Os dados foram divididos em um conjunto de treinamento com 143 compostos e um conjunto de validação externa com vinte compostos (conjunto de teste). Os mesmos conjuntos foram usados nos trabalhos de Gramatica et al. (2000), Huuskonen (2003) e Duchowicz et al. (2007).

Os compostos que constituem o conjunto de treinamento são heterogêneos e pertencentes a diversas classes de pesticidas (Tabela 1), cujos valores de $\log K_{oc}$ variam entre 0,42 e 5,31. Para o conjunto de teste, os valores variam entre 0,56 e 4,50. Assim, o conjunto de treinamento foi usado para a construção dos modelos e o conjunto de teste para avaliação das capacidades de predição. O fato desses conjuntos de dados terem sido usados em outros trabalhos possibilitou a comparação dos nossos resultados com os da literatura (Gramatica et al., 2000; Huuskonen, 2003; Duchowicz et al., 2007).

Tabela 1

Classes químicas dos pesticidas usados no conjunto de treinamento.

Classes de pesticidas	n	Faixa dos valores de logK _{oc}
Acetonitrilas	6	1,57-2,86
Carbamatos	29	0,42-3,35
Dinitroanilinas	8	2,92-4,01
Organoclorados	8	3,00-5,31
Organofosfatos	28	1,20-4,66
Feniluréias	44	1,29-3,55
Triazinas	13	1,71-3,07
Di e Triazolas	7	1,83-3,73
Total	143	0,42-5,31

2.2. Descritores dos modelos QSPR

O ponto central deste trabalho consiste na proposição do uso de um parâmetro de solubilidade alternativo e simples, como descritor, na modelagem de logK_{oc}. Este parâmetro, aqui denominado como logaritmo da solubilidade corrigida pela partição octanol/água, corresponde ao logaritmo da solubilidade (mol.L⁻¹) de um dado composto químico na fase aquosa de um sistema bifásico octanol/água, obtido a partir da solução aquosa saturada desta substância.

Esta propriedade combina matematicamente as informações dos descritores logS e logP. Assim, este descritor pode ser mais indicado para a modelagem de propriedades que envolvam simultaneamente processos de solução e de partição água/matéria orgânica. Para calcular o valor deste descritor alternativo, partimos de um volume *V* de uma solução aquosa saturada de concentração *S* (mol.L⁻¹) de um dado composto químico. Se adicionarmos um volume igual de n-octanol e se mantivermos este sistema bifásico sob agitação até se estabelecer um equilíbrio entre as quantidades de soluto presentes nas fases aquosa e orgânica, teremos uma situação na qual a quantidade total de soluto *Q* (mol) será igual a soma das quantidades de soluto presentes nas duas fases. Como o sistema é bifásico e os volumes de água e n-octanol são iguais, temos que:

$$S = S_w + S_o \quad \text{Eq. 1}$$

Onde *S_w* e *S_o* correspondem às solubilidades (mol.L⁻¹) nas fases aquosa e orgânica respectivamente. Considerando que o coeficiente de partição octanol/água *P* é definido por:

$$P = S_O/S_W \quad \text{Eq. 2}$$

Combinando as duas equações anteriores temos que:

$$S = S_W + S_O = S_W + P.S_W = S_W.(1+P) \quad \text{Eq. 3}$$

Desta maneira, a solubilidade (mol.L^{-1}) do composto na fase aquosa é definida por:

$$S_W = S/(P+1) \quad \text{Eq. 4}$$

Para evitar ambigüidades com outros descritores de solubilidade já citados na literatura, renomeamos S_W como S_P e assim temos que:

$$\log S_P = \log S - \log(P+1) \quad \text{Eq. 5}$$

Portanto, os valores de $\log S_P$ são calculados a partir dos valores de $\log P$ e $\log S$ através da Equação 5. Os valores de $\log P$ e $\log S$ foram obtidos a partir da introdução das *Simplified Molecular Input Line Entry Specification* (SMILES) de cada composto no Programa ALOGPS 2.1, disponível em <http://www.vcclab.org/lab/alogps> (Tetko et al., 2001). A opção pelo uso de algoritmos para calcular os valores de $\log P$ (AlogPs) e $\log S$ (AlogpS) foi feita em virtude da ausência de dados experimentais confiáveis para todos os compostos usados neste estudo.

Outros descritores úteis para este trabalho (descritores constitucionais, topológicos, moleculares, índices de anéis e de conectividade) foram calculados com o Programa Dragon 6.0. Todos os valores dos descritores usados nas modelagens são apresentados no Anexo 5.

2.3. Construção dos modelos QSPR

Os modelos para predição de $\log K_{oc}$ foram obtidos por regressão linear utilizando os programas Minitab 15 e QSAR Modeling (*freeware* desenvolvido e cedido pelo Laboratório de Quimiometria Teórica e Aplicada da UNICAMP). Foram obtidos três modelos com uma única variável explicativa ($\log P$, $\log S$ ou $\log S_P$) e três modelos com 4 variáveis, construído a partir da adição de descritores simples aos modelos anteriores.

Para a construção destes modelos com 4 descritores, foi utilizado o método *Stepwise* para seleção de variáveis, sendo que foram consideradas apenas classes de descritores simples e de fácil interpretação físico-química (constitucionais, topológicos, moleculares, índices de anéis e de conectividade). Além disto, foram excluídos descritores que apresentassem coeficiente de correlação linear de Pearson maior que 0,6 com outra variável

selecionada. Em seguida, os modelos foram avaliados em relação às suas qualidades estatísticas e comparados entre si e com modelos da literatura.

2.4. Avaliação das qualidades estatísticas dos modelos QSPR

O grau de ajuste de um modelo é medido em termos de quanto sua especificação se adéqua aos dados experimentais. Assim, um bom modelo de regressão é aquele no qual o valor calculado pela equação é o mais próximo possível do valor observado experimentalmente. Esta avaliação é feita através dos valores calculados do coeficiente de determinação (R^2), da soma dos quadrados dos resíduos (RSS) e do desvio padrão da calibração (SEC). Para modelos QSPR serem considerados ajustados, trabalhos recentes maia rigorosos recomendam que apresentem valores de $R^2 > 0,7$ e o mais próximo possível de 1 (100%), e que os valores de RSS e de SEC sejam os mais próximos possíveis de zero (Chirico e Gramatica, 2011; Chirico e Gramatica, 2012).

A significância dos modelos foi avaliada através do teste estatístico F . Este teste permitiu verificar se a variabilidade de Y em função de X foi significativamente explicada pela regressão. O valor da estatística F obtida na regressão deve ser maior que um valor de referência tabelado. Quanto maior for o valor da estatística F obtida na regressão em comparação com o valor tabelado, mais significativo será o modelo, sendo comumente utilizado o nível de 5% (Gaudio e Zandonade, 2001; Nantasenamat et al., 2009).

A validação interna ou validação cruzada *leave-one-out* (LOO) é feita da seguinte forma: (a) exclui-se um a um cada objeto do modelo; (b) reconstrói-se o modelo sem esse objeto e calcula-se o valor do objeto excluído. Por fim, calcula-se a soma dos quadrados das diferenças entre os valores previstos e observados ($PRESS_{cv}$), o coeficiente de determinação da validação cruzada (Q^2_{LOO}) e o desvio padrão da validação cruzada ($SECV$). Para um modelo ser aprovado, $SECV$ deve ser próximo de zero, $Q^2_{LOO} > 0,6$ e $|R^2 - Q^2_{LOO}| < 0,1$ (Chirico e Gramatica, 2011; Chirico e Gramatica, 2012).

O teste de robustez ou validação cruzada *leave-N-out* (LNO) avalia se os modelos não são significativamente afetados por pequenas e deliberadas mudanças em seus parâmetros. Neste teste, o modelo é recalculado a partir da retirada sistemática de até N elementos do conjunto de treinamento. Geralmente, usa-se N máximo na faixa de 25-30% do conjunto original. Neste estudo, N variou de 1 a 40, com doze repetições para cada valor de N . Esta técnica produz uma perturbação maior que a validação LOO. Para um modelo ser considerado robusto, o valor de Q^2_{LNO} médio deve ser o mais próximo possível do valor de Q^2_{LOO} (Kiralj e Ferreira, 2009).

A ocorrência de correlação por acaso foi verificada através do teste *y-randomization*. Este teste permite avaliar se a relação entre as variáveis explicativas e a variável resposta foi decorrente de um mero acaso. Os modelos são recalculados após a randomização do

vetor Y (variável resposta) e assim espera-se uma piora significativa nos parâmetros estatísticos destas novas regressões. Este procedimento foi repetido 50 vezes e foram construídos dois gráficos, nos quais os parâmetros R^2 e Q^2_{LOO} (eixo y) foram colocados em função, no eixo x, de valores de um coeficiente de correlação $R(Y_o, Y_r)$, obtidos a partir dos valores de Y originais e randomizados. Para descartar a possibilidade de correlação por acaso, os interceptos destes gráficos devem ser menores que 0,3 para R^2 e 0,05 para Q^2_{LOO} (Eriksson et al., 2003).

A validação externa dos modelos obtidos foi feita por meio da predição dos valores de $\log K_{oc}$ para os compostos do conjunto de teste. Assim, o poder de predição foi avaliado através dos coeficientes de determinação de predição externa (R^2_{PRED}), pela análise dos coeficientes de determinação modificados da validação externa (r^2_m) e pelo método Golbraikh-Tropsha. Recomenda-se que os valores de R^2_{PRED} sejam maiores que 0,6 (Chirico e Gramatica, 2011; Chirico e Gramatica, 2012), que r^2_m médio seja maior que 0,5 e $|\Delta r^2_m|$ seja menor que 0,2 (Ojha et al., 2011). Para o método Golbraikh-Tropsha, considerando as linhas de regressão, que passam pela origem, entre os valores observados versus previstos e previstos versus observados, recomenda-se que os valores das inclinações (k e k') estejam entre 0,85 e 1,15 e que o módulo $|R^2_o - R'^2_o|$ seja menor que 0,3 (Golbraikh e Tropsha, 2002; Golbraikh et al., 2003; Tropsha et al., 2003). Também foram determinados os valores do desvio padrão da predição (SEP), da soma dos quadrados dos resíduos da predição externa ($PRESS_{ev}$) e do erro relativo médio (ARE) que devem ser os menores possíveis.

Além destes parâmetros estatísticos clássicos, adotamos também o uso do coeficiente de correlação da concordância (CCC) proposto por Lin (1989). Este parâmetro é capaz de medir a precisão e a exatidão dos valores previstos pelo modelo, dando uma medida da concordância entre os valores calculados e experimentais. Num modelo perfeito, o valor de CCC será igual a 1. Trabalhos recentes recomendam que CCC seja maior que 0,85 para a aprovação de um modelo (Chirico e Gramatica, 2011; Chirico e Gramatica, 2012).

A análise dos gráficos de resíduos foi feita com a finalidade de se verificar se os resíduos obtidos apresentam distribuição normal e média zero. As fórmulas usadas para o cálculo de todos os parâmetros estatísticos citados acima são apresentadas no Anexo 1.

Assim, neste trabalho, foram usados alguns critérios recentes de validação que são mais rigorosos do que aqueles comumente utilizados nos estudos QSPR (Kiralj e Ferreira, 2009; Chirico e Gramatica, 2011; Chirico e Gramatica, 2012; Roy et al., 2012). Tal opção foi feita a fim de garantir a qualidade estatística dos modelos obtidos com o uso do descritor $\log S_p$ e, assim, demonstrar a viabilidade do emprego desta abordagem. Após os procedimentos de validação, os melhores modelos obtidos foram comparados com outros

modelos apresentados na literatura (Gramatica et al., 2000; Huuskonen, 2003; Duchowicz et al., 2007).

3. Resultados e Discussão

3.1. Modelos para predição de $\log K_{oc}$

Os modelos para predição de $\log K_{oc}$, são apresentados na Tabela 2. Todos os modelos são lineares e seus principais parâmetros estatísticos são mostrados na Tabela 3. Considerando os modelos com uma única variável explicativa (A1, A2 e A3), verificou-se que todos foram aprovados em relação ao ajuste, pois apresentaram R^2 maior que 0,7. Entretanto, é importante observar que o modelo A3, obtido a partir do novo descritor $\log S_p$, apresentou melhor ajuste e maior significância do que os modelos A1 e A2, obtidos a partir dos descritores clássicos $\log P$ e $\log S$. A mesma tendência também foi observada para os modelos com 4 variáveis explicativas. Assim, podemos verificar que os resultados mostraram que os modelos obtidos a partir de $\log S_p$ são mais ajustados que os demais.

Tabela 2

Modelos de regressão para predição de $\log K_{oc}$ de pesticidas não iônicos (n=143).

Modelo*	Equação
A1	$\text{Log}K_{oc} = 0,945 + 0,560 \text{ Log}P$
A2	$\text{Log}K_{oc} = 0,670 - 0,568 \text{ Log}S$
A3	$\text{Log}K_{oc} = 0,702 - 0,299 \text{ Log}S_p$
B1	$\text{Log}K_{oc} = - 1,46 + 0,386 \text{ Log}P + 4,45 \text{ Mv} + 0,0152 \text{ VAR} - 0,124 \text{ MAXDP}$
B2	$\text{Log}K_{oc} = - 0,834 - 0,435 \text{ Log}S + 3,05 \text{ Mv} + 0,00909 \text{ VAR} - 0,0850 \text{ MAXDP}$
B3	$\text{Log}K_{oc} = - 1,09 - 0,227 \text{ Log}S_p + 3,48 \text{ Mv} + 0,00987 \text{ VAR} - 0,0884 \text{ MAXDP}$

* A designa modelos com uma variável; B designa modelos com 4 variáveis explicativas.

Os valores de SEC devem ser os menores possíveis, porém dentro dos limites impostos pelo erro experimental envolvido nas medidas da variável resposta. Lohninger (1994) demonstrou que há limitações na precisão das medidas experimentais de $\log K_{oc}$. Em estudo realizado com um conjunto de pesticidas, o qual possuía no mínimo quatro valores diferentes de $\log K_{oc}$ para cada composto, determinou-se que o desvio padrão médio destas medidas era igual a 0,44 unidades logarítmicas. A modelagem de uma variável dependente é limitada à qualidade dos dados experimentais disponíveis. Portanto, a obtenção de modelos de predição, cujos erros sejam muito menores que o erro experimental, não é uma situação apropriada. Assim, é importante observar que todos os modelos obtidos possuem

valores de SEC compatíveis com este erro experimental. Portanto, podem ser considerados confiáveis para a estimação de valores de $\log K_{oc}$.

Tabela 3

Parâmetros estatísticos dos modelos para predição de $\log K_{oc}$.

Modelo	R^2	SEC	RSS	F	Q^2_{LOO}	$SECV$	$PRESS_{cv}$
A1	0,744	0,447	28,14	410,04	0,735	0,451	29,15
A2	0,788	0,407	23,35	523,02	0,782	0,410	23,99
A3	0,804	0,391	21,50	580,04	0,798	0,394	22,20
B1	0,848	0,348	16,71	192,62	0,835	0,356	18,11
B2	0,821	0,378	19,72	157,93	0,805	0,387	21,44
B3	0,852	0,343	16,24	199,14	0,840	0,351	17,62

Todos os modelos foram aprovados na validação cruzada LOO, pois apresentaram Q^2_{LOO} maior que 0,6 e $|R^2 - Q^2_{LOO}| < 0,1$. Desta forma, a partir dos dados apresentados na Tabela 3, em relação ao ajuste dos dados e à validação cruzada LOO, pode-se colocar os modelos em ordem decrescente de qualidade estatística, considerando o número de descritores nas equações:

A3 > A2 > A1 (modelos com 1 descritor)

B3 > B1 > B2 (modelos com 4 descritores)

A validação cruzada LNO e o teste de correlação por acaso são outras importantes ferramentas para verificação da qualidade interna dos modelos. Os dados da validação cruzada LNO mostraram que todos os modelos apresentaram valores de Q^2_{LNO} médio próximos ao valor de seu respectivo Q^2_{LOO} (Tabela 4). Os gráficos da validação LNO para os modelos A3 e B3 (ambos obtidos a partir do descritor $\log S_p$), apresentados nas Figuras 1 e 2, mostram que os modelos são estáveis e resistem a pequenas variações. Os gráficos dos demais modelos estão disponíveis no Anexo 3 e conduzem à mesma conclusão. Entretanto, é importante observar que o modelo A3 apresentou o melhor resultado, no teste de robustez, dentre os modelos com um único descritor e, o modelo B3 foi o melhor, dentre os modelos com 4 variáveis. Portanto, também para a validação LNO, os modelos construídos a partir de $\log S_p$ foram melhores que os demais.

A correlação por acaso foi avaliada através do teste *y-randomization* e os resultados apresentados (Tabela 4) mostraram a ausência da mesma em todos os modelos obtidos, uma vez que os valores dos interceptos dos gráficos de R^2 e de Q^2_{LOO} em função de $R(Y_o, Y_f)$

estão dentro dos limites (menores que 0,3 e 0,05 respectivamente), propostos por Ericksson et al. (2003).

Tabela 4

Dados da validação cruzada LNO e teste de correlação por acaso.

Modelos	Robustez Q^2_{LNO} médio	Correlação por acaso	
		Intercepto R^2 vs $R(Y_o, Y_i)$	Intercepto Q^2_{LOO} vs $R(Y_o, Y_i)$
A1	0,734	-0,040	-0,070
A2	0,781	-0,041	-0,070
A3	0,798	-0,046	-0,077
B1	0,834	-0,023	-0,096
B2	0,803	-0,028	-0,105
B3	0,839	-0,024	-0,100

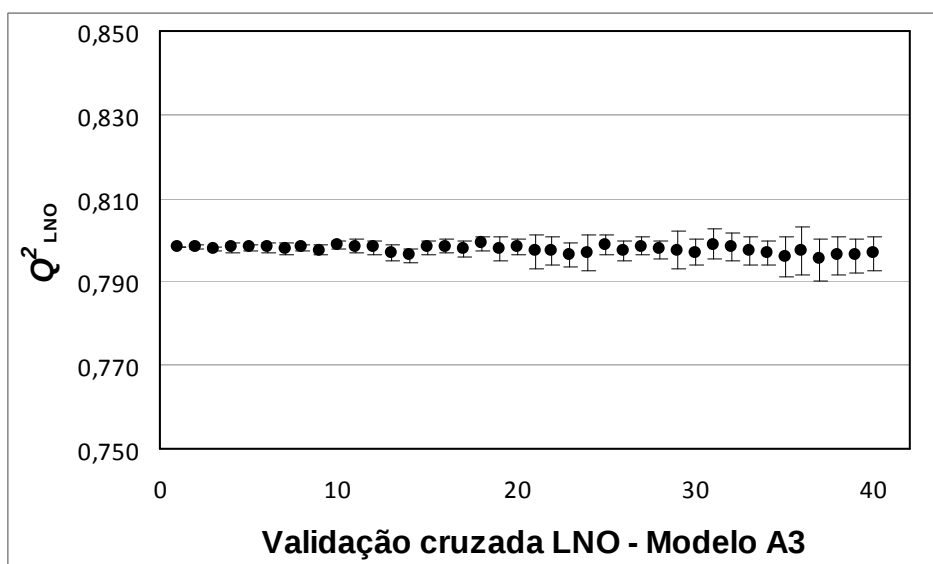


Figura 1. Gráfico da validação cruzada LNO do modelo A3.

Os resultados da validação externa (Tabela 5) mostraram que todos os modelos podem ser considerados aprovados, pois apresentaram valores de $R^2_{PRED} > 0,6$. Os valores de SEP estão compatíveis com o erro experimental associado às medidas de $\log K_{oc}$ (0,44). Desta forma, a partir dos dados da Tabela 5, os modelos podem ser colocados em ordem decrescente de poder de predição, considerando o número de descritores nas equações:

$$A3 > A2 > A1 \text{ (modelos com 1 descritor)}$$

$$B3 > B2 > B1 \text{ (modelos com 4 descritores)}$$

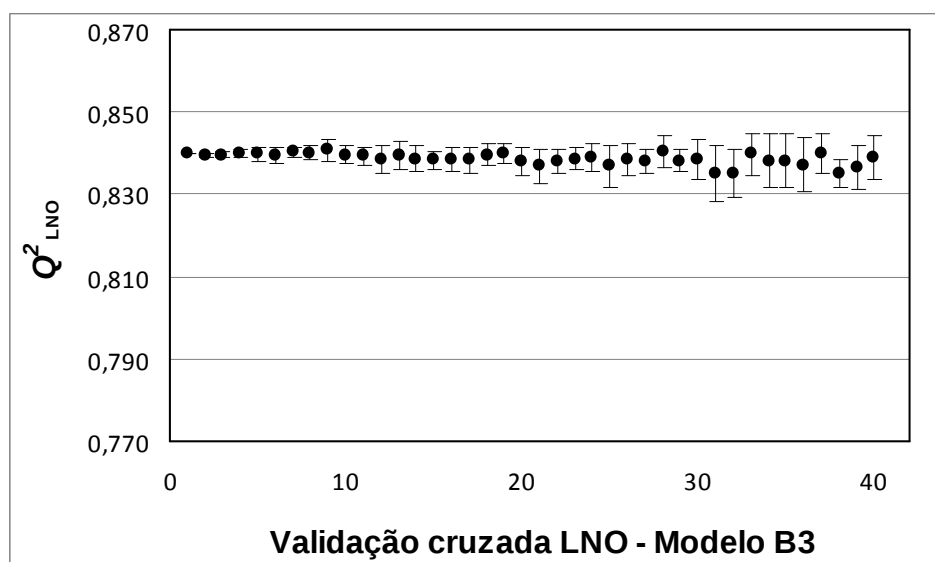


Figura 2. Gráfico da validação cruzada LNO do modelo B3.

Assim, é importante observar que, em relação ao poder de predição, o modelo A3 foi novamente o melhor dentre os modelos com um único descritor. Da mesma forma, o modelo B3 superou os demais, entre os modelos com 4 variáveis.

Os dados de CCC (Tabela 5) mostraram que o modelo A2 foi rejeitado neste critério, pois apresentou valor menor que 0,85, sendo que os demais modelos foram aprovados.

A análise dos parâmetros k , k' e $|R_o^2 - R_o'^2|$ (método Golbraikh-Tropsha) mostrou que todos os modelos estão dentro dos limites propostos (Tropsha et al.; 2003). Entretanto, é importante observar que o modelo A3 apresentou o melhor resultado, dentre os modelos com um descritor, e o modelo B3 foi novamente o melhor, dentre os modelos com 4 descritores. Esta mesma tendência também foi observada em relação aos parâmetros r_m^2 e $|\Delta r_m^2|$. Nas Figuras 3 e 4 são apresentados os gráficos dos valores observados em função dos valores previstos de $\log K_{oc}$ para os modelos A3 e B3.

Tabela 5

Dados estatísticos da validação externa (conjunto de teste, n=20).

Modelo	R^2_{PRED}	SEP	$PRESS_{ev}$	ARE (%)	CCC	r_m^2	$ \Delta r_m^2 $	k	K'	$ R_o^2 - R_o'^2 $
<i>médio</i>										
A1	0,668	0,536	5,75	16,1	0,85	0,622	0,127	0,97	0,99	0,050
A2	0,708	0,503	5,06	24,2	0,84	0,614	0,157	0,96	1,00	0,064
A3	0,723	0,490	4,80	19,9	0,86	0,654	0,004	0,96	1,00	0,001
B1	0,710	0,501	5,03	19,0	0,87	0,678	0,076	0,94	1,03	0,023
B2	0,733	0,481	4,62	23,4	0,86	0,658	0,181	0,95	1,02	0,063
B3	0,743	0,472	4,45	20,5	0,87	0,688	0,028	0,95	1,02	-0,008

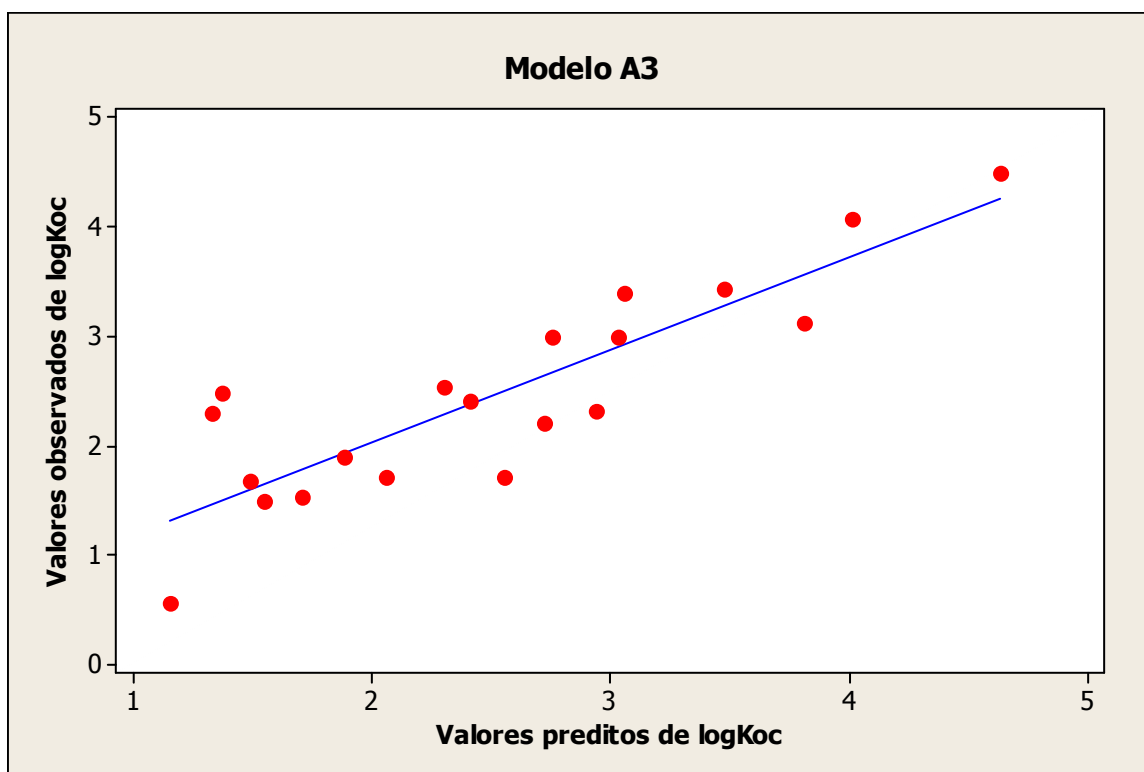


Figura 3. Gráfico dos valores observados em função dos valores previstos de $\log K_{oc}$ para o conjunto de validação externa do modelo A3.

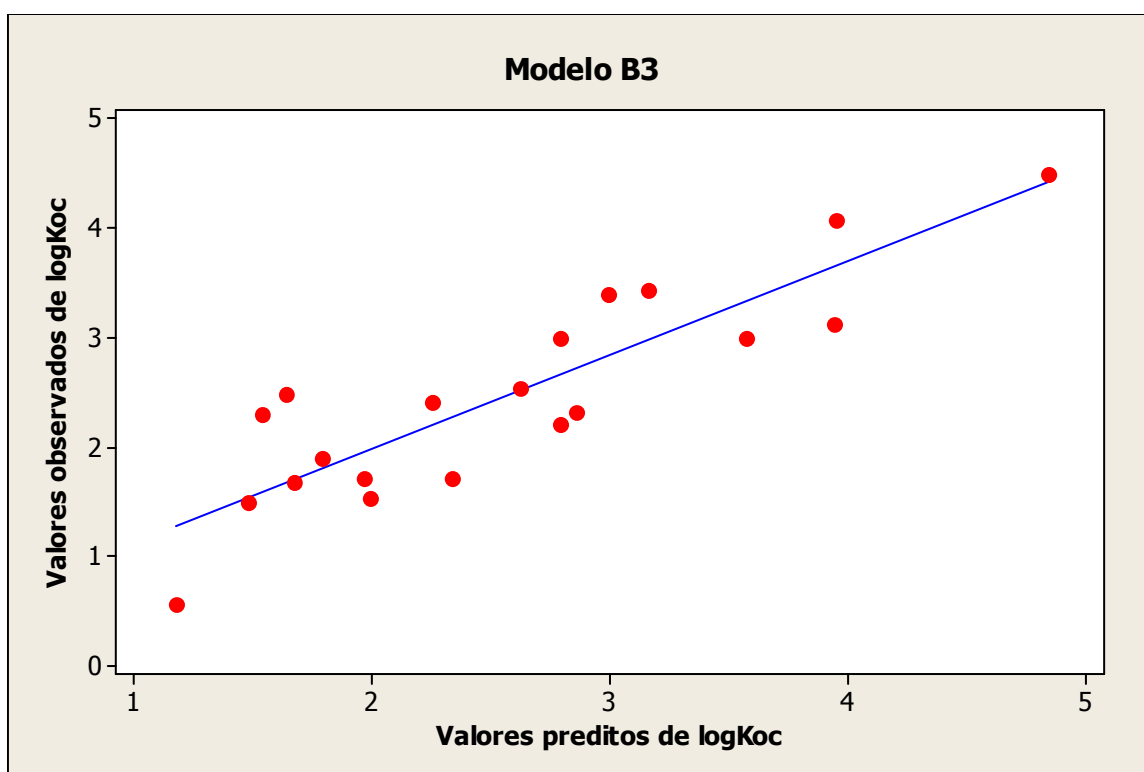


Figura 4. Gráfico dos valores observados em função dos valores previstos de $\log K_{oc}$ para o conjunto de validação externa do modelo B3.

Se os resultados forem analisados conforme critérios de r^2_m mais rigorosos (r^2_m médio $> 0,65$), os modelos A1 e A2 são rejeitados (Chirico e Gramatica, 2012). Assim, ficou evidente que o uso do descritor de solubilidade $\log S_p$ foi mais adequado do que os descritores convencionais $\log P$ e $\log S$ na modelagem de $\log K_{oc}$ de pesticidas não iônicos, com uma ou com 4 variáveis explicativas.

Quanto à análise de resíduos, os histogramas dos modelos A3 e B3, apresentados nas Figuras 5 e 6, mostram que os resíduos podem ser considerados com distribuição normal e média zero. Os histogramas dos demais modelos estão disponíveis no Anexo 3.

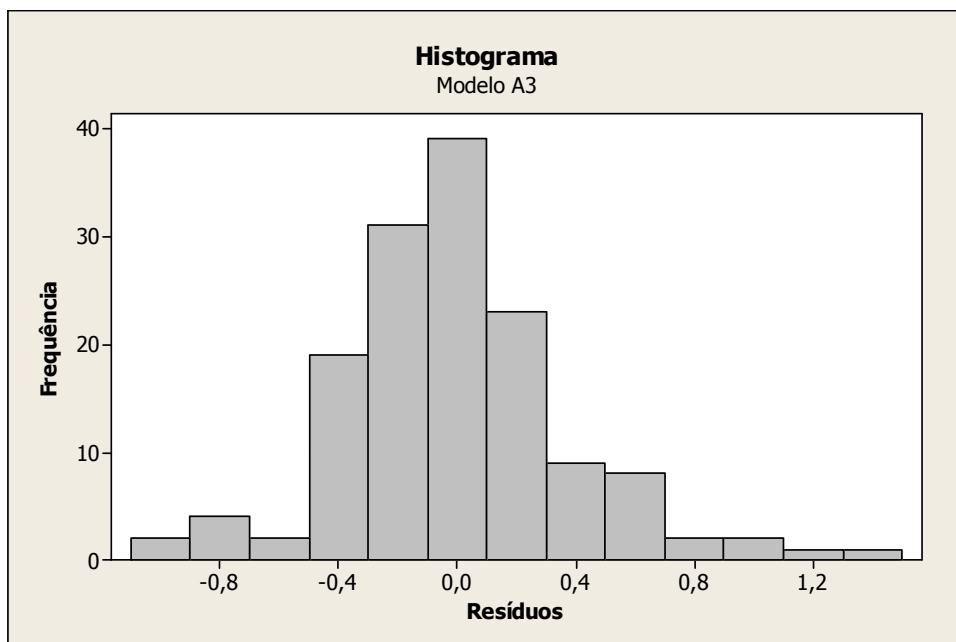


Figura 5. Histograma dos resíduos do modelo A3.

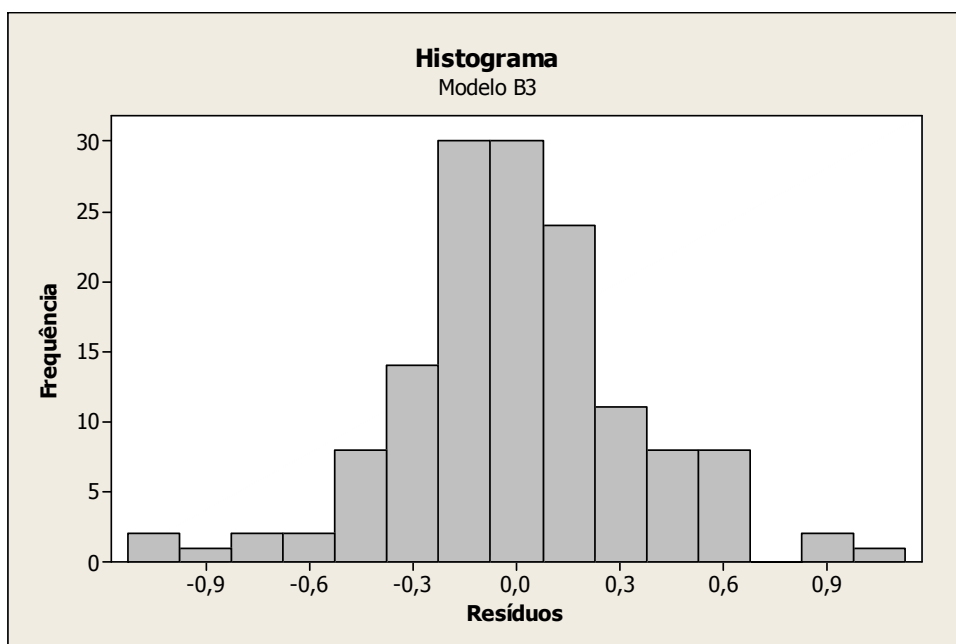


Figura 6. Histograma dos resíduos do modelo B3.

Assim, o descritor proposto $\log S_p$ mostrou-se útil na modelagem de $\log K_{oc}$, pois ao combinar as informações de $\log P$ e $\log S$, produziu modelos mais ajustados e de melhor poder de predição do que aqueles obtidos a partir dos descritores convencionais. Portanto, pode vir a ser uma alternativa a ser considerada na modelagem de propriedades que envolvam, simultaneamente, em seu mecanismo, processos de solubilidade e de partição água/matéria orgânica.

3.3. Análise e interpretação dos descritores

Os coeficientes de correlação linear de Pearson dos descritores entre si e também com a variável resposta $\log K_{oc}$ são apresentados na Tabela 6. Tais dados mostraram que a colinearidade entre as variáveis explicativas foi controlada, pois os coeficientes de correlação entre descritores são menores que 0,6. O requisito de ausência de colinearidade entre os descritores é essencial para se obter modelos estatisticamente confiáveis. Por outro lado, quanto maior for a correlação de um descritor com a variável resposta, maior será a contribuição deste parâmetro na modelagem da propriedade investigada (Gaudio e Zandonade, 2001).

A interpretação de um modelo QSPR é uma etapa importante e desejável, porém nem sempre possível. Assim, compreender como cada descritor atua sobre a variável resposta possibilita a proposição e interpretação dos mecanismos envolvidos no processo sob estudo (Melo, 2012).

Tabela 6

Coeficiente de correlação linear de Pearson entre as variáveis usadas nos modelos.

Variáveis	$\log K_{oc}$	Mv	VAR	MAXDP
$\log P$	0,863	0,300	0,522	-0,246
$\log S$	-0,888	-0,404	-0,591	0,284
$\log S_p$	-0,897	-0,361	-0,572	0,272
Mv	0,489	-	-	-
VAR	0,527	-0,055	-	-
MAXDP	-0,311	-0,206	0,222	-

A sorção de compostos orgânicos não iônicos no solo envolve um mecanismo no qual a hidrofobicidade é a força motriz do processo. Desta maneira, propriedades moleculares que estimem ou afetem a hidrofobicidade de uma substância são relevantes no processo de sorção. Assim, fica explicada a alta correlação de $\log P$, $\log S$ e $\log S_p$ com $\log K_{oc}$. Nos modelos com 4 variáveis, observou-se também a importância de outros descritores, sendo um deles constitucional (Mv) e dois topológicos (VAR e MAXDP).

O descritor constitucional Mv, volume médio de van der Waals dos átomos que compõem a molécula (escalado em função do volume do átomo de carbono), está em concordância com o mecanismo esperado de interação hidrofóbica, pois essa medida de volume pode ser associada com a capacidade de cada átomo interagir com a matéria orgânica do solo. Afinal, um átomo de maior volume teria uma maior superfície de contato.

Os descritores topológicos são variáveis que se caracterizam por descrever diferentes elementos estruturais, tais como grau de ramificação, tamanho, flexibilidade ou centricidade das moléculas (Melo, 2012). Desta maneira, o descritor VAR (variação) é um parâmetro topológico de distância e enfatiza a forte dependência entre o processo de sorção e o tamanho e simetria das moléculas.

O descritor MAXDP (diferença positiva máxima dos estados intrínsecos dos átomos) é um índice topológico proposto por Gramatica et al. (2000), deduzido a partir das eletronegatividades atômicas de Kier-Hall. Portanto, MAXDP pode ser relacionada com a eletrofilicidade da molécula e ressalta a importância de propriedades eletrônicas no sentido de dificultar o processo de sorção no solo e facilitar a lixiviação dos pesticidas.

3.3. Comparação com modelos QSPR da literatura

Os modelos A3 e B3 foram comparados com outros modelos apresentados na literatura obtidos a partir do mesmo conjunto de dados. Os principais parâmetros estatísticos destes modelos e seus respectivos poderes de predição são sumarizados na Tabela 7. Observou-se que os modelos A3 e B3 apresentaram qualidade de ajuste e capacidade de predição semelhantes a dos modelos da literatura. Todos os modelos apresentados na Tabela 7 são aprovados em relação ao grau de ajuste e possuem valores de SEC compatíveis com o erro experimental associado às medidas de $\log K_{oc}$. Entretanto, enquanto os modelos A3 e B3 são aprovados em todos os critérios de validação externa, recomendados para estudos de QSPR, os modelos de Gramatica et al. (2000) e Duchowicz et al. (2007) são rejeitados pelos seus valores de CCC (menores que 0,85). Estes modelos também são rejeitados se aplicarmos os critérios mais rigorosos de r_m^2 (valores maiores que 0,65), propostos por Chirico e Gramatica (2012). Assim, dentre os modelos da literatura, apenas o modelo de Huuskonen (2003) resistiu aos critérios mais rigorosos de validação externa atualmente recomendados.

Desta maneira, é importante observar que o modelo B3 (nosso melhor modelo) superou o modelo de Huuskonen (melhor modelo da literatura) em relação ao ajuste. Quanto aos parâmetros da validação externa, praticamente apresentaram valores próximos. Todavia, enquanto o modelo de Huuskonen foi construído com 12 variáveis explicativas, o modelo B3 é muito mais simples e foi construído apenas com quatro descritores. Ressalta-

se ainda que estas quatro variáveis estão de acordo com o mecanismo dos processos de sorção no solo de compostos não iônicos.

Tabela 7

Comparação entre modelos de $\log K_{oc}$.

Modelo	Nº de variáveis	R^2	SEC	Q^2_{LOO}	R^2_{PRED}	ARE (%)	CCC	r^2_m médio	$ \Delta r^2_m $
Modelo A3	1	0,804	0,391	0,798	0,723	19,9	0,86	0,654	0,004
Modelo B3	4	0,852	0,343	0,840	0,743	20,5	0,87	0,688	0,028
Gramatica et al. (2000)	6	0,843	0,35	0,824	0,67	27,4*	0,80*	0,539*	0,245*
Huuskonen (2003)	12	0,82	0,37	0,79	0,79	20,0*	0,87*	0,654*	0,190*
Duchowicz et al. (2007)	6	0,90	0,29	0,89	0,71	21,7*	0,83*	0,594*	0,134*

*calculados pelos autores deste trabalho, com base nos valores experimentais e preditos disponibilizados na literatura.

Nestas circunstâncias, configura-se a viabilidade da aplicação do princípio da *navalha de Occan*: se existem várias soluções possíveis para o mesmo problema, que seja adotada a explicação mais simples (Gaudio e Zandonade, 2001). Portanto, o modelo B3, proposto neste trabalho, além de ser confiável em relação às suas qualidades estatísticas e poder de predição, preenche este requisito de maior simplicidade.

4. Conclusões

Os resultados deste trabalho mostraram que o descritor de solubilidade $\log S_p$ é uma alternativa a ser considerada na modelagem de $\log K_{oc}$, uma vez que os modelos obtidos foram aprovados, sob critérios estatísticos mais rigorosos do que aqueles normalmente usados nos estudos QSPR. Além disto, todos os descritores selecionados estão consistentes com os mecanismos aceitos para o processo de sorção no solo de compostos orgânicos não iônicos.

O objetivo primordial de um estudo QSPR é a obtenção de modelos que conjuguem, dentro do possível, simplicidade matemática, facilidade de interpretação físico-química, alta qualidade estatística e efetivo poder de predição. Portanto, os modelos construídos a partir do descritor de solubilidade $\log S_p$ atingiram plenamente tais metas.

Entretanto, alguns cuidados são necessários, pois a aplicação desta abordagem não pode ser generalizada para qualquer estudo QSPR. Assim, ela deve ser testada caso a caso, desde que a propriedade a ser modelada seja dependente tanto da solubilidade em água como da partição água/matéria orgânica dos compostos sob estudo.

5. Agradecimentos

Agradecemos à Fundação Araucária e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelos recursos financeiros disponibilizados na execução deste trabalho.

Agradecemos ao Laboratório de Quimiometria Teórica e Aplicada da Universidade Estadual de Campinas pela cessão do programa QSAR Modeling.

6. Referências Bibliográficas

Andersson, P.L., Maran, U., Fara, D., Karelson, M., Hermens, J.L.M., 2002. General and class specific models for prediction of soil sorption using various physicochemical descriptors. *Journal of Chemical Information and Computer Sciences* 42 (6), 1450–1459.

Arias-Estévez, M., Lopez-Periago, E., Martinez-Carballo, E., Simal-Gandara, J., Mejuto, J.C., Garcia-Rio, L., 2008. The mobility and degradation of pesticides in soils and the pollution of groundwater resources. *Agriculture, Ecosystems & Environment* 123 (4), 247–260.

ASTM, 2001. Standard Test Method 24-Hour Batch-Type Measurement Containment Sorption by Soils and Sediments. American Society for Testing and Materials, Philadelphia, PA.

Bard, C. 2002. *Química Ambiental*. 2ª ed., Bookman, Porto Alegre.

Chirico N., Gramatica P., 2011. Real external predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. *Journal of Chemical Information and Modeling* 51 (9), 2320–2335.

Chirico N., Gramatica P., 2012. Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection. *Journal of Chemical Information and Modeling* 52 (8), 2044–2058.

- Dearden, J.C., 2002. Prediction of environmental toxicity and fate using quantitative structure-activity relationships (QSARs). *Journal of the Brazilian Chemical Society* 13 (6), 754–762.
- Doucette, W.J., 2003. Quantitative structure–activity relationships for predicting soil-sediment sorption coefficients for organic chemicals. *Environmental Toxicology and Chemistry* 22 (8), 1771–1788.
- Duchowicz, P.R., González, M.P., Helguera, A.M., Cordeiro, M.N.D.S., Castro, E.A., 2007. Application of the replacement method as novel variable selection in QSPR. 2. Soil sorption coefficients. *Chemometrics and Intelligent Laboratory Systems* 88 (2), 197–203.
- Eriksson, L., Jaworska, J., Worth, A.P., Cronin, M.T.D., McDowell, R.M., Gramatica, P., 2003. Methods for reliability and uncertainty assessment and for applicability evaluations of classification and regression-based QSARs. *Environmental Health Perspectives* 111 (10), 1361–1375.
- Gaudio, A.C., Zandonade, E., 2001. Proposition, validation and analysis of qsar models. *Química Nova* 24 (5), 658–671.
- Gawlik, B.M., Sotiriou, N., Feicht, E.A., Schulte-Hostede, S., Kettrup, A., 1997. Alternatives for the determination of the soil adsorption coefficient, KOC, of non-ionic-organic compounds - a review. *Chemosphere* 34 (12), 2525–2551.
- Golbraikh, A., Shen, M., Xiao, Z., Xiao, Y., Lee, K., Tropsha, A., 2003. Rational selection of training and test set for the development of validated QSAR models. *Journal of Computer-Aided Molecular Design* 17 (2-4), 241–253.
- Golbraikh, A., Tropsha, A., 2002. Beware of q²! *Journal of Molecular Graphics and Modelling* 20 (4), 269–276.
- Goudarzi N., Goodarzi M., Araujo M.C., Galvão R.K., 2009. QSPR modeling of soil sorption coefficients (KOC) of pesticides using SPA-ANN and SPA-MLR. *Journal of Agricultural and Food Chemistry* 57 (15), 7153–7158.

- Gramatica, P., Corradi, M., Consonni, V., 2000. Modelling and prediction of soil sorption coefficients of non-ionic organic pesticides by molecular descriptors. *Chemosphere* 41 (5), 763–777.
- Gramatica, P., Giani, E., Papa, E., 2007. Statistical external validation and consensus modeling: a QSPR case study for Koc prediction. *Journal of Molecular Graphics and Modelling* 25 (6), 755–766.
- Huuskonen, J., 2003. Prediction of soil sorption coefficient of organic pesticides from the atom-type electrotopological state indices. *Environmental Toxicology and Chemistry* 22 (4), 816–820.
- Jardim, I.C.S.F., Andrade, J.A., Queiroz, S.C.N., 2009. Residues of pesticides in food: a global environmental preoccupation - Focusing on apples. *Química Nova* 32 (4), 998–1012.
- Kahn, I., Fara, D., Karelson, M., Maran, U., 2005. QSPR treatment of the soil sorption coefficients of organic pollutants. *Journal of Chemical Information and Modeling* 45 (1), 94–105.
- Kiralj, R., Ferreira, M.M.C., 2009. Basic validation procedures for regression models in QSAR and QSPR studies: theory and application. *Journal of the Brazilian Chemical Society* 20 (4), 770–787.
- Lee, L.S., Rao, P.S.C., Brusseau, M.L., 1991. Nonequilibrium sorption and transport of neutral and ionized chlorophenols. *Environmental Science & Technology* 25 (4), 722–729.
- Lin, L., 1989: A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 45 (1), 255-268.
- Livingstone, D.J., 2003. Quantitative structure-activity relationships. In: *Medicinal Chemistry: Principles and Practice*. Org. King, F.D. 2nd ed. Cambridge: Royal Society of Chemistry.
- Lohninger, H., 1994. Estimation of soil partition coefficients of pesticides from their chemical structure. *Chemosphere* 29 (8), 1611–1626.

- Mackay, D., Webster, E., 2003. A perspective on environmental models and QSARs. *SAR and QSAR in Environmental Research* 14 (1), 7–16.
- Melo, E.B., 2012. A new quantitative structure-property relationship model to predict bioconcentration factors of polychlorinated biphenyls (PCBs) in fishes using E-state index and topological descriptors. *Ecotoxicology and Environmental Safety* 75 (1), 213-222.
- Müller, M., 1997. Quantum chemical modelling of soil sorption coefficients: multiple linear regression models. *Chemosphere* 35 (1-2), 365–377.
- Nantasenamat, C., Isarankura-Na-Ayudhya, C., Naenna, T., Prachayasittikul, V., 2009. A practical overview of quantitative structure-activity relationship. *EXCLI Journal* 8, 74–88.
- Nguyen, T.H., Goss, K.U., Ball, P.W., 2005. Polyparameter linear free energy relationships for estimating the equilibrium partition of organic compounds between water and the natural organic matter in soils and sediments. *Environmental Science & Technology* 39 (4), 913–924.
- Ojha, P.K., Mitra, I., Das, R.N., Roy, K., 2011. Further exploring rm2 metrics for validation of QSPR models. *Chemometrics and Intelligent Laboratory Systems* 107 (1), 194–205.
- Pimentel, D., Levitan, L., 1986. Pesticides: amounts applied and amounts reaching pests. *BioScience* 36 (2), 86–91.
- Razzaque M.M., Grathwohl P., 2008. Predicting organic carbon-water partitioning of hydrophobic organic chemicals in soils and sediments based on water solubility. *Water Research* 42 (14), 3775–3780.
- Roy, K., Mitra, I., Kar, S., Ojha, P.K., Das, R.N., Kabir, H., 2012. Comparative studies on some metrics for external validation of QSPR models. *Journal of Chemical Information and Modeling* 52 (2), 396–408.
- Sabljić, A., Güsten, H., Verhaar, H., Hermens, J., 1995. QSAR modeling of soil sorption. Improvements and systematic of log KOC vs log KOW correlations. *Chemosphere* 31 (11-12), 4489–4514.

- Tetko, I.V., Tanchuk, V.Y., Villa, A.E., 2001. Prediction of n-octanol/water partition coefficients from PHYSPROP database using artificial neural networks and E-state indices. *Journal of Chemical Information Computer Science* 41 (5), 1407-1421.
- Tropsha, A., Gramatica, P., Gombar, V.K., 2003. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR & Combinatorial Science* 22 (1), 69–77.
- Wen, Y., Su, L.M., Qin, W.C., Fu, L., He, J., Zhao, Y.H., 2012. Linear and non-linear relationships between soil sorption and hydrophobicity: model, validation and influencing factors. *Chemosphere* 86 (6), 634–640.

ANEXO 1 – Fórmulas dos parâmetros estatísticos

R²: coeficiente de determinação

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_{ci})^2}{\sum_i (y_i - \bar{y})^2}$$

RSS: soma dos quadrados dos resíduos

$$RSS = \sum_i (y_i - \hat{y}_{ci})^2$$

SEC: desvio padrão da calibração

$$SEC = \sqrt{\frac{\sum_i (y_i - \hat{y}_{ci})^2}{n - p - 1}}$$

Teste estatístico F

$$F_{(p, n-p-1)} = \frac{(n - p - 1) \cdot R^2}{p \cdot (1 - R^2)}$$

Q²_{LOO}: coeficiente de determinação da validação cruzada LOO

$$Q^2_{LOO} = 1 - \frac{\sum_i (y_i - \hat{y}_{vi})^2}{\sum_i (y_i - \bar{y})^2}$$

PRESS_{cv}: soma dos quadrados dos resíduos da validação cruzada

$$PRESS_{cv} = \sum_i (y_i - \hat{y}_{vi})^2$$

SECV: desvio padrão da validação cruzada

$$SECV = \sqrt{\frac{\sum_i (y_i - \hat{y}_{vi})^2}{n}}$$

R²_{PRED}: coeficiente de determinação da validação externa

$$R^2_{PRED} = 1 - \frac{\sum_i (y_i - \hat{y}_{ei})^2}{\sum_i (y_i - \bar{y})^2}$$

PRESS_{ev}: soma dos quadrados dos resíduos da validação externa

$$PRESS_{ev} = \sum_i (y_i - \hat{y}_{ei})^2$$

SEP: desvio padrão da predição

$$SEP = \sqrt{\frac{\sum_i (y_i - \hat{y}_{ei})^2}{n_{ev}}}$$

ARE: erro relativo médio

$$ARE = \frac{\sum_i \frac{|y_i - \hat{y}_{ei}|}{y_i}}{n} \times 100$$

k e k': inclinações das linhas de regressão

$$k = \frac{\sum_i y_i \hat{y}_{ei}}{\sum_i \hat{y}_{ei}^2}; \quad k' = \frac{\sum_i y_i \hat{y}_{ei}}{\sum_i y_i^2}$$

r²_m: coeficiente de determinação modificado

$$r_m^2 = r^2 \times \left(1 - \sqrt{r^2 - r_0^2}\right)$$

CCC: coeficiente de correlação da concordância

$$CCC = \frac{2 \sum_{i=1}^{n_{EXT}} (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sum_{i=1}^{n_{EXT}} (y_i - \bar{y})^2 + \sum_{i=1}^{n_{EXT}} (\hat{y}_i - \bar{\hat{y}})^2 + n_{EXT} (\bar{y} - \bar{\hat{y}})^2}$$

y_i : valor observado de $\log K_{oc}$; $\bar{y}$: valor observado médio de $\log K_{oc}$ para o conjunto de treinamento; $\hat{y}_{ci}$: valor estimado de $\log K_{oc}$ no modelo de calibração; $\hat{y}_{vi}$: valor estimado de $\log K_{oc}$ na validação cruzada; n : número de amostras; p : número de variáveis no modelo; $\hat{y}_{ei}$: valor estimado de $\log K_{oc}$ na validação externa; n_{ev} : número de amostras no conjunto de teste; r^2 : coeficiente de determinação do gráfico entre valores observados e previstos; r_0^2 : coeficiente de determinação do gráfico entre valores observados e previstos sem intercepto.

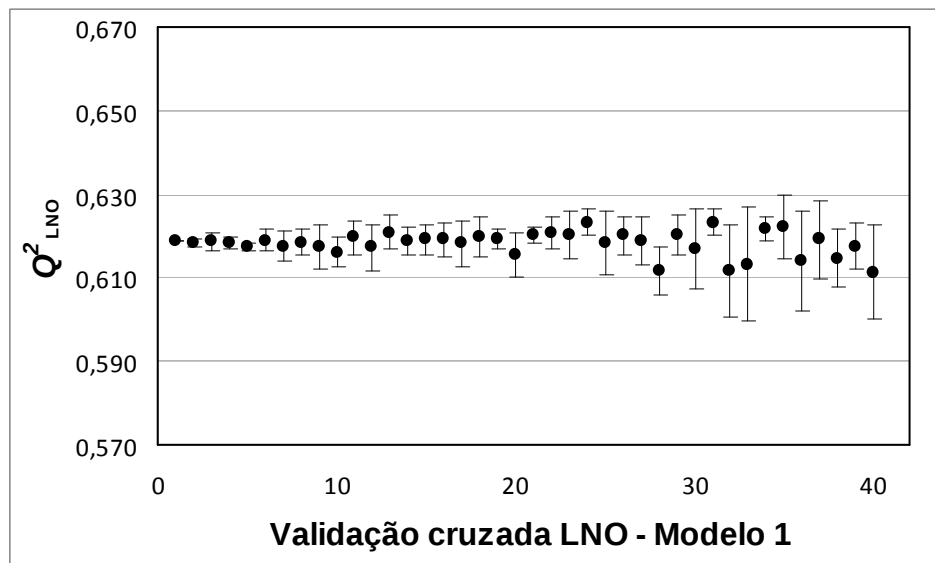
ANEXO 2 - Figuras extras do artigo 1

Figura A2.1. Gráfico da validação cruzada LNO do modelo 1 (ALOGP).

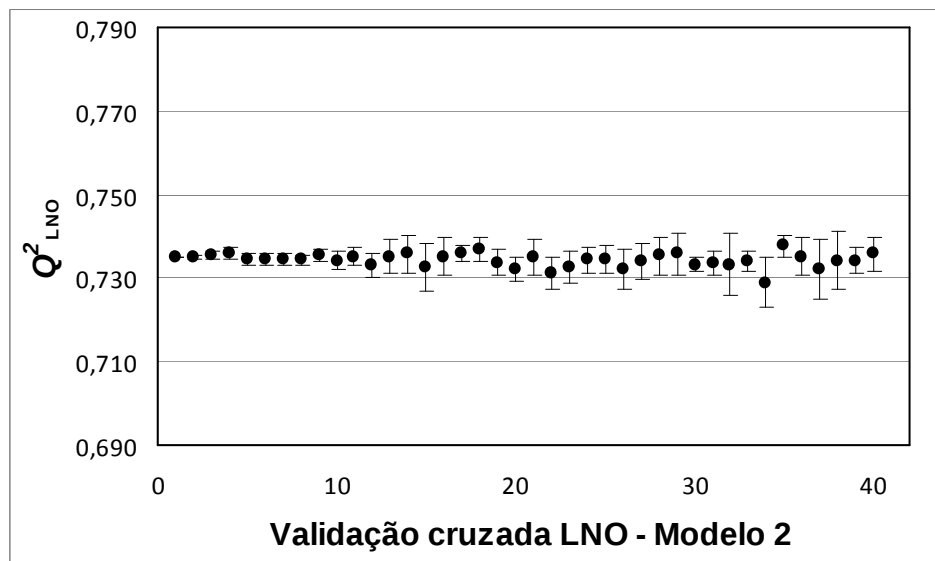


Figura A2.2. Gráfico da validação cruzada LNO do modelo 2 (ALOGPS).

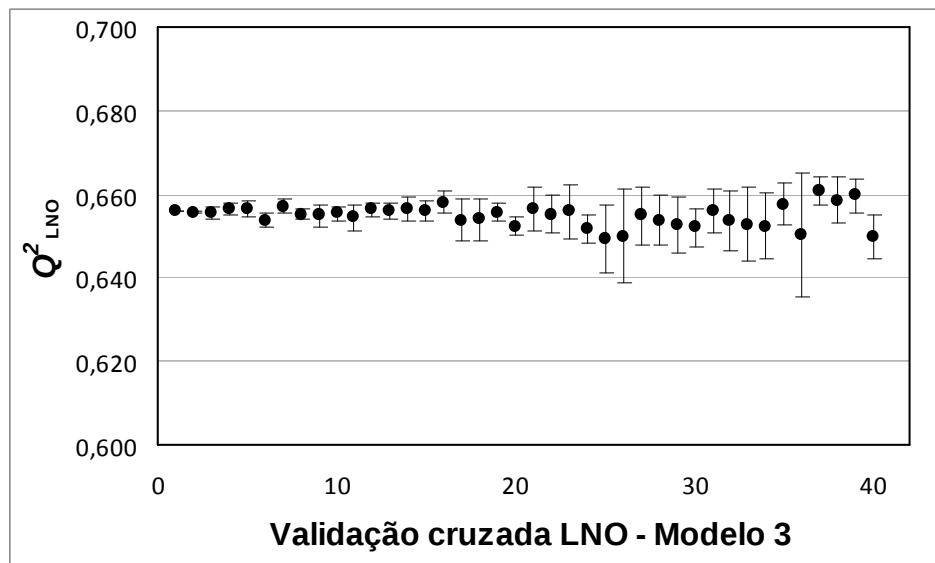


Figura A2.3. Gráfico da validação cruzada LNO do modelo 3 (ACLOGP).

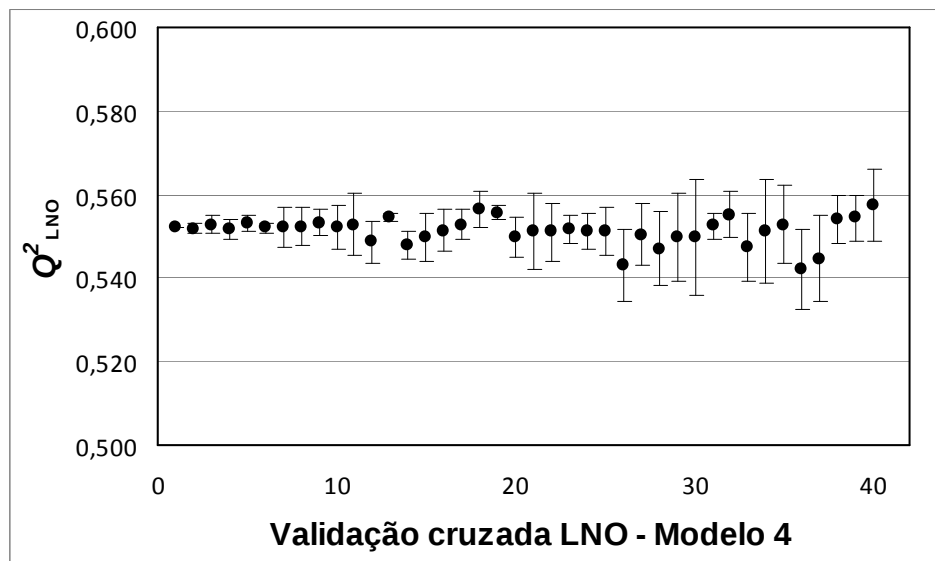


Figura A2.4. Gráfico da validação cruzada LNO do modelo 4 (MLOGP).

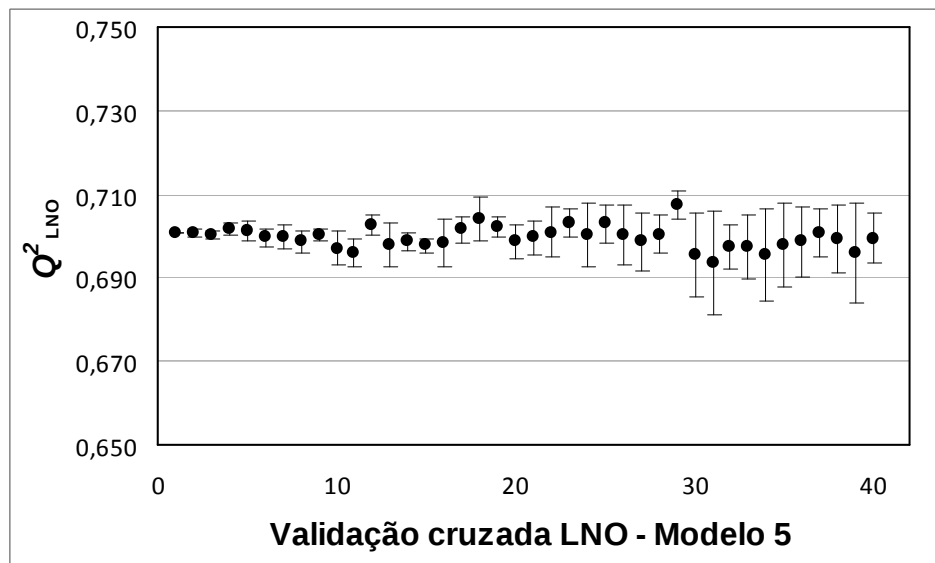


Figura A2.5. Gráfico da validação cruzada LNO do modelo 5 (KOWWIN).

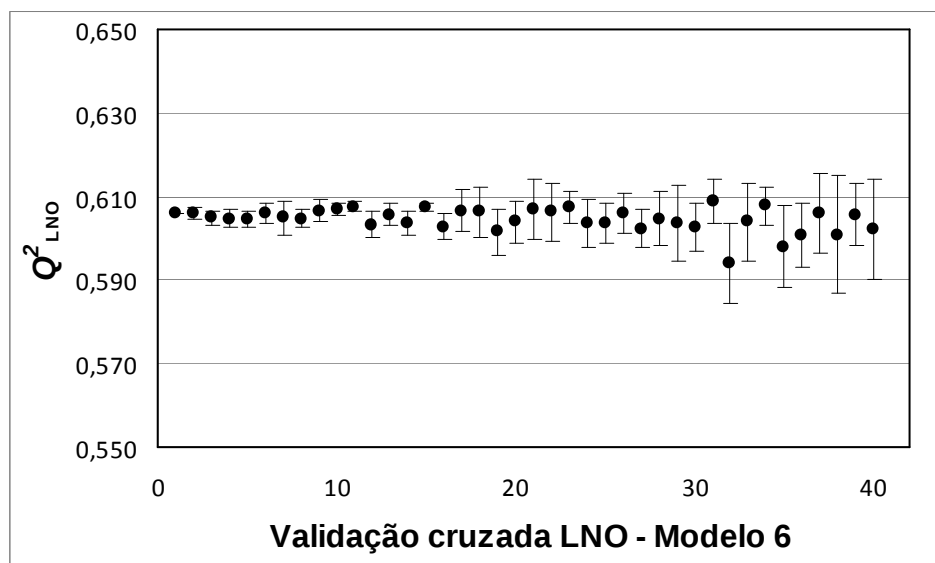


Figura A2.6. Gráfico da validação cruzada LNO do modelo 6 (XLOGP2).

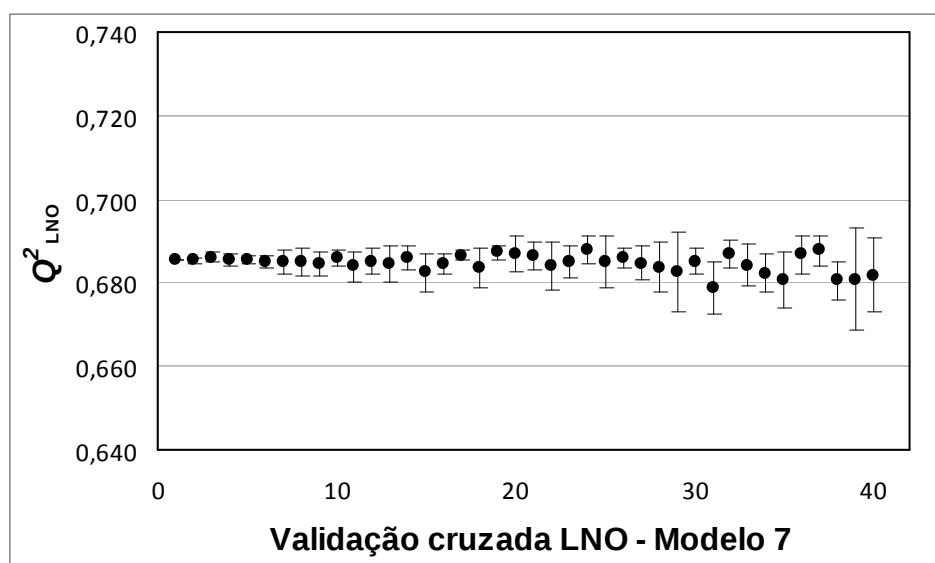


Figura A2.7. Gráfico da validação cruzada LNO do modelo 7 (XLOGP3).

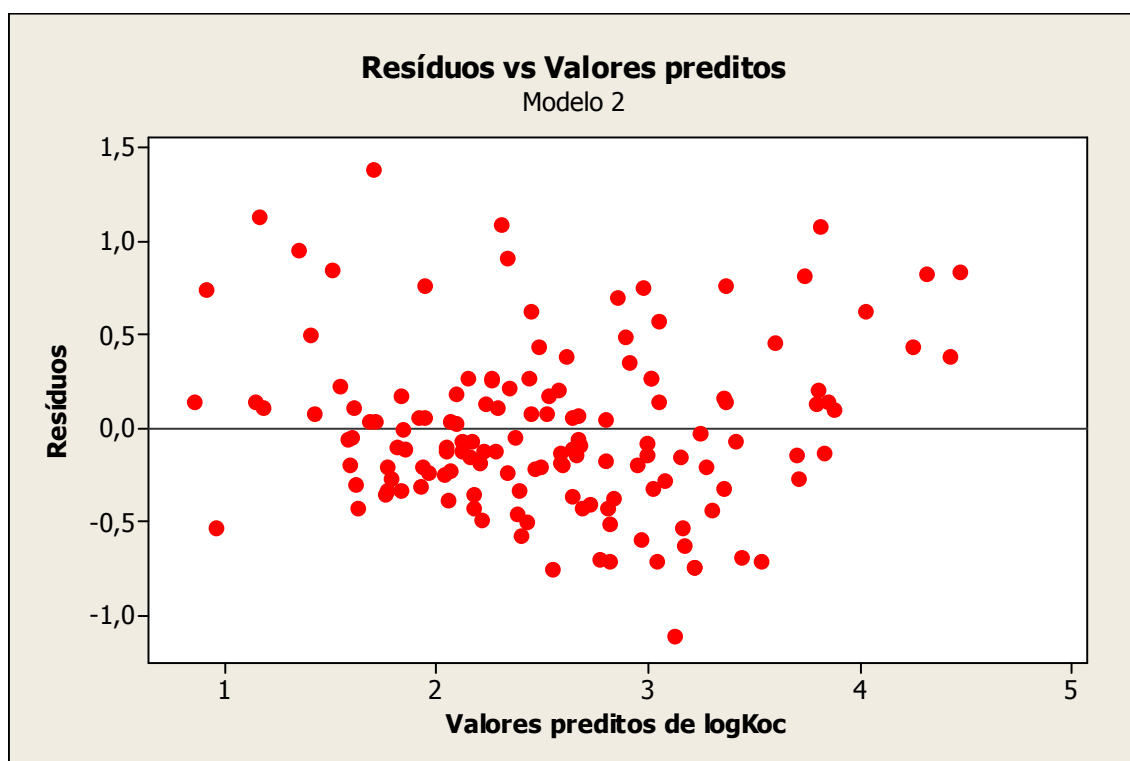


Figura A2.8. Gráfico dos resíduos versus valores ajustados do modelo 2 (ALOGPS).

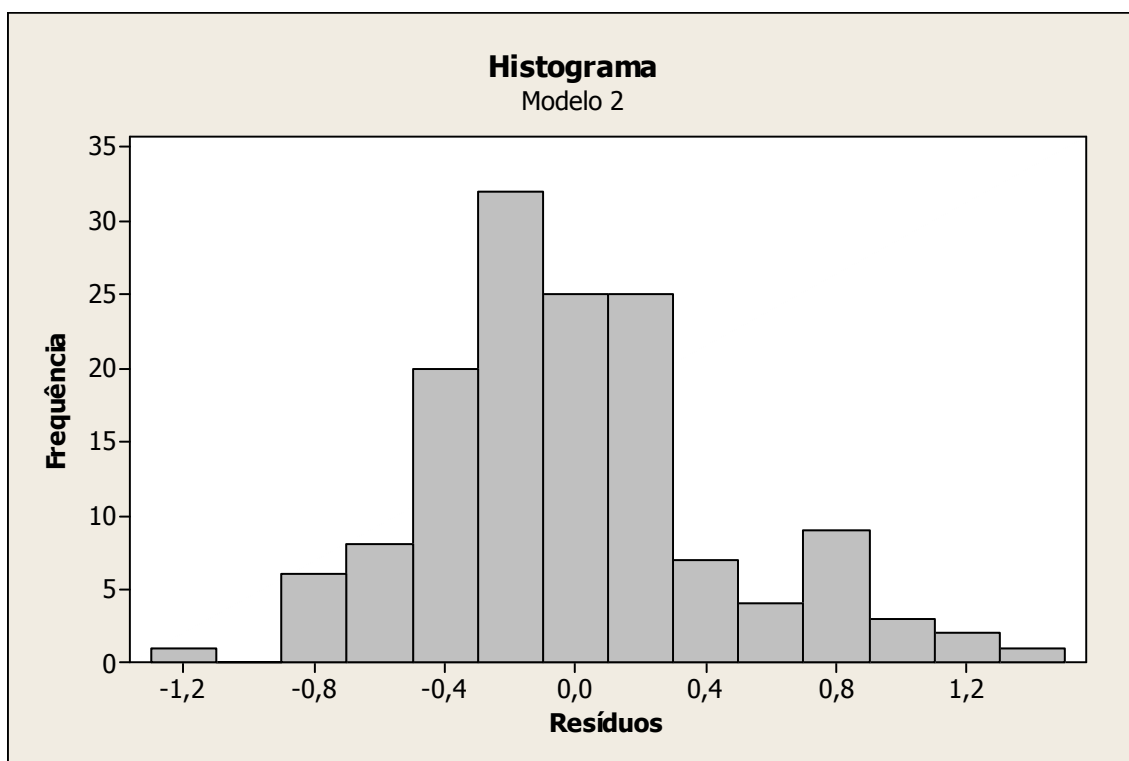


Figure A2.9. Histograma da distribuição dos resíduos do modelo 2 (ALOGPS).

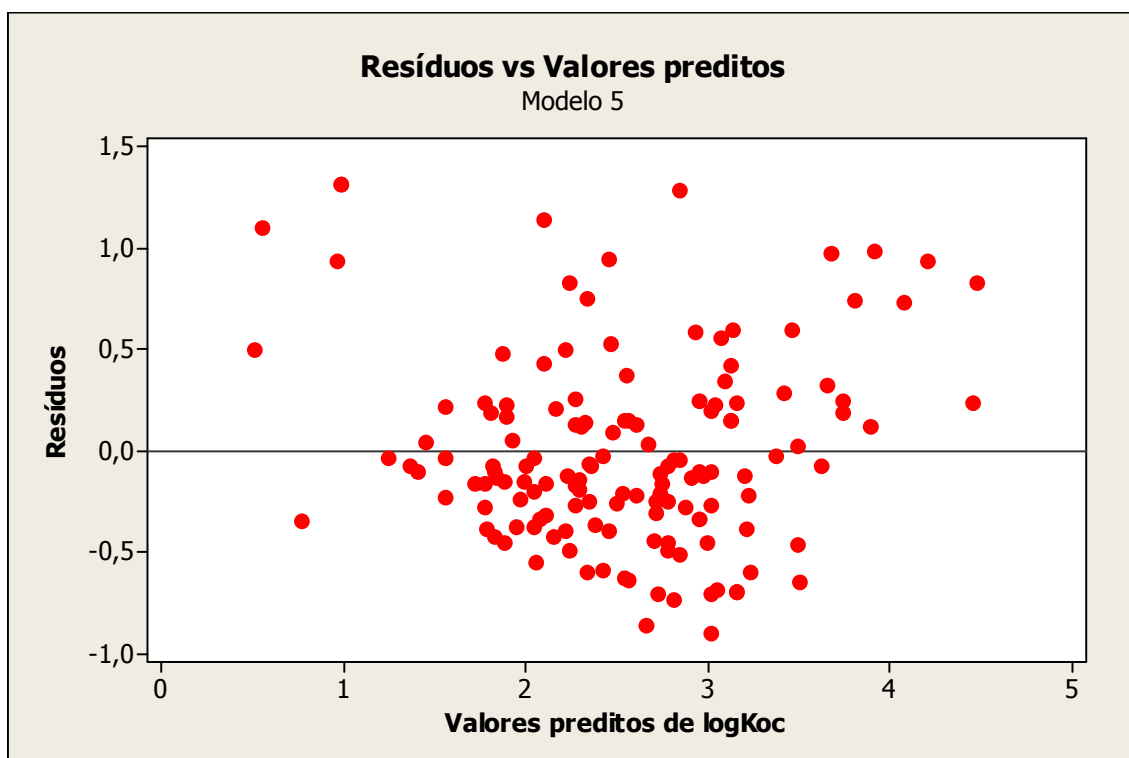


Figura A2.10. Gráfico dos resíduos versus valores ajustados do modelo 5 (KOWWIN).

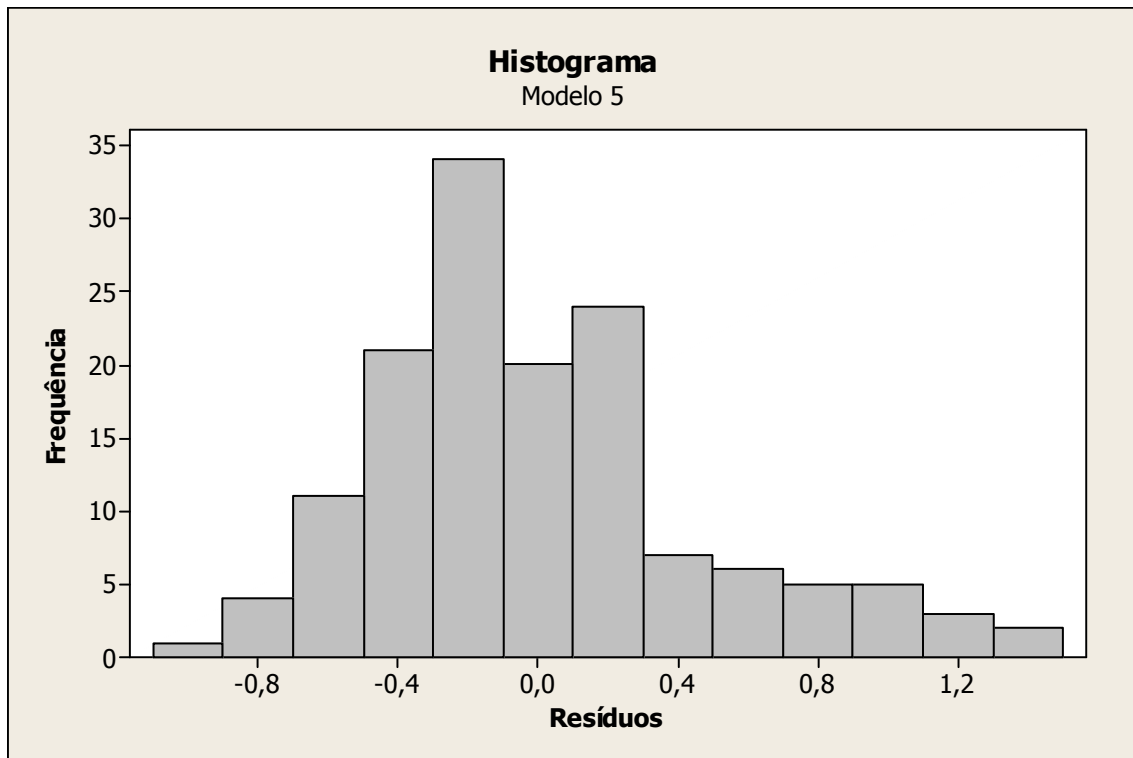


Figure A2.11. Histograma da distribuição dos resíduos do modelo 5 (KOWWIN).

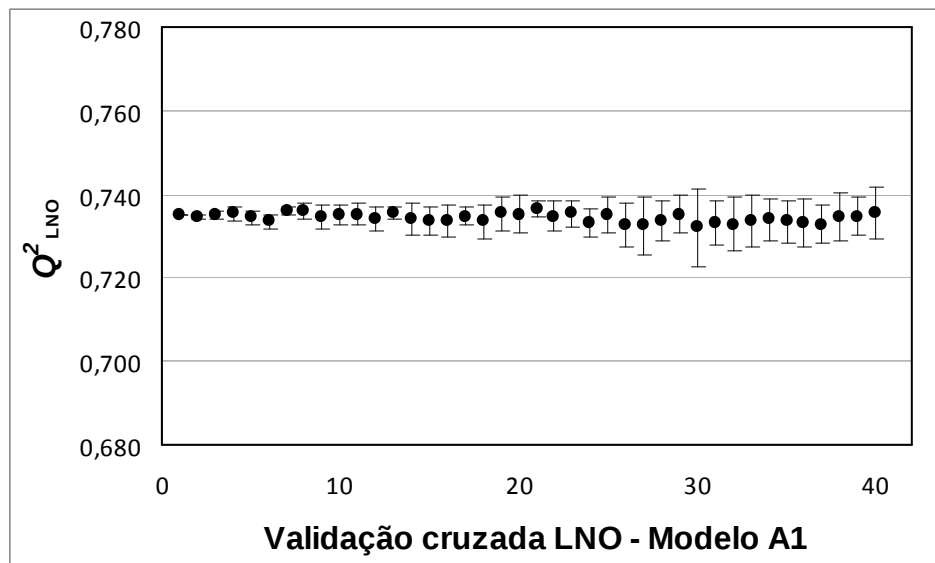
ANEXO 3 - Figuras extras do artigo 2

Figura A3.1. Gráfico da validação cruzada LNO do modelo A1.

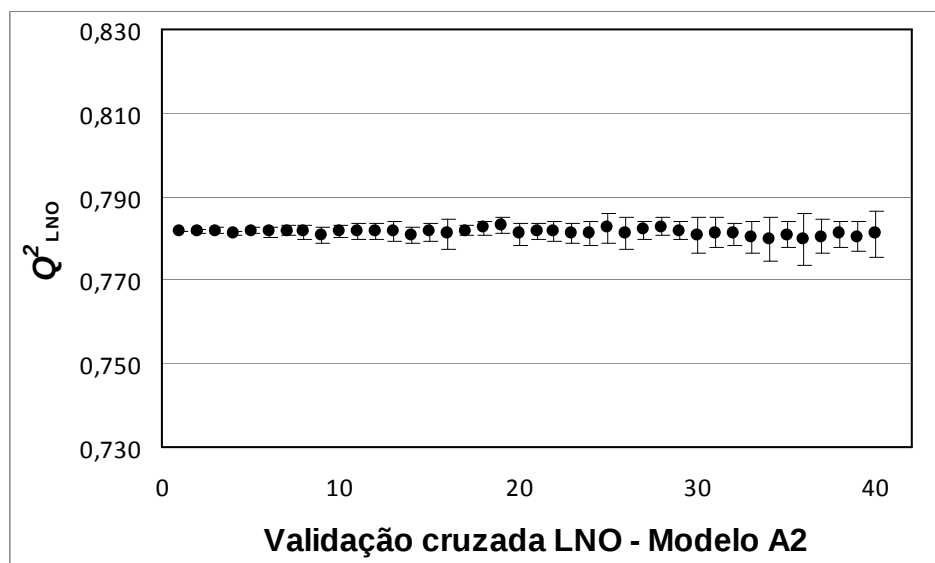


Figura A3.2. Gráfico da validação cruzada LNO do modelo A2.

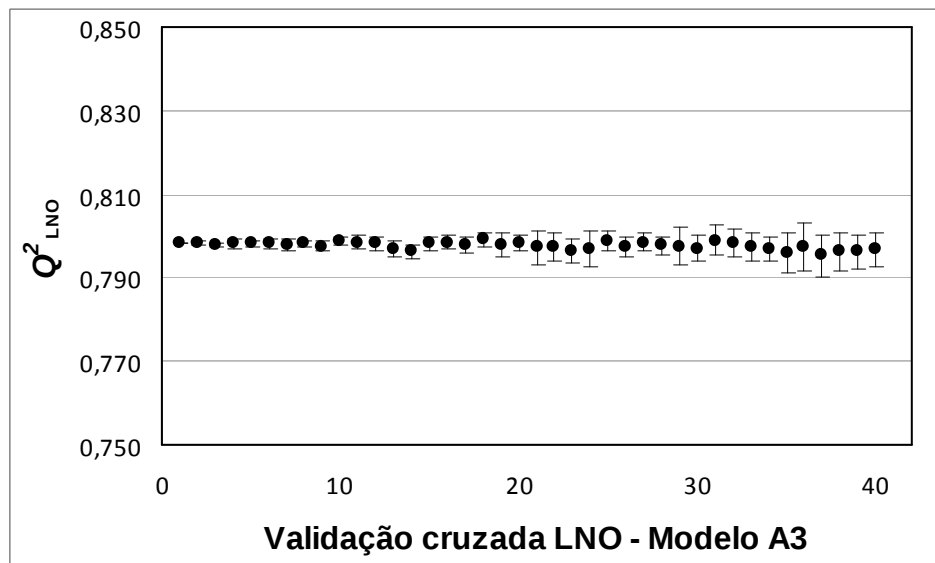


Figura A3.3. Gráfico da validação cruzada LNO do modelo A3.

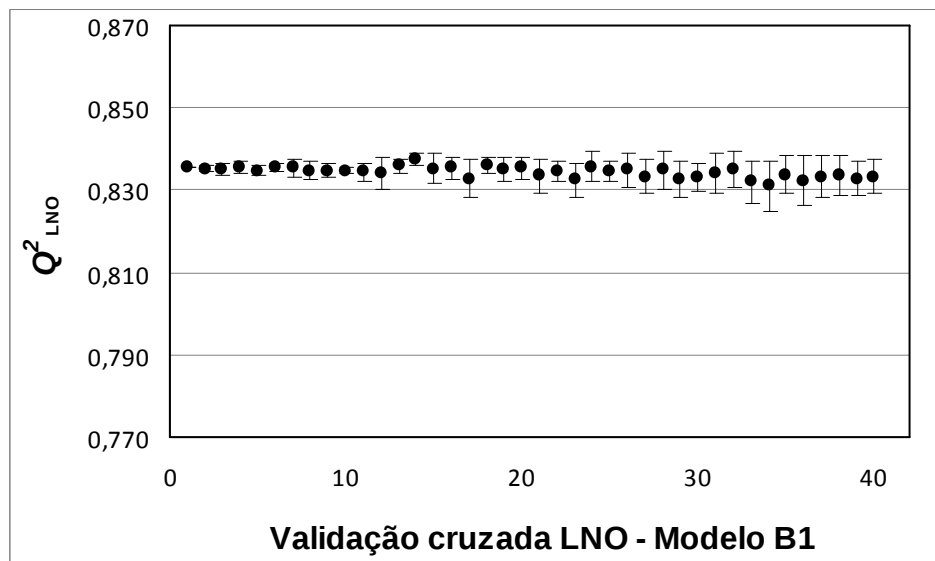


Figura A3.4. Gráfico da validação cruzada LNO do modelo B1.

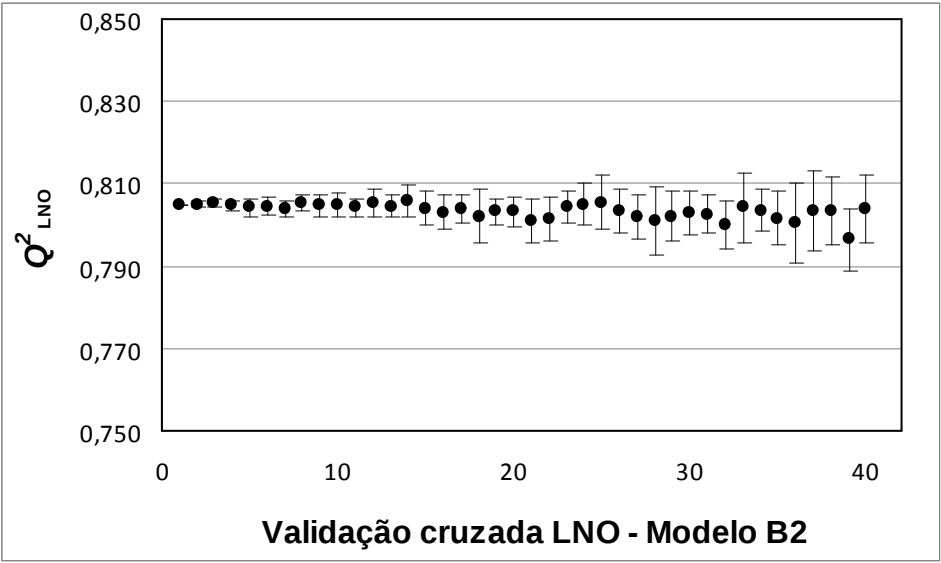


Figura A3.5. Gráfico da validação cruzada LNO do modelo B2.

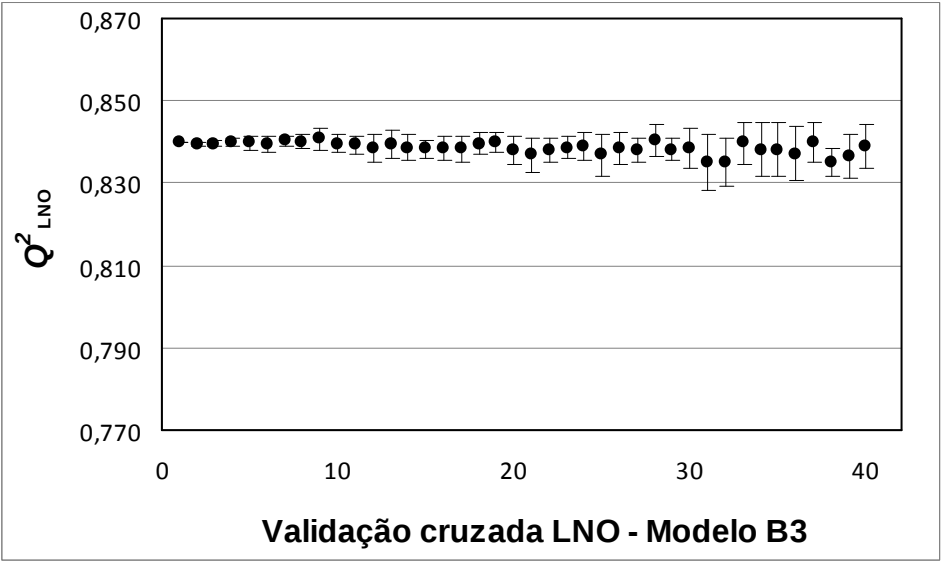


Figura A3.6. Gráfico da validação cruzada LNO do modelo B3.

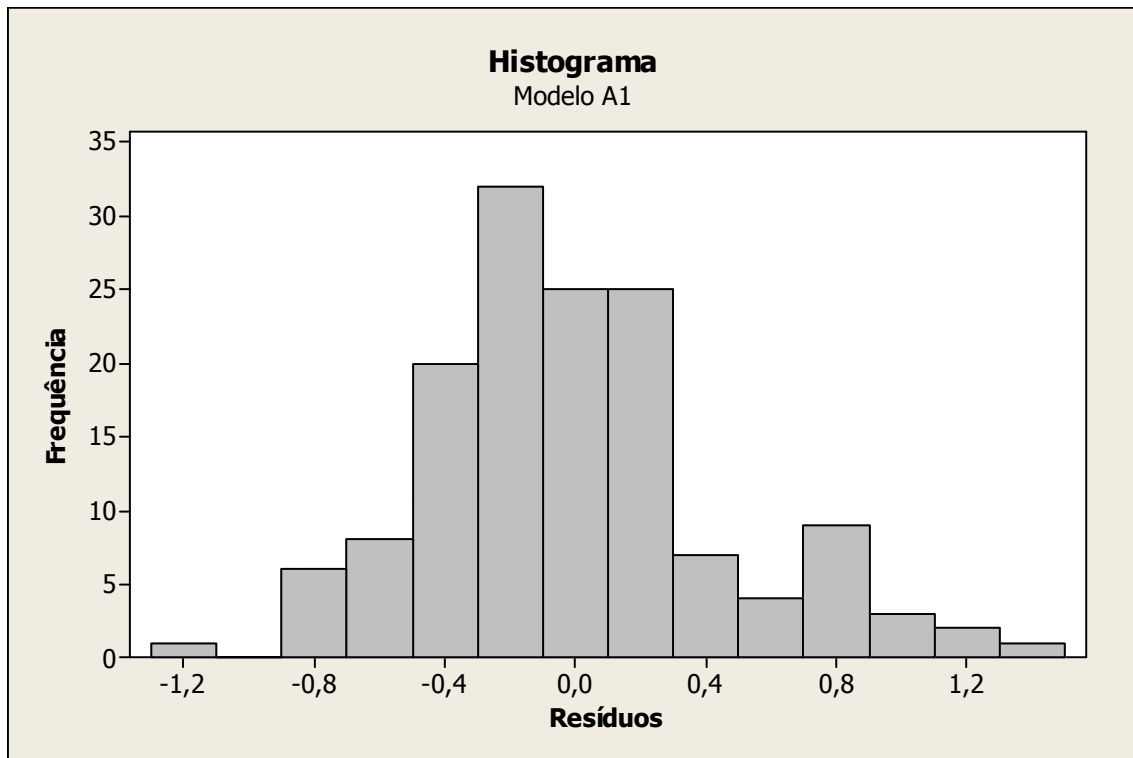


Figure A3.7. Histograma da distribuição dos resíduos do modelo A1.

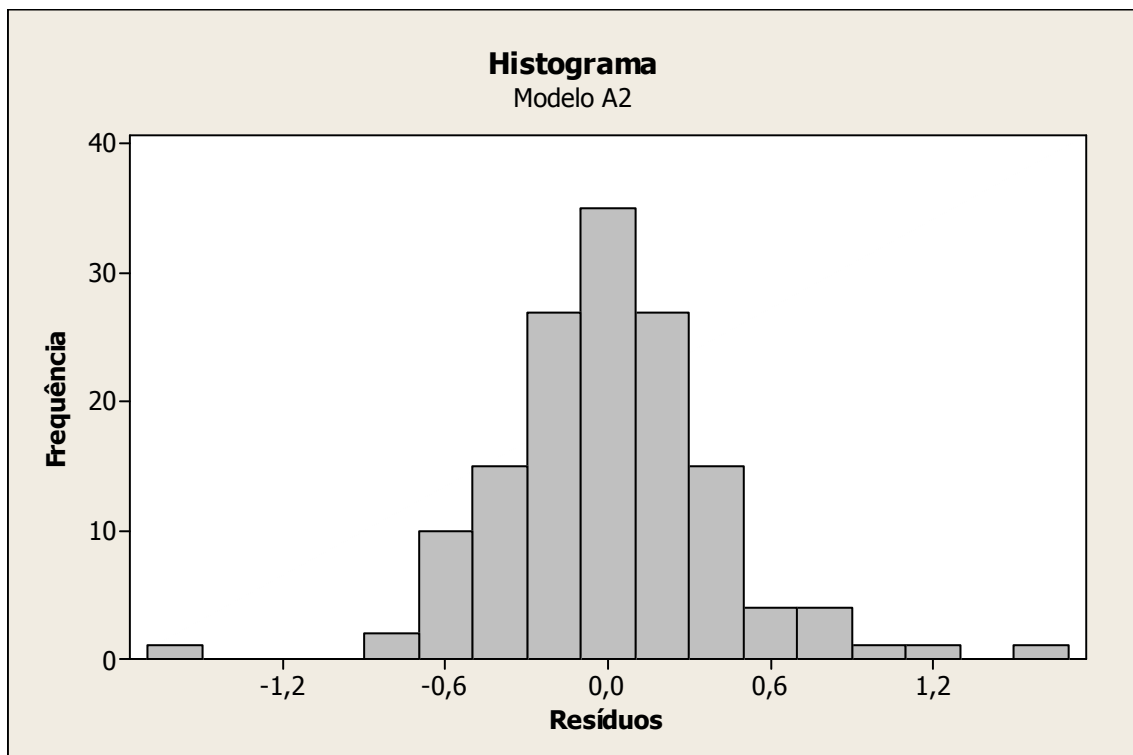


Figure A3.8. Histograma da distribuição dos resíduos do modelo A2.

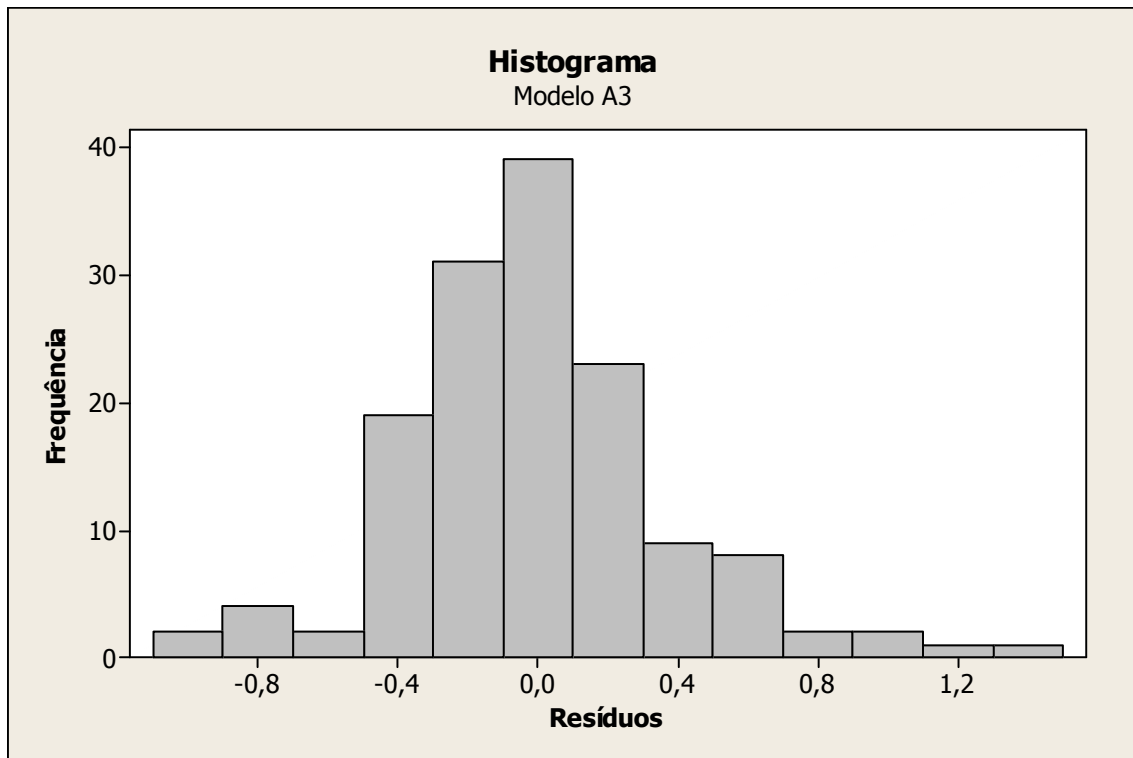


Figure A3.9. Histograma da distribuição dos resíduos do modelo A3.

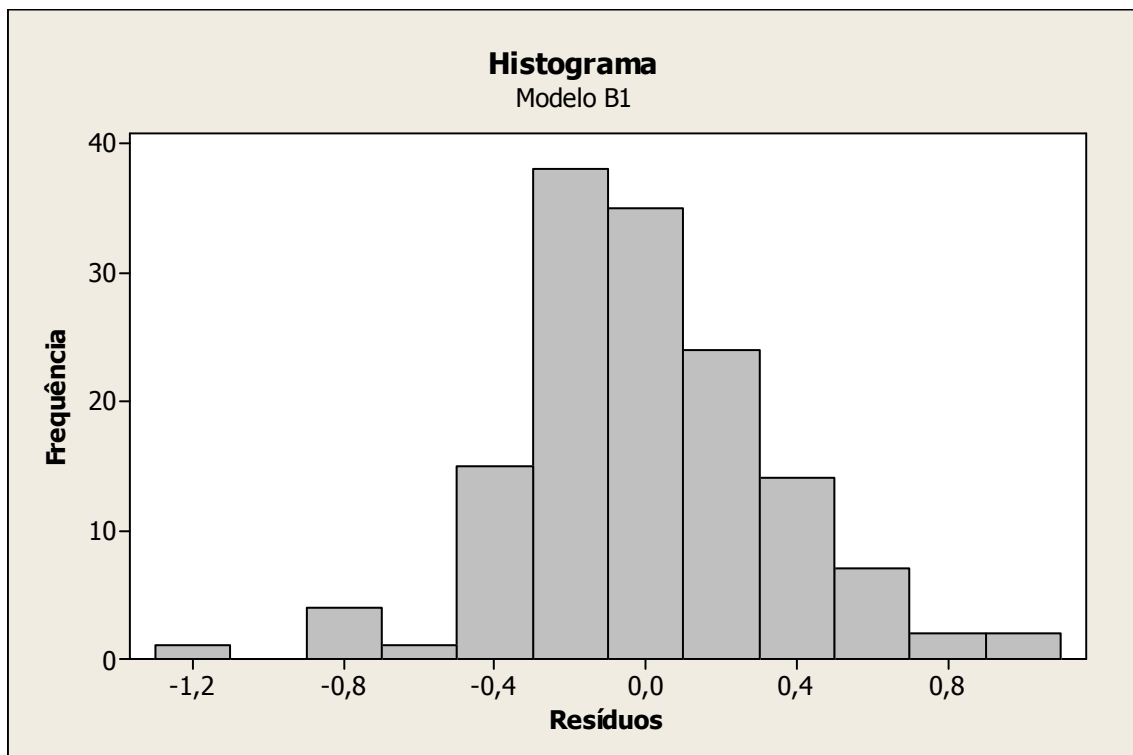


Figure A3.10. Histograma da distribuição dos resíduos do modelo B1.

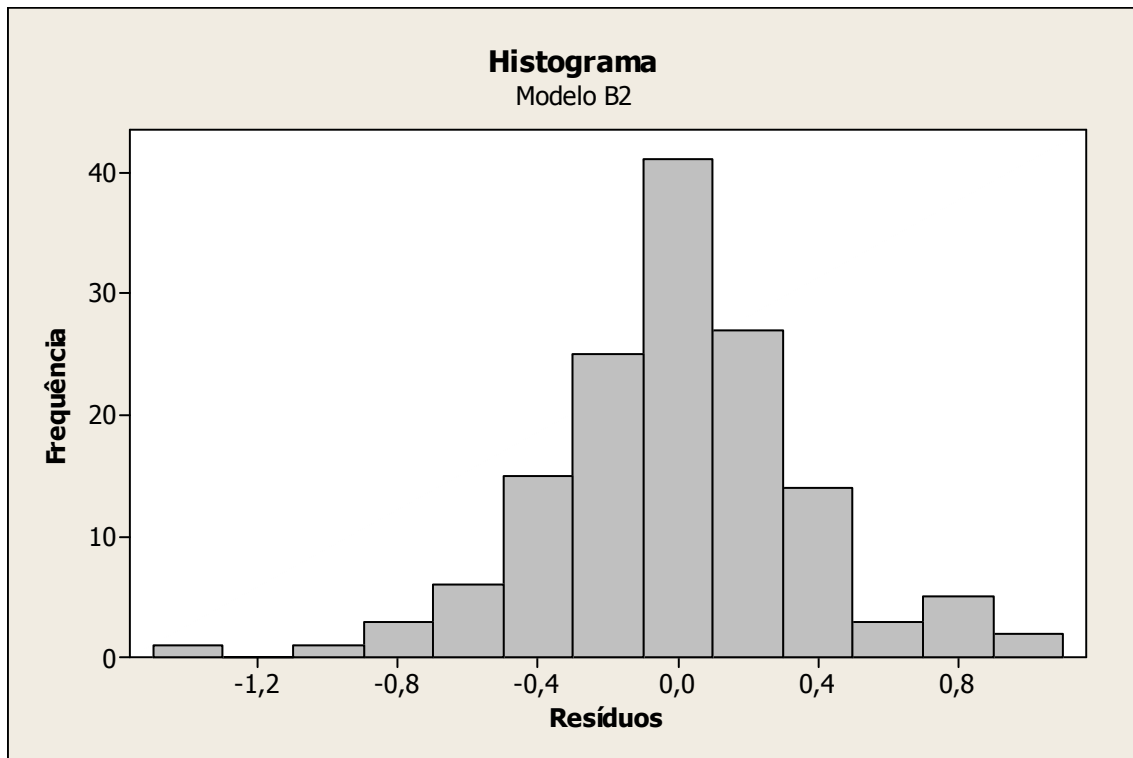


Figure A3.11. Histograma da distribuição dos resíduos do modelo B2.

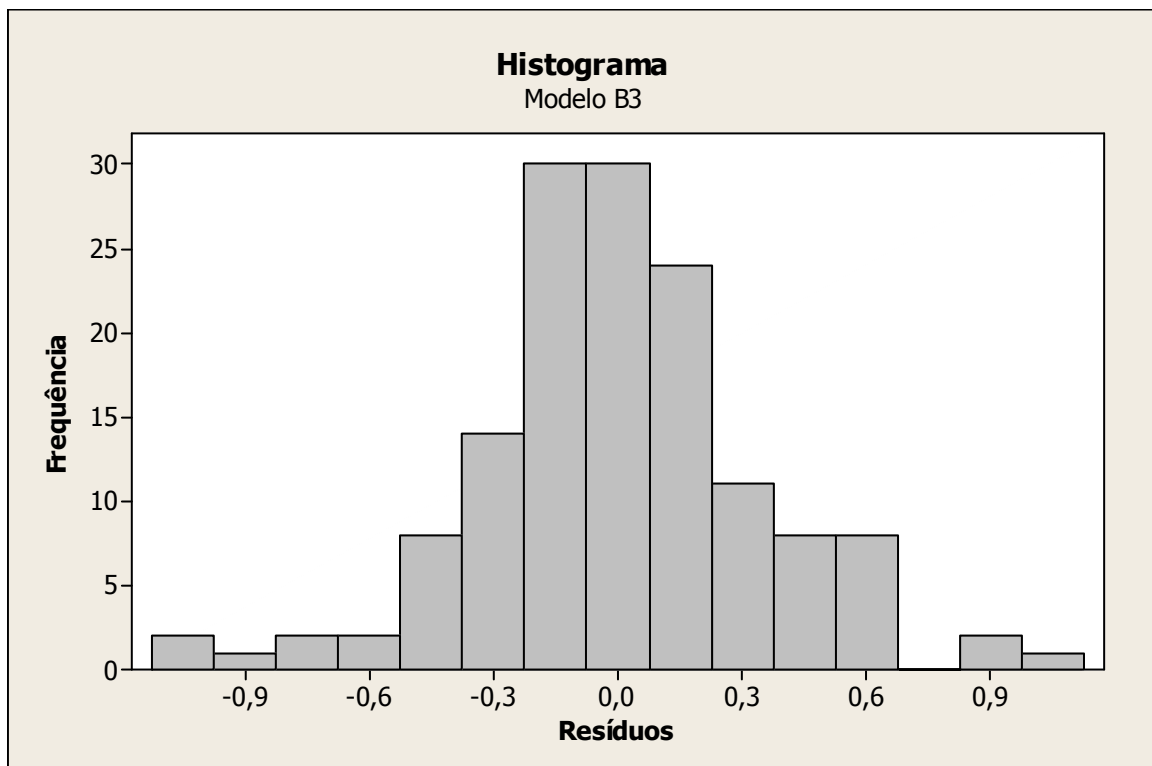


Figure A3.12. Histograma da distribuição dos resíduos do modelo B3.

ANEXO 4 - Tabela dos valores de $\log K_{oc}$ e $\log P$ de todos os pesticidas usados no artigo 1

N	Compostos do conjunto de treinamento	$\log K_{oc}$	$\log P_{exp}$	ALOGP	ALOGPS	ACLOGP	MLOGP	KOWWIN	XLOGP2	XLOGP3
1	Acetochlor	2,32	3,03	3,53	3,17	3,43	3,18	3,37	2,94	3,18
2	Alachlor	2,28	3,52	3,64	3,02	3,36	3,18	3,37	2,98	3,52
3	Butachlor	2,86	4,50	4,97	4,20	4,72	3,91	4,84	4,33	4,50
4	Metaxyl	1,57	1,65	2,38	1,47	2,08	1,91	1,70	1,58	1,65
5	Metolachlor	2,46	3,13	3,58	3,37	3,52	3,03	3,24	2,84	3,13
6	Propachlor	2,42	2,18	2,63	2,15	3,03	2,82	2,42	2,54	2,18
7	Aldicarb	1,50	1,13	1,17	1,58	2,44	0,79	1,36	1,68	1,13
8	Aldicar sulfone (= aldoxycarb)	0,42	-0,57	0,15	0,01	0,79	-0,15	-0,67	0,46	-0,57
9	Benomyl	2,71	2,12	2,96	1,78	3,29	1,66	2,24	2,83	2,12
10	Butylate	2,11	4,15	3,51	3,34	3,59	2,57	3,85	2,88	4,15
11	Carbaryl	2,40	2,36	2,50	2,39	2,88	2,22	2,35	2,71	2,36
12	Carbendazym (MBC)	2,35	1,52	1,65	0,99	2,00	1,05	1,55	1,23	1,52
13	Carbofuran	1,75	2,32	2,23	2,19	2,62	1,46	2,30	2,09	2,32
14	Chlorpropham	2,53	3,51	2,98	3,03	3,30	2,38	3,30	2,95	3,51
15	Cycloate	2,54	3,88	3,34	3,97	3,13	2,17	3,81	2,98	4,11
16	Diallate (cis)	3,28	4,49	3,27	3,68	4,04	2,74	4,08	3,44	3,61
17	Diallate (trans)	3,28	4,49	3,27	3,68	4,04	2,74	4,08	3,44	3,61
18	EPTC	2,38	3,21	2,88	3,32	2,91	1,97	3,02	2,29	3,21
19	Methiocarb	2,32	2,92	3,11	2,54	2,83	2,38	2,87	2,64	2,92
20	Methomyl	1,30	0,60	0,65	0,42	0,93	0,06	0,61	1,05	0,60
21	Molinate	1,92	3,21	2,50	2,56	2,56	1,58	2,91	1,95	3,21
22	Oxamyl	1,00	-0,48	0,21	-0,16	-0,22	-0,46	-1,20	0,33	-0,47
23	Pebulate	2,80	3,83	3,33	3,80	3,38	2,28	3,51	2,86	3,84

24	<i>Methyl-N-phenylcarbamate</i>	1,73	1,75	1,59	1,76	1,85	1,20	1,75	1,45	1,76
25	<i>Ethyl-N-phenylcarbamate</i>	1,82	2,30	1,94	2,19	2,28	1,51	2,24	1,87	2,30
26	<i>Propyl-N-phenylcarbamate</i>	2,06	2,80	2,47	2,57	2,75	1,82	2,73	2,23	2,80
27	<i>Butyl-N-phenylcarbamate</i>	2,26	3,30	2,92	3,11	3,21	2,11	3,22	2,80	3,30
28	<i>Pentyl-N-phenylcarbamate</i>	2,61	3,80	3,38	3,08	3,68	2,39	3,72	3,37	3,80
29	<i>Methyl-N-(3-chlorophenyl)carbamate</i>	2,15	2,58	2,26	2,37	2,46	1,78	2,40	2,07	2,58
30	<i>Methyl-N-(3,4-dichlorophenyl)carbamate</i>	2,74	3,32	2,92	3,07	3,07	2,35	3,04	2,69	3,54
31	<i>Proham</i>	1,83	2,60	2,32	2,60	2,69	1,82	2,66	2,33	2,60
32	<i>Propoxur</i>	1,67	1,52	2,30	1,97	2,44	1,57	1,90	1,98	1,52
33	<i>Thiobencarb</i>	3,27	3,40	3,73	3,51	3,59	3,09	3,90	3,28	3,40
34	<i>Triallate</i>	3,35	4,60	3,96	4,41	5,18	3,01	4,57	3,82	4,45
35	<i>Vernolate</i>	2,33	3,84	3,40	3,73	3,38	2,28	3,51	2,65	3,84
36	<i>Benfluralin</i>	3,99	5,29	4,40	5,18	4,55	4,21	5,31	4,71	5,29
37	<i>Butralin</i>	3,98		4,08	5,24	4,38	3,81	5,15	5,58	5,11
38	<i>Dinitramine</i>	3,63	4,30	2,67	3,76	2,90	2,93	3,96	3,61	3,19
39	<i>Fluchloralin</i>	3,55	4,63	4,26	4,91	4,27	4,21	5,07	4,49	5,04
40	<i>Nitralin</i>	2,92		3,05	2,74	2,85	2,96	2,92	3,10	3,69
41	<i>Oryzalin</i>	3,40		2,35	2,42	2,40	1,96	2,73	2,04	2,05
42	<i>Profluralin</i>	4,01	5,44	4,42	5,10	4,36	4,07	5,62	4,46	5,19
43	<i>Trifluralin</i>	3,93	5,34	4,47	5,09	4,55	4,21	5,31	4,50	5,34
44	<i>Aldrin</i>	4,69	6,50	4,26	5,90	5,31	5,36	6,75	4,16	4,55
45	<i>Chlordane</i>	5,15	6,16	4,78	6,02	5,77	5,47	6,26	4,36	4,94
46	<i>p,p-DDT</i>	5,31	6,91	6,33	6,29	6,94	6,17	6,79	6,65	6,91
47	<i>p,p-DDE</i>	4,82	6,51	5,93	6,22	6,82	5,85	6,00	6,73	6,96
48	<i>Dieldrin</i>	4,55	5,40	3,40	4,98	4,21	4,45	5,45	3,01	3,67
49	<i>Endosulfan</i>	4,13	3,83	2,14	4,32	3,82	3,12	3,50	0,84	3,83

50	<i>Lindane</i>	3,00	3,72	4,16	3,94	3,90	4,09	4,26	3,83	3,80
51	<i>Methomylchlor</i>	4,90	5,08	4,97	5,12	5,50	4,27	5,67	5,23	5,08
52	<i>Azinphos methyl</i>	2,28	2,75	2,93	2,75	1,86	0,91	2,53	2,22	2,75
53	<i>Carbophenothion</i>	4,66	5,33	4,83	5,50	5,34	3,09	5,19	5,53	5,33
54	<i>Chlorfenvinphos (cis)</i>	2,47	3,81	3,56	4,05	3,87	4,45	4,15	4,87	3,10
55	<i>Chlorfenvinphos (trans)</i>	2,47	3,81	3,56	4,05	3,87	4,45	4,15	4,87	3,10
56	<i>Chlorpyrifos</i>	3,70	4,96	4,60	5,15	5,11	3,21	4,66	5,45	5,27
57	<i>Chlorpyrifos methyl</i>	3,52	4,31	3,90	4,30	4,23	2,64	3,68	4,61	4,31
58	<i>Crotoxyphos (trans)</i>	2,00	3,30	1,98	2,10	2,49	2,56	1,89	2,28	3,30
59	<i>Diazinon</i>	2,75	3,81	3,14	4,45	3,95	2,43	3,86	3,63	3,81
60	<i>Dicrotophos (cis)</i>	1,66	-0,49	-0,42	-0,06	0,42	0,68	-1,10	-0,35	0,00
61	<i>Dimethoate</i>	1,20	0,78	0,57	1,21	0,66	-0,75	0,28	0,90	0,78
62	<i>Disulfoton</i>	3,22	4,02	2,95	4,11	3,58	1,65	3,86	4,17	4,02
63	<i>Ethion</i>	4,06	5,07	3,78	4,74	4,63	0,63	4,75	4,98	5,07
64	<i>Ethoprophos</i>	1,80	3,59	3,23	2,86	1,10	2,06	3,14	2,63	3,59
65	<i>Fenemiphos</i>	2,51	3,23	3,25	3,05	3,06	3,75	3,29	3,31	3,23
66	<i>Fenitrothion</i>	2,63	3,30	2,69	3,31	3,09	1,92	3,30	3,30	3,30
67	<i>Fensulfothion</i>	2,52	2,23	2,44	2,35	3,35	1,89	2,35	2,74	2,23
68	<i>Fonofos</i>	3,44	3,94	3,41	4,94	5,11	2,72	4,02	4,70	3,94
69	<i>Isazophos</i>	2,01	3,82	3,10	3,88	4,10	2,66	3,26	3,74	3,87
70	<i>Malathion</i>	3,07	2,36	1,89	2,67	1,66	0,49	2,29	2,01	2,36
71	<i>Mevinphos (cis)</i>	2,30	0,13	0,02	0,71	0,67	0,75	-0,24	0,04	1,20
72	<i>Mevinphos (trans)</i>	2,30		-0,05	0,38	0,08	0,75	-0,24	-0,05	0,14
73	<i>Parathions (= ethyl parathion)</i>	3,20	3,83	2,90	3,76	3,65	2,22	3,73	3,92	3,83
74	<i>Parathions methyl</i>	3,00	2,86	2,20	2,97	2,78	1,60	2,75	3,07	2,86
75	<i>Phorate</i>	2,70	3,56	2,93	3,71	3,80	1,32	3,37	3,81	3,56

76	<i>Phosalone</i>	2,63	4,38	4,00	3,96	4,26	1,65	4,29	3,96	4,36
77	<i>Profenofos</i>	3,03	4,68	4,79	4,30	4,33	3,74	4,82	4,52	4,68
78	<i>Terbufos</i>	2,82	4,48	3,52	4,61	4,85	1,97	4,24	4,50	4,48
79	<i>Trichlorfon</i>	1,90	0,51	0,75	0,81	0,65	0,49	-0,28	0,53	0,51
80	<i>Phenylurea</i>	1,50	0,83	0,74	0,85	0,86	0,86	0,71	0,40	0,83
81	<i>2-chlorophenylurea</i>	1,61	1,27	1,40	1,74	1,47	1,47	1,35	1,02	1,27
82	<i>2-fluorophenylurea</i>	1,32	0,88	0,95	1,20	0,92	1,30	0,91	0,56	0,88
83	<i>3-chlorophenylurea</i>	2,01	1,82	1,40	1,58	1,47	1,47	1,35	1,02	1,82
84	<i>3-fluorophenylurea</i>	1,77	1,29	0,95	1,07	0,92	1,30	0,91	0,56	1,29
85	<i>3-bromophenylurea</i>	2,12	2,08	1,49	2,04	1,55	1,63	1,60	1,20	2,08
86	<i>3-methyphenylurea</i>	1,56	1,29	1,23	1,17	1,17	1,20	1,25	0,84	1,29
87	<i>3-trifluoromethylphenylurea</i>	1,98	2,31	1,68	1,73	1,62	1,94	1,67	1,32	2,31
88	<i>4-fuorophenylurea</i>	1,52	1,04	0,95	1,13	0,92	1,30	0,91	0,56	1,04
89	<i>4-bromophenylurea</i>	2,06	1,98	1,49	2,10	1,55	1,63	1,60	1,20	1,98
90	<i>4-phenoxyphenylurea</i>	2,56	2,80	2,30	2,49	2,27	2,14	2,76	1,88	2,80
91	<i>3,4-dichlorophenylurea</i>	2,53	2,65	2,07	2,35	2,08	2,05	2,00	1,64	2,64
92	<i>3-chloro-4-meghoxyphylurea</i>	2,00	1,37	1,39	1,78	1,36	1,24	1,43	0,93	1,37
93	<i>3-methyl-4-fluorophenylurea</i>	1,75	1,59	1,43	1,36	1,23	1,63	1,45	0,79	1,59
94	<i>3-mthyl-4-bromophenylurea</i>	2,37	2,49	1,98	2,30	1,87	1,94	2,14	1,42	2,49
95	<i>N-phenyl-N'-cyclopropylurea</i>	1,74	1,65	1,44	1,62	2,01	1,43	1,97	1,46	1,65
96	<i>N-phenyl-N'-cyclopentylurea</i>	1,93	2,65	2,35	2,65	2,65	2,00	2,95	2,18	2,65
97	<i>N-phenyl-N'-cyclohexylurea</i>	2,07	2,77	2,81	3,25	2,96	2,27	3,44	2,74	2,91
98	<i>N-phenyl-N'-cycloheptylurea</i>	2,37		3,26	3,61	3,28	2,53	3,93	3,31	3,45
99	<i>Siduron</i>	2,31	3,80	3,12	3,34	3,16	2,53	3,86	3,04	3,34
100	<i>N-phenyl-N'-methylurea</i>	1,29	0,42	0,95	0,35	1,28	1,20	0,53	0,54	0,42
101	<i>N-(3-chlorophenyl)-N'-methylurea</i>	1,93	2,18	1,61	1,96	1,99	1,78	1,82	1,54	2,18

102	<i>N</i> -(3,4-dichlorophenyl)- <i>N'</i> -methylurea	2,46	2,94	2,28	2,93	2,60	2,35	2,46	2,16	2,94
103	<i>N</i> -(3-chloro-4-methylphenyl)- <i>N'</i> -methylurea	2,10	2,61	2,10	2,28	2,31	2,09	2,36	1,76	2,61
104	<i>N</i> -(3-chloro-4-methoxyphenyl)- <i>N'</i> -methylurea	1,84	1,76	1,59	1,99	1,88	1,55	1,90	1,45	1,76
105	<i>Fenuron</i>	1,40	0,98	1,15	1,15	1,60	1,51	1,38	1,06	0,98
106	<i>N</i> -(3-chlorophenyl)- <i>N',N'</i> -dimethylurea	1,79	2,00	1,82	1,95	2,21	2,09	2,03	1,68	2,00
107	<i>N</i> -(3-methoxyphenyl)- <i>N',N'</i> -dimethylurea	1,72	1,73	1,14	1,18	1,49	1,28	1,46	0,97	1,29
108	<i>N</i> -(3-fluorophenyl)- <i>N',N'</i> -dimethylurea	1,73	1,37	1,36	1,32	1,66	1,94	1,58	1,22	1,37
109	<i>Fluometuron</i>	2,00	2,42	2,10	2,16	2,36	2,52	2,35	1,98	2,42
110	<i>N</i> -(4-fluorophenyl)- <i>N',N'</i> -dimethylurea	1,43	1,13	1,36	1,46	1,66	1,94	1,58	1,22	1,13
111	<i>Monuron</i>	1,95	1,94	1,82	1,96	2,21	2,09	2,03	1,68	1,94
112	<i>N</i> -(4-methylphenyl)- <i>N',N'</i> -dmethylurea	1,51	1,33	1,64	1,49	1,91	1,82	1,93	1,50	1,33
113	<i>N</i> -(4-methoxyphenyl)- <i>N',N'</i> -dimethylurea	1,40	0,83	1,14	1,44	1,49	1,28	1,46	0,97	0,83
114	<i>Metoxuron</i>	1,72	1,64	1,80	2,26	2,11	1,84	2,11	1,60	1,64
115	<i>Chlorotoluron</i>	2,02	2,41	2,30	2,25	2,53	2,38	2,58	1,91	2,41
116	<i>N</i> -(3,5-dimethylphenyl)- <i>N',N'</i> -dimethylurea	1,73	1,90	2,13	1,81	2,23	2,11	2,48	1,93	1,90
117	<i>N</i> -(3,5-diMe-4-Br-phenyl)- <i>N',N'</i> -dimethylurea	2,53	2,92	2,87	2,68	2,93	2,80	3,37	2,31	2,92
118	<i>Diuron</i>	2,40	2,68	2,48	2,92	2,82	2,65	2,67	2,30	2,68
119	<i>Chloroxuron</i>	3,55	3,70	3,38	3,40	3,62	3,18	4,08	3,16	3,47
120	<i>Monolinuron</i>	2,10	2,30	1,66	1,99	1,74	2,09	2,26	1,59	2,30
121	<i>Metobromuron</i>	2,10	2,38	1,75	2,18	1,83	2,24	2,51	1,77	2,38
122	<i>Linuron</i>	2,70	3,20	2,33	2,82	2,36	2,65	2,91	2,21	3,20
123	<i>Chlorobromuron</i>	2,70	3,09	2,41	3,02	2,44	2,79	3,15	2,39	3,09
124	<i>Ametryn</i>	2,59	2,98	2,74	3,09	2,36	2,59	3,32	1,78	2,98
125	<i>Atrazine</i>	2,24	2,61	2,54	2,70	2,48	2,59	2,82	1,66	2,61
126	<i>Cyanazine</i>	2,28	2,22	2,34	2,05	2,18	2,03	2,51	0,91	2,22
127	<i>Dipropetryn</i>	3,07	3,81	3,47	4,15	3,28	3,18	4,22	2,60	3,88

128	<i>Metribuzin</i>	1,71	1,70	1,47	1,54	1,32	1,60	1,49	1,08	1,70
129	<i>Prometon</i>	2,60	2,99	2,56	2,80	2,35	2,53	3,57	2,06	2,99
130	<i>Prometry</i>	2,85	3,51	3,12	3,31	2,77	2,89	3,73	2,24	3,51
131	<i>Propazine</i>	2,40	2,93	2,91	2,94	2,88	2,89	3,24	2,12	2,93
132	<i>Simazine</i>	2,10	2,18	2,16	2,48	2,07	2,27	2,40	1,20	2,18
133	<i>Secbumeton</i>	2,78		2,71	2,90	2,41	2,53	3,64	1,95	3,22
134	<i>Terbutryn</i>	2,85	3,74	2,95	3,65	2,51	2,89	3,77	2,20	3,74
135	<i>Trietazine</i>	2,76	3,34	2,86	3,58	2,94	2,89	3,44	2,19	3,34
136	<i>Ipazine</i>	2,91	3,94	3,24	3,65	3,34	3,18	3,86	2,65	3,77
137	<i>Imazalil</i>	3,73	3,82	3,35	3,62	2,94	3,14	4,10	3,75	3,82
138	<i>Oxamdiazon</i>	3,51	4,80	5,22	4,32	5,40	4,03	4,81	4,85	4,80
139	<i>Propiconazole</i>	3,39	3,72	3,80	3,48	3,34	4,17	4,13	3,42	3,50
140	<i>Tebuthiuron</i>	1,83	1,79	1,43	1,59	2,12	1,24	1,78	1,32	1,60
141	<i>Thiabendazole</i>	3,24	2,47	2,15	2,47	2,01	1,53	2,00	2,17	2,47
142	<i>Triadimefon</i>	2,71	2,77	3,49	2,66	2,78	3,32	2,94	2,72	2,77
143	<i>Tricyclazole</i>	3,09	1,70	2,55	1,35	1,69	2,39	2,48	3,33	1,70

N	Compostos do conjunto de teste A	logK _{oc}	logP _{exp}	ALOGP	ALOGPS	ACLOGP	MLOGP	KOWWIN	XLOGP2	XLOGP3
A1	<i>Aldicarb sulfoxide</i>	0,56	Nc	0,06	-0,53	0,49	-0,10	-0,78	-0,35	-0,20
A2	<i>Asulam</i>	2,48	-0,27	0,42	0,29	0,43	-0,26	0,05	-0,39	-0,27
A3	<i>Chlorbufam</i>	2,21	Nc	3,83	2,81	2,76	2,57	3,02	2,86	3,10
A4	<i>Fenobucarb</i>	1,71	2,78	3,24	3,06	3,26	2,39	2,86	3,21	2,78
A5	<i>Pirimicarb (= pirimor)</i>	1,90	1,70	2,01	1,45	1,94	1,65	1,40	1,50	1,70
A6	<i>Triodicarb</i>	2,54	1,70	2,10	2,14	2,01	0,52	1,91	2,7	1,70
A7	<i>Xylicarb</i>	1,71	2,09	2,57	2,18	2,33	1,82	2,27	2,32	2,07
A8	<i>Demeton-S-methyl</i>	1,49	1,02	1,36	1,00	1,19	0,94	1,11	1,37	1,02
A9	<i>Dichlorvos</i>	1,67	1,47	0,47	1,27	1,66	1,25	0,60	1,21	1,43

A10	<i>EPN</i>	3,12	4,78	3,98	4,88	4,86	3,31	4,47	5,42	4,40
A11	<i>Iprobenfos</i>	2,40	3,34	3,43	2,76	3,17	2,80	3,57	3,25	3,21
A12	<i>Leptophos</i>	4,50	6,31	5,81	6,37	6,34	4,54	6,34	7,15	6,31
A13	<i>Kmethidathion</i>	1,53	2,20	2,36	1,18	1,37	-0,27	1,58	1,77	2,42
A14	<i>Piperophos</i>	3,44	4,04	3,81	4,24	4,09	1,64	4,23	3,93	4,23
A15	<i>Pirimiphos methyl</i>	3,00	4,20	2,73	3,91	3,48	2,05	3,44	3,71	4,20
A16	<i>Sulprofos</i>	4,08	5,48	4,82	5,25	5,04	3,09	5,65	5,79	4,90
A17	<i>Neburon</i>	3,40	4,10	3,81	4,12	4,19	3,46	4,15	3,65	3,80
A18	<i>Anilazine</i>	3,00	3,88	4,39	4,05	4,98	3,86	3,64	2,88	3,00
A19	<i>Cyromazine</i>	2,30	Nc	0,31	0,04	0,39	1,22	0,96	-0,36	-0,21
A20	<i>Terbutylazine</i>	2,32	3,21	2,74	3,33	2,63	2,89	3,27	2,08	3,06
N	Compostos do conjunto de teste B	logK_{oc}	logP_{exp}	ALOGP	ALOGPS	ACLOGP	MLOGP	KOWWIN	XLOGP2	XLOGP3
B1	<i>1,3-dichloropropene</i>	1,51	2,06	1,52	2,07	2,13	2,08	2,29	2,10	1,74
B2	<i>2,4,5-T</i>	1,72	3,31	3,48	3,44	2,80	2,91	3,26	3,18	3,31
B3	<i>2,6-dichloro-4-nitraniline</i>	3,00	2,80	2,31	3,01	2,49	2,57	2,76	2,34	2,89
B4	<i>2,6-dichlorobenzonitrile</i>	2,60	2,74	3,04	3,16	3,02	2,96	2,83	2,99	2,74
B5	<i>Amitrole</i>	2,00	-0,86	-0,64	-0,89	-0,47	-0,23	-0,47	0,72	-0,36
B6	<i>Acephate</i>	0,30	-0,85	-0,28	-0,66	0,11	-0,76	-0,90	-0,27	-0,85
B7	<i>Bendiocarb</i>	2,76	1,70	1,48	1,53	2,86	1,06	2,55	2,08	1,70
B8	<i>Biphenox</i>	4,00	4,48	4,47	4,31	4,60	3,72	4,15	4,57	4,48
B9	<i>Bromacil</i>	2,09	2,11	1,41	1,20	2,06	1,80	1,68	1,46	2,11
B10	<i>Bromomethane</i>	1,34	1,19	0,99	0,68	1,10	1,12	1,18	1,38	0,99
B11	<i>Carboxin</i>	2,41	2,14	1,41	1,65	1,84	1,51	1,49	2,17	2,14
B12	<i>Chinomethionate</i>	3,36	3,78	4,13	2,57	2,96	0,45	3,37	2,14	3,11
B13	<i>Chlorimuron-ethyl</i>	2,04	2,50	2,44	2,25	2,55	1,68	2,29	2,75	2,71
B14	<i>Chlorothalonil</i>	3,14	3,05	4,24	3,98	4,06	3,13	3,66	3,96	2,90

B15	<i>Cyfluthrin</i>	5,00	5,95	5,62	5,93	5,98	4,28	5,74	5,65	6,15
B16	<i>Desmedipham</i>	3,18	3,39	3,28	3,67	3,58	2,09	3,22	2,86	3,39
B17	<i>Diclofop-methyl</i>	4,20	4,62	5,00	5,13	4,45	3,66	4,54	4,89	4,80
B18	<i>Diflubenzuron</i>	4,06	3,88	3,47	3,93	3,62	3,51	3,59	4,46	3,88
B19	<i>Dimethipin</i>	0,48	-0,17	-0,37	-0,93	0,23	-0,23	0,04	-0,58	-0,50
B20	<i>Esphenvalerate</i>	3,72	6,20	6,67	6,30	6,38	4,83	6,76	6,23	6,20
B21	<i>Fenvalerate</i>	3,72	6,20	6,67	6,30	6,38	4,83	6,76	6,23	6,20
B22	<i>Fluvalinate</i>	6,00	Nc	7,50	6,65	6,46	4,81	6,81	6,65	7,72
B23	<i>Hexazinone</i>	1,73	1,85	2,60	1,34	2,24	2,29	2,15	1,24	1,26
B24	<i>Iprodione</i>	2,85	3,00	2,88	2,23	3,00	2,01	2,85	2,17	3,10
B25	<i>Isofenphos</i>	2,78	4,12	3,70	4,48	4,86	2,99	4,65	5,11	4,12
B26	<i>Cyhalothrin</i>	5,26	6,02	5,90	5,93	5,92	4,22	6,85	5,65	6,11
B27	<i>Methamidophos</i>	0,70	-0,80	-0,15	-0,94	-0,42	-1,22	-0,93	-0,55	-0,88
B28	<i>Napropamide</i>	2,85	3,36	3,28	3,43	3,62	3,10	3,33	3,89	3,36
B29	<i>Nitrapyrin</i>	2,76	3,41	3,51	3,87	3,42	2,69	3,35	2,91	3,41
B30	<i>Oxydemeton-methyl</i>	1,00	-0,75	0,24	0,81	-0,76	0,03	-1,03	-0,66	-0,69
B31	<i>Pendimethalin</i>	3,70	5,18	4,17	4,36	3,96	3,55	4,82	4,79	5,18
B32	<i>Pentachloronitrobenzene</i>	3,70	4,64	5,05	4,73	5,05	4,29	5,03	5,02	4,22
B33	<i>Propargite</i>	3,60	5,00	5,03	3,94	3,18	3,27	5,57	3,44	4,98
B34	<i>Terbacil</i>	1,95	1,89	1,01	1,78	1,70	1,65	1,75	1,34	1,89
B35	<i>Tetramethylthiuram disulfide</i>	2,83	1,73	3,16	2,18	-0,10	0,45	1,70	0,18	1,69
B36	<i>Thidiazuron</i>	2,04	Nc	1,61	1,93	-0,21	1,25	2,10	2,43	1,34
B37	<i>Tridiphane</i>	3,75	Nc	3,98	4,74	5,28	4,05	5,18	4,75	4,67
B38	<i>Triforine</i>	2,30	2,20	2,16	2,03	1,61	1,40	2,02	2,25	2,76

ANEXO 5 - Tabela dos valores de $\log K_{oc}$ e dos descritores de todos os pesticidas usados no artigo 2

N	Compostos do conjunto de treinamento	$\log K_{oc}$	$\log P$	$\log S$	$\log S_p$	M_v	VAR	MAXDP
1	<i>Acetochlor</i>	2,32	3,17	-3,04	-6,21	0,593	52	4,953
2	<i>Alachlor</i>	2,28	3,02	-3,11	-6,13	0,593	44	4,953
3	<i>Butachlor</i>	2,86	4,20	-4,64	-8,84	0,577	84	5,226
4	<i>Metaxyl</i>	1,57	1,47	-2,32	-3,80	0,589	58	5,303
5	<i>Metolachlor</i>	2,46	3,37	-3,43	-6,80	0,587	54	5,188
6	<i>Propachlor</i>	2,42	2,15	-2,37	-4,52	0,616	28	4,548
7	<i>Aldicarb</i>	1,50	1,58	-2,22	-3,81	0,570	24	3,532
8	<i>Aldicar sulfone (= aldoxycarb)</i>	0,42	0,01	-2,33	-2,64	0,580	34	4,151
9	<i>Benomyl</i>	2,71	1,78	-2,91	-4,70	0,613	82	5,305
10	<i>Butylate</i>	2,11	3,34	-3,18	-6,52	0,533	34	4,703
11	<i>Carbaryl</i>	2,40	2,39	-3,34	-5,73	0,657	36	4,092
12	<i>Carbendazym (MBC)</i>	2,35	0,99	-2,37	-3,40	0,673	32	3,845
13	<i>Carbofuran</i>	1,75	2,19	-2,55	-4,74	0,608	40	4,154
14	<i>Chlorpropham</i>	2,53	3,03	-2,95	-5,98	0,632	28	4,158
15	<i>Cycloate</i>	2,54	3,97	-3,14	-7,11	0,548	34	4,762
16	<i>Diallate (cis)</i>	3,28	3,68	-4,03	-7,71	0,604	39	4,806
17	<i>Diallate (trans)</i>	3,28	3,68	-4,03	-7,71	0,604	39	4,806
18	<i>EPTC</i>	2,38	3,32	-2,30	-5,62	0,538	26	4,412
19	<i>Methiocarb</i>	2,32	2,54	-2,88	-5,42	0,611	36	4,020
20	<i>Methomyl</i>	1,30	0,42	-1,17	-1,73	0,588	16	3,383
21	<i>Molinate</i>	1,92	2,56	-2,10	-4,66	0,556	26	4,440
22	<i>Oxamyl</i>	1,00	-0,16	-1,59	-1,82	0,594	34	4,338
23	<i>Pebulate</i>	2,80	3,80	-2,94	-6,74	0,535	33	4,528

24	<i>Methyl-N-phenylcarbamate</i>	1,73	1,76	-1,37	-3,14	0,628	20	3,648
25	<i>Ethyl-N-phenylcarbamate</i>	1,82	2,19	-1,61	-3,80	0,612	26	3,879
26	<i>Propyl-N-phenylcarbamate</i>	2,06	2,57	-2,01	-4,58	0,600	33	4,038
27	<i>Butyl-N-phenylcarbamate</i>	2,26	3,11	-2,56	-5,67	0,591	40	4,154
28	<i>Pentyl-N-phenylcarbamate</i>	2,61	3,08	-3,06	-6,14	0,583	48	4,242
29	<i>Methyl-N-(3-chlorophenyl)carbamate</i>	2,15	2,37	-2,21	-4,58	0,669	24	3,716
30	<i>Methyl-N-(3,4-dichlorophenyl)carbamate</i>	2,74	3,07	-3,22	-6,29	0,711	28	3,768
31	<i>Proham</i>	1,83	2,60	-1,97	-4,57	0,600	24	4,090
32	<i>Propoxur</i>	1,67	1,97	-2,47	-4,44	0,595	36	4,030
33	<i>Thiobencarb</i>	3,27	3,51	-4,16	-7,67	0,624	40	4,708
34	<i>Triallate</i>	3,35	4,41	-4,58	-8,99	0,630	36	4,875
35	<i>Vernolate</i>	2,33	3,73	-2,96	-6,69	0,535	33	4,571
36	<i>Benfluralin</i>	3,99	5,18	-6,00	-11,18	0,623	81	4,833
37	<i>Butralin</i>	3,98	5,24	-4,44	-9,68	0,587	60	4,289
38	<i>Dinitramine</i>	3,63	3,76	-5,02	-8,78	0,636	50	4,893
39	<i>Fluchloralin</i>	3,55	4,91	-5,91	-10,82	0,655	66	4,824
40	<i>Nitralin</i>	2,92	2,74	-4,78	-7,52	0,613	66	4,656
41	<i>Oryzalin</i>	3,40	2,42	-4,35	-6,77	0,616	66	4,434
42	<i>Profluralin</i>	4,01	5,10	-5,52	-10,62	0,632	68	4,917
43	<i>Trifluralin</i>	3,93	5,09	-5,77	-10,86	0,623	66	4,845
44	<i>Aldrin</i>	4,69	5,90	-6,65	-12,55	0,794	24	2,651
45	<i>Chlordane</i>	5,15	6,02	-6,77	-12,79	0,846	34	2,508
46	<i>p,p-DDT</i>	5,31	6,29	-8,00	-14,29	0,779	55	2,002
47	<i>p,p-DDE</i>	4,82	6,22	-6,64	-12,86	0,787	50	1,860
48	<i>Dieldrin</i>	4,55	4,98	-6,27	-11,25	0,791	31	2,713
49	<i>Endosulfan</i>	4,13	4,32	-3,16	-7,48	0,818	44	4,316

50	<i>Lindane</i>	3,00	3,94	-4,73	-8,67	0,785	10	1,770
51	<i>Methomylchlor</i>	4,90	5,12	-6,84	-11,96	0,685	74	2,073
52	<i>Azinphos methyl</i>	2,28	2,75	-3,96	-6,71	0,682	47	5,166
53	<i>Carbophenothion</i>	4,66	5,50	-6,11	-11,61	0,661	52	2,059
54	<i>Chlorfenvinphos (cis)</i>	2,47	4,05	-4,33	-8,38	0,677	56	5,299
55	<i>Chlorfenvinphos (trans)</i>	2,47	4,05	-4,33	-8,38	0,677	56	5,299
56	<i>Chlorpyrifos</i>	3,70	5,15	-5,38	-10,53	0,705	44	1,945
57	<i>Chlorpyrifos methyl</i>	3,52	4,30	-4,99	-9,29	0,756	32	1,762
58	<i>Crotoxyphos (trans)</i>	2,00	2,10	-2,76	-4,86	0,612	62	4,750
59	<i>Diazinon</i>	2,75	4,45	-3,94	-8,39	0,589	49	2,185
60	<i>Dicrotophos (cis)</i>	1,66	-0,06	-1,17	-1,44	0,572	30	4,480
61	<i>Dimethoate</i>	1,20	1,21	-2,04	-3,28	0,609	24	3,819
62	<i>Disulfoton</i>	3,22	4,11	-4,47	-8,58	0,581	42	2,007
63	<i>Ethion</i>	4,06	4,74	-5,26	-10,00	0,604	49	2,055
64	<i>Ethoprophos</i>	1,80	2,86	-2,36	-5,22	0,562	32	5,035
65	<i>Fenemiphos</i>	2,51	3,05	-2,98	-6,03	0,587	56	5,498
66	<i>Fenitrothion</i>	2,63	3,31	-4,07	-7,38	0,651	37	3,632
67	<i>Fensulfothion</i>	2,52	2,35	-2,76	-5,11	0,626	46	4,254
68	<i>Fonofos</i>	3,44	4,94	-5,02	-9,96	0,629	30	2,171
69	<i>Isazophos</i>	2,01	3,88	-3,67	-7,55	0,610	44	1,948
70	<i>Malathion</i>	3,07	2,67	-3,30	-5,97	0,601	61	4,802
71	<i>Mevinphos (cis)</i>	2,30	0,71	-1,27	-2,06	0,589	32	4,371
72	<i>Mevinphos (trans)</i>	2,30	0,38	-1,29	-1,82	0,589	30	4,303
73	<i>Parathions (= ethyl parathion)</i>	3,20	3,76	-4,37	-8,13	0,637	46	3,504
74	<i>Parathions methyl</i>	3,00	2,97	-3,74	-6,71	0,667	36	3,412
75	<i>Phorate</i>	2,70	3,71	-3,85	-7,56	0,589	35	1,970

76	<i>Phosalone</i>	2,63	3,96	-5,03	-8,99	0,673	60	4,919
77	<i>Profenofos</i>	3,03	4,30	-4,32	-8,62	0,662	52	5,427
78	<i>Terbufos</i>	2,82	4,61	-4,84	-9,45	0,575	33	2,038
79	<i>Trichlorfon</i>	1,90	0,81	-1,56	-2,43	0,671	18	4,350
80	<i>Phenylurea</i>	1,50	0,85	-1,24	-2,15	0,630	14	3,280
81	<i>2-chlorophenylurea</i>	1,61	1,74	-1,88	-3,63	0,676	17	3,373
82	<i>2-fluorophenylurea</i>	1,32	1,20	-1,44	-2,67	0,651	17	4,680
83	<i>3-chlorophenylurea</i>	2,01	1,58	-1,91	-3,50	0,676	17	3,348
84	<i>3-fluorophenylurea</i>	1,77	1,07	-1,47	-2,58	0,651	17	4,425
85	<i>3-bromophenylurea</i>	2,12	2,04	-2,54	-4,58	0,687	17	3,376
86	<i>3-methyphenylurea</i>	1,56	1,17	-1,75	-2,95	0,613	17	3,391
87	<i>3-trifluoromethylphenylurea</i>	1,98	1,73	-2,69	-4,43	0,667	26	4,161
88	<i>4-fuorophenylurea</i>	1,52	1,13	-1,51	-2,67	0,651	18	4,280
89	<i>4-bromophenylurea</i>	2,06	2,10	-2,54	-4,64	0,687	18	3,353
90	<i>4-phenoxyphenylurea</i>	2,56	2,49	-3,30	-5,79	0,659	44	3,625
91	<i>3,4-dichlorophenylurea</i>	2,53	2,35	-2,84	-5,19	0,722	20	3,400
92	<i>3-chloro-4-meghoxyphylurea</i>	2,00	1,78	-2,39	-4,18	0,655	25	3,471
93	<i>3-methyl-4-fluorophenylurea</i>	1,75	1,36	-1,71	-3,09	0,631	20	4,692
94	<i>3-mthyl-4-bromophenylurea</i>	2,37	2,30	-2,96	-5,26	0,661	20	3,464
95	<i>N-phenyl-N'-cyclopropylurea</i>	1,74	1,62	-2,07	-3,70	0,616	23	4,241
96	<i>N-phenyl-N'-cyclopentylurea</i>	1,93	2,65	-2,75	-5,40	0,595	31	4,546
97	<i>N-phenyl-N'-cyclohexylurea</i>	2,07	3,25	-3,03	-6,28	0,587	36	4,658
98	<i>N-phenyl-N'-cycloheptylurea</i>	2,37	3,61	-3,30	-6,91	0,581	41	4,771
99	<i>Siduron</i>	2,31	3,34	-3,09	-6,43	0,581	41	4,804
100	<i>N-phenyl-N'-methylurea</i>	1,29	0,35	-0,76	-1,27	0,613	15	3,648
101	<i>N-(3-chlorophenyl)-N'-methylurea</i>	1,93	1,96	-2,47	-4,43	0,652	24	3,827

102	<i>N</i> -(3,4-dichlorophenyl)- <i>N'</i> -methylurea	2,46	2,93	-3,13	-6,06	0,691	28	3,879
103	<i>N</i> -(3-chloro-4-methylphenyl)- <i>N'</i> -methylurea	2,10	2,28	-3,02	-5,30	0,634	28	3,912
104	<i>N</i> -(3-chloro-4-methoxyphenyl)- <i>N'</i> -methylurea	1,84	1,99	-3,02	-5,01	0,637	32	3,950
105	<i>Fenuron</i>	1,40	1,15	-1,25	-2,43	0,600	18	4,127
106	<i>N</i> -(3-chlorophenyl)- <i>N',N'</i> -dimethylurea	1,79	1,95	-2,48	-4,43	0,634	22	4,195
107	<i>N</i> -(3-methoxyphenyl)- <i>N',N'</i> -dimethylurea	1,72	1,18	-1,66	-2,87	0,594	30	4,286
108	<i>N</i> -(3-fluorophenyl)- <i>N',N'</i> -dimethylurea	1,73	1,32	-1,56	-2,90	0,616	22	4,654
109	<i>Fluometuron</i>	2,00	2,16	-3,24	-5,40	0,632	34	4,341
110	<i>N</i> -(4-fluorophenyl)- <i>N',N'</i> -dimethylurea	1,43	1,46	-1,54	-3,01	0,616	26	4,460
111	<i>Monuron</i>	1,95	1,96	-2,42	-4,38	0,634	26	4,179
112	<i>N</i> -(4-methylphenyl)- <i>N',N'</i> -dmethylurea	1,51	1,49	-1,91	-3,41	0,590	26	4,212
113	<i>N</i> -(4-methoxyphenyl)- <i>N',N'</i> -dimethylurea	1,40	1,44	-1,62	-3,08	0,594	34	4,250
114	<i>Metoxuron</i>	1,72	2,26	-2,71	-4,97	0,624	35	4,318
115	<i>Chlorotoluron</i>	2,02	2,25	-2,87	-5,12	0,620	26	4,280
116	<i>N</i> -(3,5-dimethylphenyl)- <i>N',N'</i> -dimethylurea	1,73	1,81	-2,07	-3,89	0,582	26	4,350
117	<i>N</i> -(3,5-diMe-4-Br-phenyl)- <i>N',N'</i> -dimethylurea	2,53	2,68	-3,22	-5,90	0,616	30	4,423
118	<i>Diuron</i>	2,40	2,92	-3,23	-6,15	0,669	26	4,247
119	<i>Chloroxuron</i>	3,55	3,40	-4,06	-7,46	0,657	64	4,491
120	<i>Monolinuron</i>	2,10	1,99	-2,69	-4,68	0,637	30	4,285
121	<i>Metobromuron</i>	2,10	2,18	-2,90	-5,08	0,645	30	4,306
122	<i>Linuron</i>	2,70	2,82	-3,40	-6,22	0,670	35	4,353
123	<i>Chlorobromuron</i>	2,70	3,02	-3,26	-6,28	0,678	35	4,375
124	<i>Ametryn</i>	2,59	3,09	-3,63	-6,72	0,577	35	1,264
125	<i>Atrazine</i>	2,24	2,70	-3,90	-6,60	0,592	32	1,616
126	<i>Cyanazine</i>	2,28	2,05	-2,74	-4,79	0,623	42	2,865
127	<i>Dipropetryn</i>	3,07	4,15	-4,25	-8,40	0,566	41	1,360

128	<i>Metribuzin</i>	1,71	1,54	-2,20	-3,75	0,594	28	4,717
129	<i>Prometon</i>	2,60	2,80	-3,69	-6,49	0,557	32	1,516
130	<i>Prometry</i>	2,85	3,31	-3,84	-7,15	0,571	32	1,304
131	<i>Propazine</i>	2,40	2,94	-3,99	-6,93	0,584	29	1,664
132	<i>Simazine</i>	2,10	2,48	-3,24	-5,72	0,602	27	1,568
133	<i>Secbumeton</i>	2,78	2,90	-3,70	-6,60	0,557	40	1,515
134	<i>Terbutryn</i>	2,85	3,65	-3,85	-7,50	0,571	40	1,309
135	<i>Trietazine</i>	2,76	3,58	-3,01	-6,59	0,584	37	1,704
136	<i>Ipazine</i>	2,91	3,65	-3,56	-7,21	0,577	34	1,751
137	<i>Imazalil</i>	3,73	3,62	-3,92	-7,54	0,670	56	2,277
138	<i>Oxamdiazon</i>	3,51	4,32	-5,07	-9,39	0,640	62	5,059
139	<i>Propiconazole</i>	3,39	3,48	-4,15	-7,63	0,650	76	2,731
140	<i>Tebuthiuron</i>	1,83	1,59	-2,89	-4,49	0,585	34	4,325
141	<i>Thiabendazole</i>	3,24	2,47	-3,16	-5,63	0,752	22	1,388
142	<i>Triadimefon</i>	2,71	2,66	-3,37	-6,03	0,639	64	5,495
143	<i>Tricyclazole</i>	3,09	1,35	-2,07	-3,44	0,715	17	1,029

N	Compostos do conjunto de teste A	logK_{oc}	logP	logS	logS_p	Mv	VAR	MAXDP
A1	<i>Aldicarb sulfoxide</i>	0,56	-0,53	-1,42	-1,53	0,575	29	4,074
A2	<i>Asulam</i>	2,48	0,29	-1,79	-2,26	0,648	36	4,460
A3	<i>Chlorbufam</i>	2,21	2,81	-3,96	-6,77	0,677	37	4,221
A4	<i>Fenobucarb</i>	1,71	3,06	-3,15	-6,21	0,583	36	4,111
A5	<i>Pirimicarb (= pirimor)</i>	1,90	1,45	-2,48	-3,95	0,577	38	4,487
A6	<i>Triodicarb</i>	2,54	2,14	-3,23	-5,37	0,620	75	4,591
A7	<i>Xylicarb</i>	1,71	2,18	-2,36	-4,54	0,600	28	3,844
A8	<i>Demeton-S-methyl</i>	1,49	1	-1,8	-2,84	0,580	30	4,423
A9	<i>Dichlorvos</i>	1,67	1,27	-1,35	-2,64	0,671	18	4,027

A10	<i>EPN</i>	3,12	4,88	-5,51	-10,39	0,677	65	3,648
A11	<i>Iprobenfos</i>	2,40	2,76	-2,97	-5,73	0,591	48	5,526
A12	<i>Leptophos</i>	4,50	6,37	-6,79	-13,16	0,764	50	2,370
A13	<i>Kmethidathion</i>	1,53	1,18	-2,17	-3,38	0,667	39	4,415
A14	<i>Piperophos</i>	3,44	4,24	-5,04	-9,28	0,568	65	5,368
A15	<i>Pirimiphos methyl</i>	3,00	3,91	-2,96	-6,87	0,591	45	2,065
A16	<i>Sulprofos</i>	4,08	5,25	-5,84	-11,09	0,627	56	2,401
A17	<i>Neburon</i>	3,40	4,12	-3,77	-7,89	0,625	54	4,753
A18	<i>Anilazine</i>	3,00	4,05	-3,74	-7,79	0,792	30	1,843
A19	<i>Cyromazine</i>	2,30	0,04	-1,79	-2,11	0,599	19	1,373
A20	<i>Terbutylazine</i>	2,32	3,33	-4,15	-7,48	0,584	37	1,661
