

QSAR MODELING: UM NOVO PACOTE COMPUTACIONAL OPEN SOURCE PARA GERAR E VALIDAR MODELOS QSAR

João Paulo A. Martins e Márcia M. C. Ferreira*

Instituto de Química, Universidade Estadual de Campinas, CP 6154, 13084-971 Campinas – SP, Brasil

Amostra	SpPosA_RG	SM11	AEA(...)	SpMin3_Bh...	RDF040u	E1m	H8u	y
Amostra1	0,423	1,618	1,295	5,008	0,266	0	3,35	
Amostra2	0,424	1,618	1,53	6,324	0,275	0	3,87	
Amostra3	0,423	1,618	1,425	5,457	0,267	0	4	
Amostra4	0,425	1,618	1,622	9,568	0,269	0	4,39	
Amostra5	0,425	1,618	1,557	6,325	0,31	0	4,38	
Amostra6	0,424	1,948	1,533	5,814	0,242	0	4,31	
Amostra7	0,424	1,952	1,57	6,657	0,267	0	4,44	
Amostra8	0,423	1,618	1,539	7,197	0,247	0	4,31	
Amostra9	0,424	1,737	1,569	7,059	0,269	0	4,31	
Amostra10	0,42	2,044	1,572	7,657	0,276	0,31	4,73	
Amostra11	0,419	1,802	1,417	8,02	0,33	0	4,5	
Amostra12	0,421	2,362	1,672	7,875	0,308	0	5,69	
Amostra13	0,423	1,806	1,526	9,015	0,301	0	4,52	
Amostra14	0,422	1,927	1,544	9,753	0,303	0	5,08	
Amostra15	0,424	2,028	1,598	9,053	0,279	0	5,24	
Amostra16	0,424	2,08	1,623	9,594	0,306	0,033	5,15	
Amostra17	0,417	2,342	1,681	11,128	0,371	0,091	5,76	
Amostra18	0,42	2,26	1,681	11,999	0,353	0,081	5,91	
Amostra19	0,421	2,158	1,65	11,782	0,341	0,053	5,86	
Amostra20	0,42	2,517	1,876	9,035	0,253	0	5,49	
Amostra21	0,419	2,247	1,683	8,537	0,257	0	5	
Amostra22	0,418	2,556	1,827	15,161	0,395	0,137	6,81	
Amostra23	0,421	2,473	1,804	15,211	0,383	0,24	5,8	
Amostra24	0,419	2,353	1,701	12,351	0,312	0,023	5,97	
Amostra25	0,418	2,496	1,745	12,891	0,325	0	6,25	

Figura 1S. Tela principal do programa QSAR modeling

Validação cruzada

Número de variáveis latentes:

Número de amostras a retirar na validação cruzada:

Figura 2S. Janela do programa QSAR modeling na qual o usuário escolhe o número máximo de variáveis latentes e o número de amostras a serem removidas durante a validação cruzada

OPS

Número de variáveis latentes para o algoritmo OPS:

Número de variáveis latentes para o modelo:

Número de amostras a retirar na validação cruzada:

Janela: Incremento:

Porcentagem de variáveis:

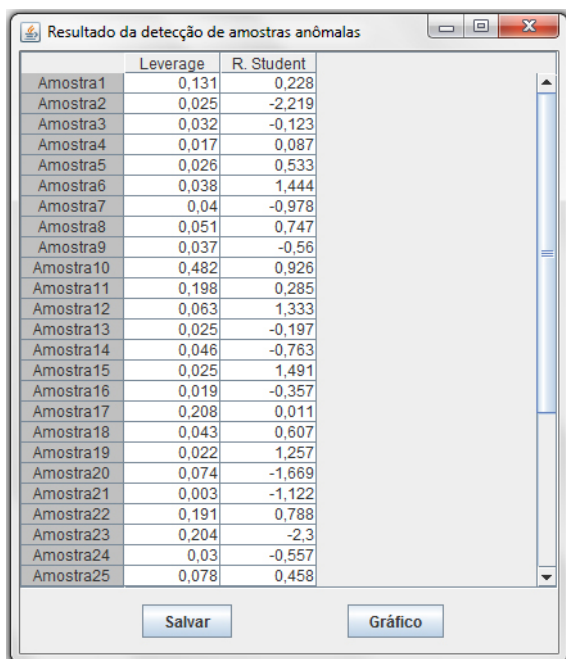
Vetor

- Correlograma
- Vetor de regressão
- Produto

Critério para classificar modelos

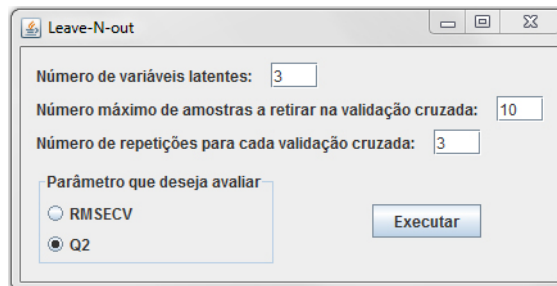
- RMSECV
- Rcv
- Q2
- Spres

Figura 3S. Janela do programa QSAR modeling na qual o usuário escolhe as opções de execução do algoritmo OPS



	Leverage	R. Student
Amostra1	0,131	0,228
Amostra2	0,025	-2,219
Amostra3	0,032	-0,123
Amostra4	0,017	0,087
Amostra5	0,026	0,533
Amostra6	0,038	1,444
Amostra7	0,04	-0,978
Amostra8	0,051	0,747
Amostra9	0,037	-0,56
Amostra10	0,482	0,926
Amostra11	0,198	0,285
Amostra12	0,063	1,333
Amostra13	0,025	-0,197
Amostra14	0,046	-0,763
Amostra15	0,025	1,491
Amostra16	0,019	-0,357
Amostra17	0,208	0,011
Amostra18	0,043	0,607
Amostra19	0,022	1,257
Amostra20	0,074	-1,669
Amostra21	0,003	-1,122
Amostra22	0,191	0,788
Amostra23	0,204	-2,3
Amostra24	0,03	-0,557
Amostra25	0,078	0,458

Figura 4S. Resultado da detecção de amostras anômalas mostrando os valores de influência e dos resíduos de Student para os compostos do conjunto de treinamento



Leave-N-out

Número de variáveis latentes: 3

Número máximo de amostras a retirar na validação cruzada: 10

Número de repetições para cada validação cruzada: 3

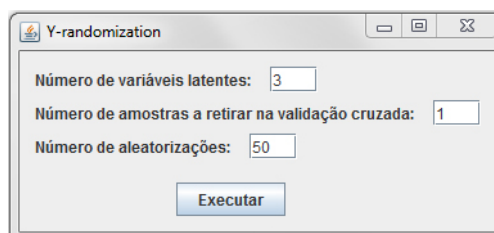
Parâmetro que deseja avaliar:

RMSECV

Q2

Executar

Figura 5S. Procedimento de validação leave-N-out para garantir a robustez de um modelo usando o programa QSAR modeling



Y-randomization

Número de variáveis latentes: 3

Número de amostras a retirar na validação cruzada: 1

Número de aleatorizações: 50

Executar

Figura 6S. Procedimento de validação de aleatorização de y para verificar a correlação ao acaso de um modelo usando o programa QSAR modeling