



UNIVERSITAT^{DE}
BARCELONA

Aplicació de la metodologia QSPR al càlcul de propietats de compostos inorgànics i de sistemes multicomponents

Jesús Jover Modrego



Aquesta tesi doctoral està subjecta a la llicència **Reconeixement 4.0. Espanya de Creative Commons.**

Esta tesis doctoral está sujeta a la licencia **Reconocimiento 4.0. España de Creative Commons.**

This doctoral thesis is licensed under the **Creative Commons Attribution 4.0. Spain License.**

TD Jov



Departament de Química Inorgànica

**Aplicació de la metodologia QSPR al càlcul
de propietats de compostos inorgànics i de
sistemes multicomponents**

Jesús Jover Modrego

Tesi Doctoral

BIBLIOTECA DE LA UNIVERSITAT DE BARCELONA



0700561294

A large, stylized handwritten signature in black ink is positioned at the bottom right of the page, overlapping the library information.

Universitat de Barcelona
Departament de Química Inorgànica

Programa de Doctorat de Química Inorgànica
Bienni 2003/2005

**Aplicació de la metodologia QSPR al càlcul
de propietats de compostos inorgànics i de
sistemes multicomponents**

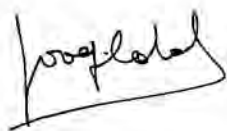
Memòria presentada per en Jesús Jover Modrego
per optar al Títol de Doctor per la Universitat de Barcelona

Joaquim Sales Cabré. Catedràtic del Departament de Química Inorgànica de la Universitat de Barcelona.

Ramón Bosque Pueyo. Professor titular del Departament de Química Inorgànica de la Universitat de Barcelona.

Certifiquen: Que el treball titulat “Aplicació de la metodologia QSPR al càlcul de propietats de compostos inorgànics i de sistemes multicomponents” que presenta Jesús Jover Modrego, per aspirar al grau de doctor per la Universitat de Barcelona, ha estat realitzat sota la seva direcció en el Departament de Química Inorgànica de la Universitat de Barcelona.

Barcelona, Desembre de 2007.



Dr. Joaquim Sales Cabré



Dr. Ramón Bosque Pueyo

La realització d'aquesta Tesi Doctoral ha estat possible gràcies a la concessió d'una Beca de Formació en la Recerca i la Docència (BRD) de la Universitat de Barcelona (2003/2007).

Fa més de cinc anys que vaig aterrar en aquest departament, en tot aquest temps m'han passat un munt de coses i he conegut un gran nombre de persones que, d'una manera o una altra m'han marcat profundament. Aquest petit espai ha de servir com a agraïment a totes elles i, també, per a totes les que de segur estaré oblidant, encara que no sigui voluntàriament.

En primer lloc, als meus directors de Tesi, Joaquim Sales i Ramón Bosque. Per haver-me ensenyat que no tota la Química es fa en els laboratoris, per anar sempre un pas més enllà del que jo hauria estat capaç, per la resolució de dubtes, el programes i mil coses més que no caben en aquestes pàgines.

Als meus companys fixos de "camarote". Al Jordi Cirera per TOT, és a dir, els moments d'humor compartits, els concerts, el Mario Kart, les sortides però, sobretot, per ser un bon amic. Al David Casanova per aguantar-nos a l'Avi i a mi quan ens posàvem en plan "salvaje", pels seus apunts humorístics inesperats, per les consoles i veduretes (hem de fer alguna més) i també per les seves deduccions matemàtiques increïbles. Si és cert això que diuen, que gaudir fa viure més anys, nosaltres arribarem als cent.

Als altres companys que es van traslladar quan van poder (no sé si per culpa dels que ens vam quedar). Al Thomas Cauchy per les classes de francès-chicano, els balls regionals i les crepes (sobretot això últim). Al Javi Tercero per ser capaç se sumar-se a nosaltres.

A los "niños", Toni, Benji, Jorge i Edu, perquè en vosaltres veig des de fora el reflex del que per a mi han estat aquests anys, per la mítica organització de la tesi d'en Tommy i per tots els bons moments que m'han ajudat a acabar la meua. A la meua nova companya de despatx, la Sílvia, que ha aparegut quan ja portava quatre mesos sol al despatx (començava a sentir-me com un profe, amb despatx propi) per fer-me veure que sempre és millor estar acompanyat.

A tot el grup d'Estructura Electrònica per demostrar que ser un "membre annex" té molts avantatges, com que et convidin als aniversaris on l'Elisco posa els pastissos. Al Joan Cano per les visites al despatx i per ser, de tant en tant, com el prof. Z.

Al grup de Materials Catalítics, al Narcís Homs i la Pilar Ramírez de la Piscina, per acollir-me en les seves files (estic segur que aquesta dada ha agafat desprevingut a més d'un).

To all the people of the ITODYS lab in Paris with a special memory to Prof. Bo-Tao Fan (or just Fan as you said). A Cátia por ayudarme mientras estuve allí, por venir a verme aquí, por el día de la música (que llovió) y por acompañarme a ver la final de la Champions que el Barça ganó en París. I want to thank also Joseph, Florent, Lina, Ana and Cyril.

A mi familia, por haberme ayudado a llegar hasta aquí, por hacerme ser la persona que soy y por tantas otras cosas que, si quisiera ponerlas por escrito, abarcarían más espacio que el que ocupa esta Tesis entera.

A Nuria, por ser capaz de aguantarme y de reir conmigo, por compartir su vida con la mía y por hacerme ver que, a veces, yo también me equivoco. Espero que siempre estés a mi lado para apoyarme, echarme la bronca o lo que toque en ese momento.

A la familia política, porque se diga lo que se diga, ellos también son importantes, Antonio, Nieves, Fuen y Ernesto. Un agradecimiento especial para los miembros de la familia que han cambiado mi manera de ver el mundo y que aún no saben leer: Àlex, Marc i Alba (espero que cuando seáis mayores os leáis este tocho que escribió vuestro tío).

A tota la gent del departament de Química Inorgànica per totes les coses que heu compartit amb mi, les pràctiques, els sopars de nadal i els partits de futbol (¡estúpida rodilla!). Un agraïment molt especial per a l'Òscar i el Fran d'Estat Sòlid, per totes les coses que hem passat junts.

"Es simple. La estupidez real siempre vence a la inteligencia artificial."

Mustrum Ridcully en *Papá Puerco*, Terry Pratchett

"Lo que llamamos "casualidad" no es más que la ignorancia de las causas físicas."

Gottfried Wilhelm Leibniz.

"¿Quería matarme? Bajo esta capa no queda carne ni sangre que se pueda matar.
Sólo hay una idea, y las ideas son a prueba de balas."

V de Vendetta, Alan Moore y David Lloyd.

Índex

1. Introducció	1
1.1. Antecedents	6
1.2. Objectius	12
1.3. Bibliografia	17
2. Fonaments i metodologia	25
2.1. Esquema general de treball	27
2.1.1. Dades d'entrada	27
2.1.2. Optimització de la geometria	28
2.1.3. Càlcul de descriptors	29
2.1.4. Establiment dels models	34
2.1.5. Validacions	40
2.2 Annexos	
Annex 1. Criteris estadístics	44
Annex 2. Selecció de descriptors i construcció de models	51
2.3. Bibliografia	79
3. Compostos inorgànics	85
3.1 Viscositat de compostos organometàl·lics	89
3.1.1 Viscositat a 20°C	92
Model lineal	93
Model no lineal: Xarxes neuronals BPNN	96
Model no lineal: <i>Support Vector Machines</i>	100
3.1.2. Viscositat a temperatura variable	110
Model lineal	111
Model no lineal: Xarxes neuronals BPNN	114
Model no lineal: <i>Support Vector Machines</i>	119
3.2 Tensió superficial de compostos organometàl·lics	135
3.2.1. Tensió superficial a 20°C	138
Model lineal	139
Model no lineal: Xarxes neuronals BPNN	142
Model no lineal: <i>Support Vector Machines</i>	145

3.2.2. Tensió superficial a temperatura variable	155
Model lineal	156
Model no lineal: Xarxes neuronals BPNN	159
Model no lineal: <i>Support Vector Machines</i>	163
3.3 Entalpia de formació de compostos organometàl·lics	179
Compostos organometàl·lics	182
Compostos orgànics	189
3.4. Basicitat catiònica de compostos orgànics front liti	202
Model lineal	204
Models no lineals	210
3.5. Bibliografia	221
4. Sistemes multicomponents	227
4.1. Afinitat i basicitat catiòniques dels aminoàcids habituals	232
4.1.1. Afinitat i basicitat amb càlculs teòrics	234
4.1.2. Afinitat amb la metodologia QSPR	267
4.2. Constant d'acidesa de compostos orgànics en diferents solvents	277
4.2.1. pK_a de fenols	284
Model no lineal: Xarxes neuronals BPNN	285
Model no lineal: <i>Support Vector Machines</i>	291
4.2.2. pK_a d'àcids benzoics	300
Model no lineal: Xarxes neuronals BPNN	301
Model no lineal: <i>Support Vector Machines</i>	312
4.2.3. Anàlisi conjunta de fenols i àcids benzoics	326
4.2.4. pK_a d'àcids carboxílics alifàtics	349
Model no lineal: Xarxes neuronals BPNN	350
Model no lineal: <i>Support Vector Machines</i>	356
4.2.5. pK_b d'anilines	364
Model no lineal: Xarxes neuronals BPNN	366
Model no lineal: <i>Support Vector Machines</i>	373
4.3. Bibliografia	379
5. Conclusions i resum	389

Capítol 1. Introducció

La Química pot definir-se, d'una manera planera, com la ciència que s'ocupa dels compostos, les seves propietats i les seves transformacions. El coneixement que es té avui dia de la naturalesa íntima de la matèria és, encara, molt limitat, i si bé és cert que l'estructura, l'energia i moltes altres propietats d'un compost poden descriure's per l'equació de Schrödinger, aquesta equació només es pot resoldre de manera exacta per a sistemes molt senzills. En la seva aplicació amb caràcter general cal recórrer a solucions aproximades que, gràcies als poderosos mètodes de càlcul i als no menys potents ordinadors actuals, són cada cop més precises.

A pesar d'aquestes limitacions i de la complexitat pròpia de la Química, el desenvolupament d'aquesta ciència ha estat i continua sent extraordinari. La raó d'aquest èxit ha estat la capacitat d'aprendre dels fets experimentals. Els químics han estat capaços de deduir models a partir dels resultats experimentals, han comprovat la seva bondat amb nous experiments especialment dissenyats i, finalment, han deduït comportaments generals i proposat lleis capaces de predir l'activitat dels compostos químics ja siguin nous o prou ben coneguts. Aquest mètode inductiu de desenvolupament ha estat emprat des del bell inici de la Química com a ciència. Una de les principals ocupacions dels químics es preparar compostos amb unes determinades propietats i, a poder ser, relacionar-les amb la seva estructura. Quant millor sigui el coneixement d'aquesta relació més fàcil serà preparar nous compostos amb noves propietats.

Una aproximació que s'ocupa d'aprofundir en aquests lligams i d'expressar-ho de manera mesurable és la metodologia anomenada "Relacions estructura-propietat quantitatives", més coneguda per l'acrònim en anglès QSPR (*Quantitative Structure-Property Relationships*). Aquests mètodes tenen una llarga tradició i el més popular és el QSAR (*Quantitative Structure-Activity Relationships*) que correspon al cas particular en què la propietat estudiada és l'activitat biològica i per tant el seu desenvolupament ha estat molt lligat al disseny de nous fàrmacs.

La metodologia QSPR, tal i com es representa en la Figura 1.1, proposa l'existència d'una relació directa entre l'estructura d'un compost i les seves propietats físiques, químiques, biològiques, etc. Partint d'aquesta premissa es pot dir que si es coneix l'estructura d'un compost químic se'n poden estimar les propietats.

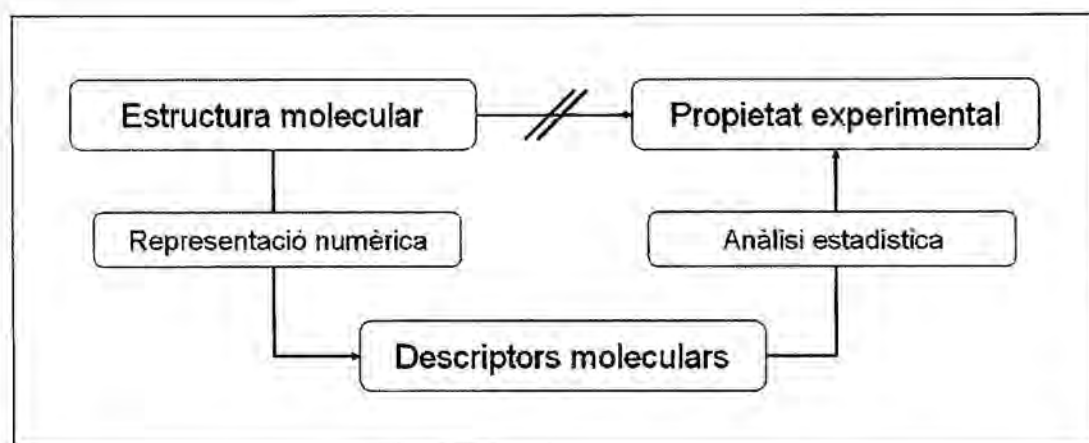


Figura 1.1. Esquema de la metodologia QSPR.

El problema d'aquest plantejament apareix a l'hora d'establir la relació directa entre l'estructura i la propietat estudiada; de vegades aquesta relació pot ser força evident però, en general, no és així. L'objectiu principal de la metodologia QSPR és establir un camí alternatiu que permeti relacionar l'estructura d'un compost amb les seves propietats. Aquest camí alternatiu consisteix a buscar, a través de l'anàlisi estadística, una relació matemàtica, que pot ser lineal o no, entre la propietat estudiada i un o més paràmetres, anomenats descriptors, derivats de l'estructura molecular dels compostos químics; és a dir, l'objectiu últim de la metodologia QSPR consisteix a establir una relació del tipus:

$$propietat = f(descriptors)$$

Aquests descriptors moleculars que s'extreuen directament de l'estructura són, simplement, el resultat d'un procediment matemàtic que transforma la informació química continguda en la representació simbòlica d'una molècula en un nombre útil. El terme útil d'aquesta definició s'ha d'entendre en els dos sentits, ser útil en la

interpretació de la propietat estudiada i/o ser útil en la predicció d'aquesta propietat per a altres molècules, que no sempre hauran estat sintetitzades prèviament. Si la relació matemàtica que s'estableix, també anomenada model, és acceptable pot servir no tan sols per estimar la propietat de compostos anàlegs, sinó també per suggerir, o fins i tot proposar mecanismes pels quals la propietat estudiada es relaciona amb l'estructura molecular.

A tall de presentació, en la Figura 1.2 es mostra un esquema amb els diferents passos que conformen l'aplicació de la metodologia QSPR a un estudi concret.

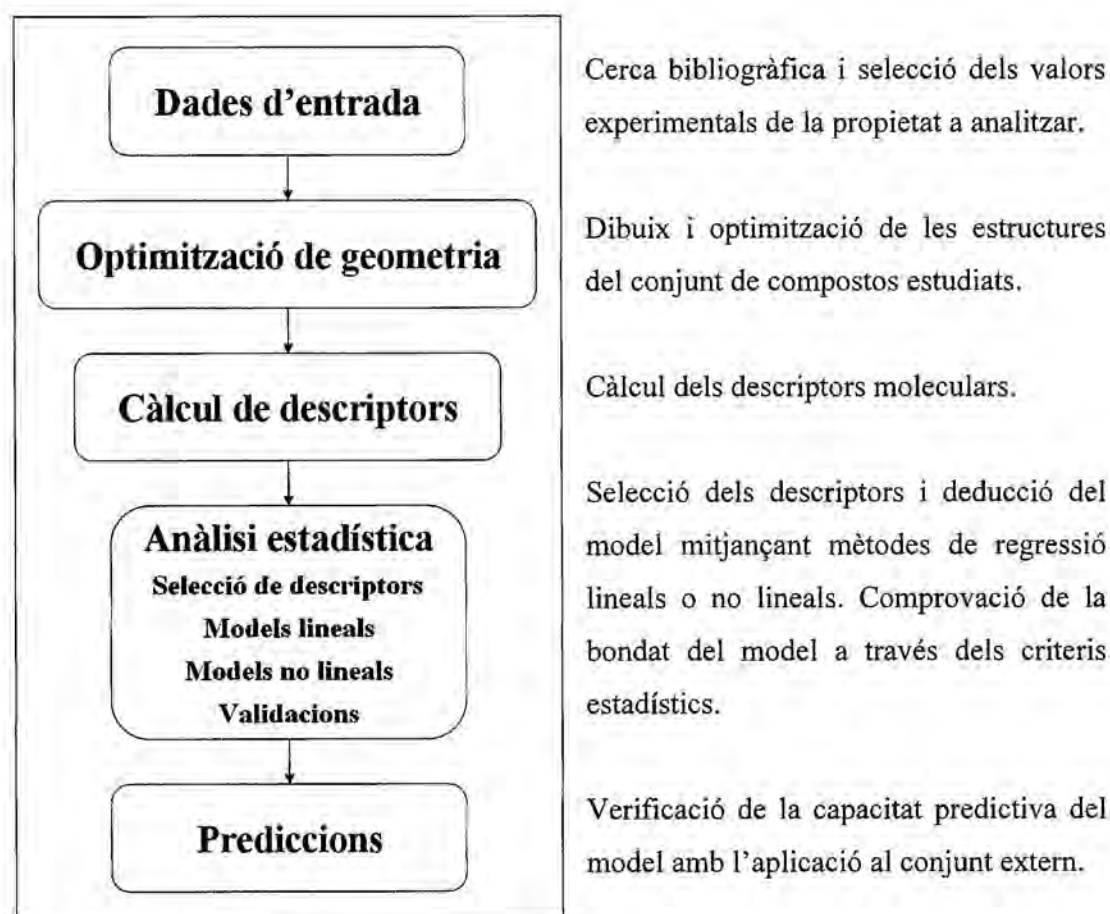


Figura 1.2. Passos d'un estudi QSPR.

1.1. Antecedents

La metodologia QSAR/QSPR no té un punt de partida històric clar.¹⁻⁶ Es creu que les seves arrels es remunten gairebé a fa més d'un segle; el primer registre conegut on es relaciona una propietat biològica amb l'estructura dels compostos químics es troba en la tesi doctoral de A. F. A. Cross, presentada a la Universitat d'Estrasburg, l'any 1863. En aquest estudi es descriu l'increment de la toxicitat de diferents alcohols en mamífers amb la disminució de la solubilitat d'aquests compostos en aigua.⁷ L'any 1868 apareix el que es considera el primer treball de QSAR, Crum Brown i Fraser van estudiar l'efecte biològic d'alguns alcaloides, abans i després de metilar el nitrogen bàsic.⁸ Les pronunciades diferències entre els compostos bàsics i els compostos quaternaris carregats van fer pensar que l'activitat fisiològica dels alcaloides Φ ha de ser funció de la seva constitució química C .

$$\phi = f(C)$$

Gairebé al mateix temps, Richardson va demostrar que l'activitat narcòtica dels alcohols és proporcional al seu pes molecular.⁹ Posteriorment, l'any 1893, Richet va observar que la toxicitat de diversos compostos químics: èters, aldehids, cetones i d'altres, es relaciona inversament amb la seva solubilitat en aigua: "*Plus ils sont solubles, moins ils sont toxiques*"; com més solubles, menys tòxics són.¹⁰ Altres teories generals d'aquest mateix procediment de narcosi van ser formulades independentment per Meyer^{11,12} i Overton^{13,14} a inicis del segle XX. Totes dues teories proposen que la toxicitat dels compostos orgànics, en forma neutra, depèn de la seva capacitat de partició entre l'aigua i una altra biofase lipòfila, on s'exerceix la seva activitat biològica. El model emprat en aquests estudis és el sistema aigua/oli d'oliva.

Més o menys en aquesta mateixa època es van desenvolupar les primeres hipòtesis sobre la naturalesa de les interaccions específiques. L'any 1894, Fischer va estudiar el trencament heterolític de diversos glicòsids amb diferents enzims. A partir de l'especificitat observada va concloure: "*Um ein Bild zu gebrauchen, will ich sagen, dass Enzym und Glucosid wie Schloss und Schlüssel zueinander passen müssen, um eine chemische Wirkung aufeinander ausüben zu können*"; que ve a dir: l'enzim i el glicòsid

han d'encaixar, com un pany i una clau, per tal d'exercir algun efecte químic l'un sobre l'altre.¹⁵⁻¹⁷ Alguns anys més tard, Ehrlich va formular: "*Corpora non agunt nisi fixata*"; les substàncies no actuen si no es troben enllaçades.¹⁸

El desenvolupament de la metodologia QSAR/QSPR va romandre aturat durant els següents 40-50 anys. Les úniques contribucions van ser realitzades per Fúhner i Neubauer, en estudis d'increments de l'activitat narcòtica de sèries homòlogues de compostos; i Ferguson, que va proporcionar una interpretació termodinàmica per a les relacions no específiques entre lipofília i activitat.¹⁻³ Mentrestant, Hammet¹⁹ va proposar, als anys 30, una de les majors contribucions a la química orgànica teòrica: utilitzant les constants d'ionització d'àcids benzoics substituïts com a sistema de referència, va postular que les constant d'equilibri (K) dels compostos aromàtics es podien descriure amb la següent equació:

$$\log K_{R-X} - \log K_{R-H} = \rho\sigma$$

on ρ és una constant de reacció específica per a cada reacció i σ , l'anomenada constant de Hammet, és un paràmetre que depèn només de la capacitat donadora del substituent estudiat. Naturalment, es va intentar traslladar aquestes equacions a sistemes biològics; correlacionant propietats biològiques amb els valors de les constants de Hammet²⁰ o bé definint "equacions biològiques de Hammet".²¹⁻²³

De tota manera, cap a principis dels anys seixanta, la metodologia QSAR/QSPR semblava haver arribat a un atzucac, i no va ser fins a l'any 1964, quan els estudis Free-Wilson i Hansch-Fujita van provocar el següent pas endavant. El model desenvolupat per Free i Wilson²⁴ diu que tots els valors d'activitat biològica són la suma, en escala logarítmica, de l'activitat d'un compost de referència i de les contribucions de grup dels diferents substituents enllaçats a aquest compost. Aquest tipus de model és molt senzill d'utilitzar ja que no són necessàries altres propietats fisicoquímiques dels compostos, només calen els valors 0 i 1 per indicar la presència d'un determinat substituent en una posició de la molècula.

El model desenvolupat per Hansch i Fujita: “ ρ - σ - π Analysis. A method for the Correlation of Biological Activity and Chemical Structure”²⁵ és considerat com el punt de partida del QSAR clàssic tal i com s'utilitza actualment. La base d'aquest model proposa combinar diversos paràmetres físicoquímics en una única equació multilinear, per exemple, l'equació següent és el model que permet calcular la dosi molar d'un compost que produeix una certa resposta biològica:²⁶

$$\log\left(\frac{1}{C}\right) = a\pi + b\sigma + \dots + ct.$$

on π i σ són els anomenats paràmetres d'hidrofòbia i electrònic de Hansch, respectivament. Els avantatges que presenta aquest model, anomenat anàlisi de Hansch (o LFER o relació extratermodinàmica), són la seva flexibilitat i la seva capacitat per descriure un gran ventall d'activitats biològiques diferents, especialment dades *in vitro* com les d'inhibició d'enzims.²⁷⁻²⁹

Tot just després de l'aparició d'aquest model, només es disposava d'uns pocs paràmetres: $\log P$ i π com a mesures de hidrofòbia, valors de f per a la lipofília, les escales σ , F i R per a propietats electròniques, E_s per a propietats estèriques i la refractivitat molar com a mesura de les forces de dispersió.³⁰ Tot i així, ben aviat es van desenvolupar multitud d'altres descriptors.³¹

A finals dels anys 70 va néixer el 3D QSAR, que intenta relacionar les propietats de les molècules mitjançant camps tridimensionals calculats dintre d'una reixeta regular.³² Més endavant, aquesta metodologia va rebre el nom DYLOMMS (*Dynamic Lattice-Oriented Molecular Modeling System*) però no va ser fins a l'any 1988 quan es van desenvolupar un gran nombre de correlacions entre els valors dels descriptors calculats a partir dels camps tridimensionals i diverses activitats biològiques. En aquest moment, aquesta metodologia passa a anomenar-se CoMFA (*Comparative Molecular Field Analysis*)³³ i es converteix en un dels mètodes més utilitzats actualment per establir models quantitius. Després d'aquest punt es van desenvolupar altres metodologies 3D QSAR: CoMSIA (*Comparative Molecular Similarity Indices Analysis*),³⁴ CoMMA (*Comparative Molecular Moment Analysis*),³⁵ Compass,³⁶ MS-WHIM,³⁷ E-State³⁸ i PARM.³⁹

En els últims anys, la química combinatòria i la química computacional s'han unit per produir els mètodes de *screening* amb els quals és possible, a partir d'un farmacòfor seleccionat, buscar fàrmacs potencials entre un gran nombre de compostos inclosos en diferents bases de dades.^{40,41} Diversos models han estat desenvolupats seguint aquesta metodologia per tal de predir la biodisponibilitat, la distribució, la capacitat de metabolització per diferents citocroms, l'eliminació, la penetració a través de les barreres del cervell i la placenta, la toxicitat i la capacitat cancerígena.⁴²⁻⁵²

Fins a aquest moment només s'ha parlat de la part més QSAR de la metodologia QSAR/QSPR, és important recordar que la seva evolució ha estat fortament lligada al disseny i desenvolupament de nous fàrmacs, però no s'ha de perdre de vista que també és possible utilitzar-la per predir altres propietats físiques o químiques dels compostos, aspecte que es presenta sota el nom de QSPR; no cal dir que aquesta és la vessant que s'ha estudiat en aquest treball. La metodologia QSPR ha estat aplicada àmpliament en l'estimació de gran quantitat de propietats físiques i químiques de compostos orgànics. Recentment s'han publicat unes revisions bastant completes sobre el seu ús en propietats físiques amb importància tecnològica.⁵³⁻⁵⁵ En aquests treballs s'ordenen les aplicacions segons si analitzen propietats que depenen de molècules individuals, de la interacció entre molècules i propietats biològiques. A més dels conjunts formats per compostos moleculars, hi ha referències a l'anàlisi de polímers.

El primer intent de correlacionar una propietat experimental amb l'estructura dels compostos químics data del 1925 quan Langmuir va intentar establir una relació entre les interaccions intermoleculars que es donen en un líquid i la seva energia superficial.⁵⁶ A partir d'aquell moment, un gran nombre de propietats que depenen de molècules aïllades, és a dir, de les interaccions intermoleculars que es donen en els compostos químics, van començar a ser estudiades mitjançant la metodologia QSPR. El punt d'ebullició dels compostos químics ha estat una propietat molt estudiada des dels inicis d'aquesta metodologia, ja el 1947 Wiener, en va fer un.⁵⁷ En general, la majoria dels treballs publicats treballen amb conjunts petits de compostos anàlegs. No és fins als anys 90, paral·lelament a la generalització dels ordinadors, quan es comencen a desenvolupar models lineals i no lineals que permeten fer estimacions d'aquesta propietat per a conjunts de compostos grans i de naturalesa química més variada.⁵⁸⁻⁶¹ Els

resultats que s'obtenen dels models construïts solen ser de gran qualitat, amb errors al voltant dels 8-12 K.

Una altra propietat d'aquest mateix tipus és el punt de fusió. En aquest cas, la majoria dels models construïts són de tipus lineal.^{62,63} Un dels millors models desenvolupats ha estat construït per a un conjunt de 366 alcans.⁶⁴ Els resultats obtinguts amb aquest model lineal d'onze descriptors són prou bons amb uns errors propers a 17 K.

La temperatura crítica és una altra propietat que depèn de les interaccions intermoleculars dels compostos químics. S'han trobat diversos treballs publicats en els quals s'utilitza la metodologia QSPR per calcular-la.⁶⁵⁻⁶⁸ El grup de Jurs ha establert models de vuit descriptors lineals i no lineals- amb xarxes neuronals-, per estimar aquesta propietat per a un conjunt de més de 150 compostos orgànics.^{59,69}

Altres propietats d'aquest mateix tipus que han estat analitzades amb la metodologia QSPR són la pressió de vapor,⁷⁰⁻⁷² la densitat,⁷³⁻⁷⁵ l'índex de refracció,^{73,74,76} el punt d'inflamació⁷⁷ i la temperatura d'autoignició.⁷⁸⁻⁸⁰

Existeix un gran nombre de propietats que, tot i pertànyer a molècules aïllades, depenen fortament de les interaccions que s'estableixen entre dues espècies químiques. El seu estudi s'ha centrat majoritàriament en compostos amb aplicacions industrials, com poden ser els surfactants i els polímers. Una de les propietats destacades dels surfactants és la concentració micel·lar crítica, que ha estat tractada amb diferents models lineals.^{81,82} Una altra propietat d'aquest mateix tipus de compostos és el punt de núvol, que es defineix com la temperatura en la qual una solució monofàsica micel·lar es converteix en bifàsica. Aquesta propietat s'ha estimat tant amb models lineals com amb no lineals.^{83,84}

Les propietats més estudiades per a famílies de polímers són la temperatura de transició vítria,⁸⁵⁻⁸⁸ l'índex de refracció,⁸⁹ la velocitat de polimerització radicalària⁹⁰ i l'acceleració del procés de vulcanització del cautxú.⁹¹

Les propietats relacionades amb els processos que es donen en solució són un grup de gran importància i també àmpliament estudiat. Una de les més destacades és el coeficient de partició octanol/aigua (*logP*), molt important en medicina i toxicologia ja que se sol utilitzar com a mesura d'hidrofobicitat dels compostos químics. La metodologia QSPR ha estat molt emprada per estimar-la tant amb models lineals com no lineals.⁹²⁻¹⁰⁰

Una altra propietat d'aquest tipus, que presenta gran interès i també gran complexitat, és la solubilitat dels compostos químics. La solubilitat aquosa de sòlids i líquids ha estat àmpliament estudiada per diversos mètodes, entre els que s'inclou la metodologia QSPR.¹⁰¹⁻¹¹⁰ Algun dels treballs publicats presenta resultats espectaculars, amb coeficients de determinació (R^2) majors que 0.98 i errors semblants als de les mesures experimentals.¹⁰⁶ La solubilitat aquosa de gasos també ha estat força estudiada.¹¹¹⁻¹¹⁴ Una aportació molt destacable és el model desenvolupat per Abraham amb els seus cinc paràmetres solvatocròmics per mesurar la solubilitat de 408 gasos diferents en aigua, amb el qual s'obtenen resultats molt acurats.¹¹⁵

Altres propietats que es relacionen amb els processos en solució, i que també han estat avaluades per la metodologia QSPR, són les escales de polaritat dels solvents¹¹⁶⁻¹¹⁹ i paràmetres de cromatografia de gasos i líquids.¹²⁰⁻¹²⁴

Altres camps en els que s'ha aplicat la metodologia QSPR són ser la mesura de velocitats de reacció,¹²⁵ predicció de constants de RMN de diferents nuclis,¹²⁶ i propietats de cristalls líquids.¹²⁷

En la revisió més recent⁵⁵ es comenta la necessitat de trobar una manera d'aplicar aquest tipus de metodologia a sistemes més complicats: sistemes multicomponents o mesclades ja que, fins al dia d'avui, aquest tipus d'estudi s'ha mostrat pràcticament inaccessible. De la mateixa manera, l'ús d'aquesta metodologia en compostos inorgànics es troba molt restringida. Els estudis més rellevants en aquest camp han estat realitzats per Cundari, establint models lineals i no lineals que permeten l'estimació de diverses propietats: distància d'enllaç, freqüència de vibració i energia de dissociació, de molècules inorgàniques diatòmiques.^{128,129} Alguns treballs han estudiat algunes propietats atòmiques, com ara els potencials d'ionització en funció de característiques estructurals senzilles: electrons de valència en cada capa i càrrega.¹³⁰

Més propers en el temps són alguns estudis amb compostos organometàl·lics. Així, es troben treballs en què es comencen a estudiar propietats interessants com la seva toxicitat¹³¹⁻¹³⁵ i el comportament com a catalitzadors homogenis en reaccions d'acoblaments de Heck,¹³⁶ *cross-coupling*,¹³⁷ metàtesi d'olefines¹³⁸ i catàlisi asimètrica.¹³⁹⁻¹⁴²

El grup de treball en què es s'ha fet aquesta Tesi ha aplicat aquesta metodologia a l'estudi del desplaçament químic en ressonància magnètica nuclear de ^{31}P de fosfines terciàries,¹⁴³ al càlcul de la polaritzabilitat de solvents orgànics,¹⁴⁴ a l'estimació d'energies d'enllaç O-H de fenols,¹⁴⁵ a la predicció del paràmetre de polaritat de soluts p relacionat amb els temps de retenció en cromatografia de líquids¹⁴⁶ (HPLC) i al càlcul dels paràmetres solvatocròmics d'Abraham.¹⁴⁷

1.2. Objectius

Els objectius proposats en la present Tesi poden resumir-se, d'una manera general, en l'extensió de la metodologia QSPR a sistemes en què fins ara la seva utilització havia estat molt limitada o inexistent. En concret, s'ha aplicat als dos tipus de sistemes següents:

-) Compostos inorgànics, que de fet són compostos de coordinació o compostos organometàl·lics, agafant el terme organometàl·lic en un sentit ampli.
-) Sistemes multicomponents, entesos com aquells casos en què la propietat experimental estudiada no depèn exclusivament de la naturalesa dels compostos que conformen el conjunt analitzat, sinó que depèn, també, d'altres elements presents, els quals en condicionen clarament el valor que té.

En l'estudi dels conjunts de compostos de coordinació i organometàl·lics, s'ha seguit el tractament usual, és a dir s'ha deduït el model a partir dels descriptors moleculars habituals; en canvi, l'anàlisi dels sistemes multicomponents requereix l'ús de descriptors externs. Per descriptors externs s'entenen aquells que caracteritzen els altres elements, els quals s'importen prèviament al programari emprat en la deducció del model final proposat.

En la deducció del model poden aplicar-se relacions lineals o bé no lineals. En general, en l'anàlisi dels sistemes habituals, és a dir quan la propietat depèn només dels compostos estudiats, inicialment es fa una regressió lineal múltiple i posteriorment s'hi apliquen mètodes no lineals, mitjançant xarxes neuronals, que sovint donen millors

estimacions. Amb els sistemes multicomponents, quan la propietat estudiada depèn de la naturalesa dels compostos analitzats i d'algun altre factor, la complexitat de les interaccions que es poden donar entre els diferents elements fixa l'ús d'aproximacions no lineals.

És evident que en el camp de la Química Inorgànica, constituïda en bona part per compostos sòlids amb un caràcter iònic important, no es fàcil disposar d'un conjunt de compostos moleculars d'estructura anàloga i dels quals se'n conegui, amb prou exactitud, una propietat física o química d'interès. Així doncs, la selecció de compostos organometàl·lics o de coordinació com a base de l'estudi aquí descrit és quasi obligada, ja que és en aquest tipus de derivats on s'hi poden trobar conjunts de compostos suficientment grans i amb estructures moleculars prou semblants que permetin l'ús de l'aproximació QSPR.

Pel que fa als conjunts de compostos organometàl·lics estudiats, s'ha de remarcar d'entrada que el terme organometàl·lic s'ha agafat en un sentit ampli, és a dir, s'hi inclouen derivats d'elements semimetàl·lics com bor, silici, germani, arsènic, tel·luri i fins i tot algun compost de fòsfor o de seleni. També s'han inclòs alguns derivats de fòsfor i estany en què els grups orgànics s'hi uneixen a través d'àtoms d'oxigen i nitrogen i que no contenen enllaços metall-carboni. Aquesta interpretació laxa del terme organometàl·lic ha permès disposar de conjunts de valors prou grans per poder aplicar-hi la metodologia QSPR.

Dels compostos organometàl·lics se n'ha estudiat la viscositat, η , la tensió superficial, σ , i l'entalpia de formació en fase gasosa, $\Delta_f H^p$. De la viscositat i la tensió superficial, les Taules Landolt-Börsntein forneixen d'un bon conjunt de dades experimentals ben contrastades que facilita l'anàlisi QSPR. A més, en disposar de valors a diferents temperatures, s'ha pogut ampliar l'anàlisi considerant la temperatura com una variable, i fer estimacions d'aquestes propietats en l'interval de temperatures estudiat. S'han analitzat conjunts bastant grans de derivats principalment de P, As, Si, Ge i Sn. D'altra banda, el fet que tant la viscositat com la tensió superficial de diferents conjunts de compostos orgànics hagués estat examinada per mètodes QSPR ha de permetre fer-ne el corresponent estudi comparatiu. La situació és ben diferent amb les entalpies de formació. Actualment, el nombre de valors experimentals disponibles no és gaire elevat i, a més, els valors descrits són bastant dispars segons els mètodes

experimentals emprats o els autors responsables de la determinació. En particular s'ha estudiat la $\Delta_f H^\circ$ en fase gasosa d'un conjunt de compostos organometàl·lics de fórmula general MR_nX_m , on M és un metall o semimetall dels grups 12 al 16 de la Taula periòdica, R és un grup orgànic tipus alquil, aril, etc. i X és Cl, Br, I o H. Tot i això, aquest primer intent mitjançant la metodologia QSPR, d'estimació d'una propietat tan fonamental com és la $\Delta_f H^\circ$, pot considerar-se prou reeixida; ja que si bé els errors no són petits, el model proposat permet computar-la a partir d'uns descriptors que es poden calcular fàcilment a partir de la fórmula dels compostos.

Una altra propietat estudiada ha estat la basicitat catiònica front el liti, LCB, d'un conjunt bastant nombrós de compostos orgànics, L, que contenen diferents grups funcionals com: alcohols, èters, cetones, àcids, amines, nitrils, etc., i fins i tot alguns aminoàcids, s'estudia, doncs, l'energia associada a la formació dels compostos de coordinació tipus $[LiL]^+$.

Pel que fa als anomenats sistemes multicomponents, l'estudi s'ha centrat en els següents processos:

-) l'afinitat i la basicitat catiòniques dels 20 aminoàcids habituals en fase gasosa
-) la constant d'acidesa d'un conjunt d'àcids orgànics en diferents solvents.

Els bons resultats obtinguts en l'estudi de la basicitat catiònica de l'extens conjunt de molècules orgàniques, van permetre el plantejament d'un tema més ampli com és el de les interaccions d'aminoàcids front cations monovalents. La cerca bibliogràfica pertinent va mostrar que no es disposa d'un conjunt de valors experimentals coherents de les corresponents afinitats i basicitats catiòniques, i que hi ha diferències significatives en els valors proposats per diferents autors, incoherències agreujades pel fet que en alguns casos els valors determinats experimentalment són valors relatius, és a dir, cal l'establiment d'algun valor per poder fixar els valors corresponents a la basicitat de cadascun dels aminoàcids estudiats. Paral·lelament, a la bibliografia es troben valors calculats per diferents mètodes, la major part d'ells d'exigència elevada com DFT i anàlegs, d'un bon grapat d'entalpies i energies lliures de reacció entre els aminoàcids i alguns cations. Per tal de poder aplicar els mètodes QSPR a un conjunt coherent de dades, s'ha portat a terme un estudi teòric detallat i sistemàtic d'aquestes magnituds termodinàmiques dels vint aminoàcids habituals amb els cations

Li^+ , Na^+ , K^+ , Cu^+ i Ag^+ . Aquests càlculs s'han fet amb DFT i bases rigoroses, i els valors d'afinitat i basicitat catióniques així calculats s'han pres com a valors de referència i s'hi ha aplicat la metodologia QSPR, amb bons resultats. En l'anàlisi d'aquest sistema multicomponent, a més dels descriptors moleculars habituals dels aminoàcids, deduïts després d'una acurada optimització de l'estructura, s'han considerat com a descriptors externs algunes propietats dels cations metàl·lics com són, entre altres, radis, relacions càrrega/radi, potencial d'ionització, electronegativitat segons diferents escales, etc. Els resultats obtinguts són bons, més tenint en compte la varietat i complexitat de les interaccions existents entre els aminoàcids i els cations metàl·lics en la formació dels corresponents $[\text{M}(\text{AA})]^+$.

L'altre gran grup de sistemes multicomponents abordat ha estat l'anàlisi de les constants de dissociació, $\text{p}K_a$, d'àcids orgànics en diferents solvents alhora. En aquest sentit s'han estudiat conjunts de fenols, àcids benzoics, àcids carboxílics alifàtics i anilines; aquests tipus de compostos són especialment adequats per a estudis QSPR atès el gran nombre i varietat de derivats que es coneixen. Aquests derivats que contenen distints grups substituents amb característiques electròniques i estèriques particulars en les posicions *orto*-, *meta*- i *para*- de l'anell benzènic, generen uns valors de $\text{p}K_a$ prou diferents, molt apropiat per al tipus d'estudi que aquí es vol portar a terme. El conjunt de solvents en què s'han determinat els $\text{p}K_a$ és prou variat i en conté de pròtics, com aigua, metanol, etanol, i d'apròtics com dimetilsulfòxid, acetonitril, dimetilformamida, acetona. Com a descriptors externs, per aportar el paper del solvent, s'han emprat algunes propietats físiques com, la constant dielèctrica, la polaritzabilitat, la densitat, el moment dipolar, l'entalpia de vaporització, així com els paràmetres de les escales empíriques de polaritat més conegudes, com són les de Kamlet i Taft, Reichardt, Koppel i Palm, Gutmann, Drago, entre d'altres. Els resultats obtinguts són, també, prou bons i els models deduïts contenen descriptors dels soluts i dels solvents que poden relacionar-se amb les complicades i subtils interaccions solut/solvent pròpies del procés de dissociació estudiat.

En general s'han deduït models que prediuen amb bons resultats les propietats analitzades i que contenen descriptors capaços, en la majoria dels casos, d'interpretar-les. Durant tot el procés de realització d'aquesta Tesi, i molt especialment en la part final de redacció, a l'hora d'analitzar, resumir i de posar en net tota la feina d'aquests quatre anys, mai no s'ha oblidat una antiga regla d'or QSPR:

Proposeu un model que:

*sigui el més simple i contingui quants menys descriptors millor
permeti una interpretació fisicoquímica de la propietat estudiada
demonstri un bon poder predictiu amb un conjunt de validació extern.*

Unger, S. H.; Hansch, C. *J. Med. Chem.* **1973**, 16, 745.

1.3. Bibliografia

1. Purcell, W. P.; Bass, G. E; Clayton, J. M. *Strategy of Drug Design. A Molecular Guide to Biological Activity*, Wiley: New York, **1973**.
2. Tute, M. S. *History and Objectives of Quantitative Drug Design*, Ramsden, C. A.; Ed.; *Quantitative Drug Design*, Volum 4; Hansch, C.; Sammes, P. G; Taylor, J. B.; Eds.; *Comprehensive Medicinal Chemistry. The Rational Design, Mechanistic Study & Therapeutic Application of Chemical Compounds*, Pergamon Press: Oxford, **1990**, 1.
3. Rekker, R. F. *Quant. Struct.-Act. Relat.* **1992**, 11, 195.
4. van de Waterbeemd, H. *Quant. Struct.-Act. Relat.* **1992**, 11, 200.
5. Kubinyi, H. *Quant. Struct.-Act. Relat.* **2002**, 21, 348.
6. Selassie, C. D. *History of Quantitative Structure-Activity Relationships Burger's Medicinal Chemistry and Drug Discovery*, 6th Ed., *Drug Discovery*: Volum 1, Abraham, D. J.; Ed; John Wiley & Sons, **2003**.
7. Cross, A. F. A. *PhD Thesis*, University of Strasbourg, **1863**. Citat de: Borman, S. *New QSAR Techniques Eyed for Environmental Assessments*, *Chem. & Eng. News*, **1990**, 20.
8. Crum Brown, A.; Fraser, T. R. *Trans. Roy. Soc. Edinburgh* **1868**, 25, 151.
9. Richardson, B. J. *Medical Times and Gazette* **1869**, 2, 703.
10. Richet, M. C. *Compt. Rend. Soc. Biol. (Paris)* **1893**, 45, 775.
11. Meyer, H. H. *Arch. Exp. Path. Pharm.* **1899**, 42, 109.
12. Lipnick, R. L. *Trends Pharmacol. Sci.* **1989**, 10, 265.
13. Overton, E. *Studien über die narkose, zugleich ein Beitrag zur allgemeinen Pharmakologie*, Fischer, G., **1901**. Traducció anglesa: Lipnick, R. L., *Studies on Narcosis, Charles Ernest Overton*, Chapman and Hall: London, **1991**.
14. Lipnick, R. L. *Trends Pharmacol. Sci.* **1986**, 7, 161.
15. Fischer, E. *Ber. Dtsch. Chem. Ges.* **1894**, 27, 2985.
16. Lichtenthaler, F. W. *Angew. Chemie* **1994**, 106, 2456; *Angew. Chem. Int. Ed. Engl.* **1994**, 33, 2364.
17. Folkers, G. *Pharm. Acta Helv.* **1995**, 69, 175.
18. Ehrlich, P. *Lancet II* **1913**, 445.

19. Hammett, L. P. *Physical Organic Chemistry. Reaction Rates, Equilibria and Mechanism*, 2nd Ed., McGraw-Hill: New York, **1970**.
20. Hansen, O. R. *Acta Chem. Scand.* **1962**, 16, 1593.
21. Zahradnik, R.; Chvapil, M. *Experientia* **1960**, 16, 511.
22. Zahradnik, R. *Arch. Int. Pharmacodyn. Ther.* **1962**, 135, 311.
23. Zahradnik, R. *Experientia* **1962**, 18, 534.
24. Free, S. M. Jr.; Wilson, J. W. *J. Med. Chem.* **1964**, 7, 395.
25. Hansch, C.; Fujita, T. *J. Am. Chem. Soc.* **1964**, 86, 1616.
26. Hansch, C. *Acc. Chem. Res.* **1969**, 2, 232.
27. Martin, Y. C. *Quantitative Drug Design. A Critical Introduction (Medicinal Research Series, Volume 8)*, Marcel Dekker: New York, **1978**.
28. Kubinyi, H. *QSAR: Hansch Analysis and Related Approaches*, Volum 1, Mannhold, R.; Krosggaard-Larsen, P; Timmerman, H; Eds.; *Methods and Principles in Medicinal Chemistry*, VCH: Weinheim, **1993**.
29. Hansch C.; Leo, A. *Exploring QSAR. Fundamentals and Applications in Chemistry and Biology*, American Chemical Society: Washington, **1995**.
30. Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR. Hydrophobic, Electronic and Steric Constants*, American Chemical Society: Washington, **1995**.
31. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*, Volum 11 de: Mannhold, R.; Kubinyi, H.; Timmerman, H.; Eds.; *Methods and Principles in Medicinal Chemistry*, Wiley-VCH: Weinheim, **2000**.
32. Cramer, R. D. III; Milne, M. *The lattice model: a general paradigm for shaped-related structure/activity*, *Abstracts of Papers of the Am. Chem. Soc. Meeting*, Abril **1979**, comp. 44.
33. Cramer, R. D. III; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, 110, 5959.
34. Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem.* **1994**, 37, 4130.
35. Silverman, B. D.; Platt, D. E. *J. Med. Chem.* **1996**, 39, 2129.
36. Jain, A. N.; Koile, K.; Chapman, D. *J. Med. Chem.* **1994**, 37, 2315.
37. Bravi, G.; Gancia, E.; Mascagni, P.; Pegna, M.; Todeschini, R.; Zaliani, A. *J. Comput.-Aided Mol. Design* **1997**, 11, 79.
38. Kellogg, G. E.; Kier, L. B.; Gaillard, P.; Hall, L. H. *J. Comput.-Aided Mol. Design* **1996**, 10, 513.
39. Chen, H.; Zhou, J.; Xie, G. *J. Chem. Inf. Comput. Sci.* **1998**, 38, 243.
40. Ajay, M.; Walters, W. P.; Murcko, M. A. *J. Med. Chem.* **1998**, 41, 3314.

41. Sadowski, J.; Kubinyi, H. *J. Med. Chem.* **1998**, 41, 3325.
42. Navia, M. A.; Chaturvedi, P. R. *Drug Disc. Today* **1996**, 1, 179.
43. Wessel, M. D.; Jurs, P. C.; Tolan, J. W.; Muskal, S. M. *J. Chem. Inf. Comput. Sci.* **1998**, 38, 726.
44. Oprea, T. I.; Gottfries, J. *J. Mol. Graph. Model.* **1999**, 17, 261.
45. Lucò, J. M. *J. Chem. Inf. Comput. Sci.* **1999**, 39, 396.
46. Kelder, J.; Grootenhuis, P. D. J.; Bayada, D. M.; Delbressine, L. P. C.; Ploemen, J. *P. Pharm. Res.* **1999**, 16, 1514.
47. Ekins, S.; Waller, C. L.; Swaan, P. W.; Cruciani, G.; Wrighton, S. A.; Wikel, J. H. *J. Pharmacol. Toxicol. Methods* **2000**, 44, 251.
48. Andrews, C. W.; Bennett, L.; Yu, L. X. *Pharm. Res.* **2000**, 17, 639.
49. Yoshida, F.; Topliss, J. G. *J. Med. Chem.* **2000**, 43, 2575.
50. Egan W. J.; Merz, K. M. Jr.; Baldwin, J. J. *J. Med. Chem.* **2000**, 43, 3867.
51. Zhao, Y. H.; Le, J.; Abraham, M. H.; Hersey, A.; Eddershaw, P. J.; Luscombe, C. N.; Boutina, D.; Beck, G.; Sherborne, B.; Cooper, I.; Platts, J. A. B. *J. Pharm. Sci.* **2001**, 90, 749.
52. van de Waterbeemd, H.; Smith, D. A.; Beaumont, K.; Walker, D. K. *J. Med. Chem.* **2001**, 44, 1313.
53. Karelson, M.; Maran, U.; Wang, Y.; Katritzky, A. R. *Collect. Czech. Chem. Commun.* **1999**, 64, 1551.
54. Katritzky, A. R.; Maran, U.; Lobanov, V. S.; Karelson, M. *J. Chem. Inf. Comput. Sci.* **2000**, 40, 1.
55. Katritzky, A. R.; Fara, D. C. *Energy & Fuels* **2005**, 19, 922.
56. Langmuir, I. *Colloid Symp. Monogr.* **1925**, 3, 49.
57. Wiener, H. *J. Am. Chem. Soc.* **1947**, 69, 17.
58. Egolf, L. M.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **1993**, 33, 616.
59. Egolf, L. M.; Wessel, M. D.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **1994**, 34, 947.
60. Katritzky, A. R.; Mu, L.; Lobanov, V. S.; Karelson, M. *J. Phys. Chem.* **1996**, 100, 10400.
61. Katritzky, A. R.; Lobanov, V. S.; Karelson, M. *J. Chem. Inf. Comput. Sci.* **1998**, 38, 28.
62. Katritzky, A. R.; Maran, U.; Karelson, M.; Lobanov, V. S. *J. Chem. Inf. Comput. Sci.* **1997**, 37, 913.
63. Gramatica, P.; Navas, N.; Todeschini, R. *Chemometr. Intell. Lab.* **1998**, 40, 53.

64. Charton, M.; Charton, B. *J. Phys. Org. Chem.* **1994**, 7, 196.
65. Grigoras, S. A. *J. Comput. Chem.* **1990**, 11, 493.
66. Murray, J. S.; Lane, P.; Brinck, T.; Paulsen, K.; Grice, E.; Politzer, P. *J. Phys. Chem.* **1993**, 97, 9369.
67. Hall, L. H.; Story, C. T. *J. Chem. Inf. Comput. Sci.* **1993**, 36, 1004.
68. Katritzky, A. R.; Mu, L.; Karelson, M. *J. Chem. Inf. Comput. Sci.* **1998**, 38, 293.
69. Turner, B. E.; Costello, C. L.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **1998**, 38, 639.
70. Chastrette, M.; Cretin, D.; Tiyal, F. C. *R. Acad. Sci. II* **1994**, 318, 1059.
71. Basak, S. C.; Gute, B. D.; Grunwald, G. D. *J. Chem. Inf. Comput. Sci.* **1997**, 37, 651.
72. Liang, C. Gallagher, D. A. *J. Chem. Inf. Comput. Sci.* **1998**, 38, 321.
73. Gakh, A. A.; Gakh, E. G.; Stumper, B. G.; Noid, D. W. *J. Chem. Inf. Comput. Sci.* **1994**, 34, 832.
74. Liu, S.; Zhang, R.; Liu, M.; Hu, Z. *J. Chem. Inf. Comput. Sci.* **1997**, 37, 1146.
75. Karelson, M.; Perkson, A. *Comput. Chem.* **1999**, 23, 49.
76. Katritzky, A. R.; Sild, S.; Karelson, M. *J. Chem. Inf. Comput. Sci.* **1998**, 38, 840.
77. Katritzky, A. R.; Lobanov, V. S.; Karelson, M.; Murugan, R.; Grendze, M. P.; Toomey, J. E. Jr. *Rev. Roum. Chim.* **1996**, 41, 851.
78. Egolf, L. M.; Jurs, P. C. *Eng. Chem. Res.* **1992**, 31, 1798.
79. Tetteh, J.; Metcalfe, E.; Howells, S. L. *Chemometr. Intell. Lab.* **1996**, 32, 177.
80. Tetteh, J.; Howells, S. L.; Metcalfe, E.; Suzuki, T. *Chemometr. Intell. Lab.* **1998**, 41, 17.
81. Huibers, P. D. T.; Lobanov, V. S.; Katritzky, A. R.; Shah, D. O.; Karelson, M. *Langmuir* **1996**, 12, 1462.
82. Huibers, P. D. T.; Lobanov, V. S.; Katritzky, A. R.; Shah, D. O.; Karelson, M. *J. Colloid Interface Sci.* **1997**, 187, 113.
83. Huibers, P. D. T.; Shah, D. O.; Katritzky, A. R. *J. Colloid Interface Sci.* **1997**, 193, 132.
84. Bünz, A. P.; Braun, B.; Janowsky, R. *Ind. Eng. Chem. Res.* **1998**, 1075.
85. Joyce, S. J.; Osguthorpe, D. J.; Padgett, J. A.; Price, G. J. *J. Chem. Soc. Faraday Trans.* **1995**, 91, 2491.
86. Tan T. T. M.; Rode, B. M. *Macromol. Theor. Simulat.* **1996**, 5, 467.
87. Katritzky, A. R.; Sild, S.; Lobanov, V. S.; Karelson, M. *J. Chem. Inf. Comput. Sci.* **1998**, 38, 300.

88. Ulmer, C. W.; Smith, D. A.; Sumpter, B. G.; Noid, D. I. *Comput. Theor. Polym.* **1998**, *8*, 311.
89. Katritzky, A. R.; Sild, S.; Karelson, M. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 1171.
90. Ignatz-Hoover, F.; Petrukhin, R.; Karelson, M.; Katritzky, A. R. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 295.
91. Ignatz-Hoover, F.; Katritzky, A. R.; Lobanov, V. S.; Karelson, M. *Rubber Chem. Technol.* **1999**, *72*, 318.
92. Klopman, G.; Iroff, L. D. *J. Comput. Chem.* **1981**, *2*, 157.
93. Klopman, G.; Namboordiri, K.; Schochet, M. *J. Comput. Chem.* **1985**, *6*, 28.
94. Bodor, N.; Gabanyi, Z.; Wong, C.-K. *J. Am. Chem. Soc.* **1989**, *111*, 3783.
95. Kantola, A.; Villar, H. O.; Loew, G. H. *J. Comput. Chem.* **1991**, *12*, 681.
96. Bodor, N.; Huang, M. J. *J. Pharm. Sci.* **1992**, *81*, 272.
97. Grombar, V. K.; Enslein, K. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1127.
98. Bodor, N.; Buchwald, P. *J. Phys. Chem. B* **1997**, *101*, 3404.
99. Grunenberg, J.; Herges, R. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 905.
100. Breindl, A.; Beck, B.; Clark, T.; Glen, R. C. *J. Mol. Model.* **1997**, *3*, 142.
101. Patil, G. S. *Chemosphere* **1991**, *22*, 723.
102. Bodor, N.; Huang, M.-J. *J. Pharm. Sci.* **1992**, *81*, 954.
103. Patil, G. S. *J. Hazard. Mater.* **1994**, *36*, 35.
104. Nelson, T. M.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 601.
105. Sutter, J. M.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 100.
106. Huibers, P. H. T.; Katritzky, A. R. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 283.
107. Bodor, N.; Harget, A.; Huang, M. *J. Am. Chem. Soc.* **1991**, *113*, 9480.
108. Huuskonen, J.; Salo, J.; Taskinen, J. *J. Pharm. Sci.* **1997**, *86*, 450.
109. Huuskonen, J.; Salo, J.; Taskinen, J. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 450.
110. Mitchell, B. E.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 489.
111. Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. *J. Solution Chem.* **1981**, *10*, 563.
112. Nirmalakhandan, N. N.; Speece, R. E. *Environ. Sci. Technol.* **1988**, *22*, 1349.
113. Russell, C. J.; Dixon, S. L.; Jurs, P. C. *Anal. Chem.* **1992**, *64*, 1350.
114. Katritzky, A. R.; Mu, L.; Karelson, M. A. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1162.
115. Abraham, M. H.; Andonian-Haftvan, J.; Whiting, G. S.; Leo, A.; Taft, R. S. *J. Chem. Soc., Perkin Trans. 2* **1994**, 1777.
116. Katritzky, A. R.; Mu, L.; Karelson, M. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 756.

117. Mu, L.; Drago, R. S.; Richardson, D. E. *J. Chem. Soc., Perkin Trans. 2* **1998**, 159.
118. Katritzky, A. R.; Tamm, T.; Wang, Y.; Sild, S.; Karelson, M. *J. Chem. Inf. Comput. Sci.* **1999**, 39, 684.
119. Katritzky, A. R.; Tamm, T.; Wang, Y.; Karelson, M. *J. Chem. Inf. Comput. Sci.* **1999**, 39, 692.
120. Katritzky, A. R.; Ignatchenko, E. S.; Barcock, R. A.; Lobanov, V. S.; Karelson, M. *Anal. Chem.* **1994**, 66, 1799.
121. Lucic, B.; Trinajstić, N.; Sild, S.; Karelson, M.; Katritzky, A. R. *J. Chem. Inf. Comput. Sci.* **1999**, 39, 610.
122. Sutter, J. M.; Peterson, T. A.; Jurs, P. C. *Anal. Chim. Acta* **1997**, 342, 113.
123. Pompe, M.; Razinger, M.; Nović, M.; Veber, M. *Anal. Chim. Acta* **1997**, 348, 215.
124. Baczek, T.; Kalisz, R. *J. Chromatogr. A* **2003**, 987, 29.
125. Bakken, G.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **1999**, 39, 1064.
126. Clouser, D. L.; Jurs, P. C. *Anal. Chim. Acta* **1996**, 321, 127.
127. Johnson, S. R.; Jurs, P. C. *Chem. Mater.* **1999**, 11, 1007.
128. Cundari, T. R.; Moody, E. W. *J. Chem. Inf. Comput. Sci.* **1997**, 37, 871.
129. Cundari, T. R.; Moody, E. W. *J. Mol. Struct. (Theochem)* **1998**, 425, 43.
130. Sigman, M. E.; Rives, S. S. *J. Chem. Inf. Comput. Sci.* **1994**, 34, 617.
131. Eng, G.; Brinckman, F. E.; Olson, G. J.; Tierney, E. J.; Bellama, J. M. *App. Organometallic Chem.* **1991**, 5, 33.
132. Ownby, D. R.; Newman, M. C. *QSAR Comb. Sci.* **2003**, 22, 241.
133. Song, X.; Zapata, A.; Eng, G. *J. Organomet. Chem.* **2006**, 691, 1756.
134. Pavan, M.; Worth, A.P. EUROPEAN COMMISSION, DIRECTORATE GENERAL, JOINT RESEARCH CENTRE, EUR 22355 EN, **2006**.
135. Pavan, M.; Worth, A.P.; Netzeva, T. EUROPEAN COMMISSION, DIRECTORATE GENERAL, JOINT RESEARCH CENTRE, EUR 22327 EN, **2006**.
136. Burello, E.; Rothenberg, G. *Adv. Synth. Catal.* **2003**, 345, 1334.
137. Burello, E.; Farriseng, D.; Rothenberg, G. *Adv. Synth. Catal.* **2004**, 346, 1844.
138. Occhipinti, G.; Björsvik, H.-R.; Jensen, V. R. *J. Am. Chem. Soc.* **2006**, 128, 6952.
139. Prieto, O.; Woodward, S. *J. Organomet. Chem.* **2006**, 691, 1515.
140. Chen, J.; Jiwu, W.; Mingzong, L.; You, T. *Journal of Molecular Catalysis A: Chemical* **2006**, 258, 191.

141. Ianni, J. C.; Annamalai, V.; Phuan, P.-W.; Panda, M.; Kozlowski, M. C. *Angew. Chem. Int. Ed.* **2006**, *45*, 5502.
142. Chen, J.; Jiwu, W.; Daliang, L.; You, T. *J Mol. Model* **2007**, *13*, 91.
143. Bosque, R.; Sales, J. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 225.
144. Bosque, R.; Sales, J. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1154.
145. Bosque, R.; Sales, J. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 637.
146. Bosque, R.; Sales, J.; Bosch, E.; Rosés, M.; García-Álvarez-Coque, M. C.; Torres-Lapasió, J. R. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1240.
147. Jover, J.; Bosque, R.; Sales, J. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1098.

Capítol 2. Fonaments i metodologia

En aquest capítol es volen presentar els fonaments teòrics en què es basa la metodologia QSPR; donada l'especificitat d'alguns dels conceptes implicats s'ha dividit en un apartat i dos annexos. En primer lloc s'explica, de manera simple, el procediment general que s'ha seguit en l'aplicació de la metodologia QSPR als diferents problemes plantejats: selecció de dades; dibuix i optimització de les estructures; càlcul de descriptors i l'anàlisi estadística final amb la selecció i validació del model proposat. Els criteris estadístics emprats habitualment en l'anàlisi del model deduït es defineixen en l'Annex 1 mentre que la descripció detallada de tots els conceptes relacionats amb la selecció de descriptors i l'establiment del model es troba en l'Annex 2.

2.1. Esquema general de treball

Aquest apartat ha de servir com a introducció a la metodologia QSPR. En ell s'explicarà, pas a pas, el procediment d'un estudi realitzat amb aquesta metodologia, incloent-hi els mètodes i el programari utilitzats en aquesta Tesi.

2.1.1. Dades d'entrada

Un cop triada la propietat a estudiar cal fer una cerca bibliogràfica exhaustiva per obtenir els valors experimentals del major nombre de compostos possible, en general, aquest nombre sol estar entre un mínim de cent i un màxim d'entre tres-cents i cinc-cents valors. El següent pas consisteix en crear una base de dades pròpia (un full de càlcul pot ser suficient) que contingui el nom dels compostos, el valor de la propietat estudiada i la font bibliogràfica de la qual s'ha obtingut l'esmentat valor experimental. És pertinent fer un comentari general sobre la dificultat de disposar de valors experimentals fiables de propietats físiques o químiques de compostos químics en general, i d'inorgànics en particular. La consulta de més d'una base de dades o *handbooks* acostuma a proporcionar més d'un valor numèric, i sovint, fins i tot, en una mateixa recopilació poden trobar-s'hi dos o més valors diferents d'una mateixa

propietat determinada, per diferents autors, en les mateixes condicions experimentals. Crida també l'atenció que al dia d'avui, quan la Química disposa de fonts bibliogràfiques extraordinàries com el *Chemical Abstracts*, la *Web of Knowledge*, el *DiscoveryGate* - que conté entre altres el *Gmelin* i el *Beilstein* -, i de bases de dades estructurals, també excepcionals, com la *Cambridge Structural Data Base (CSD)*, la *Inorganic Crystal Structure Database (ICSD)* i el *Protein Data Bank (PDB)*, entre altres, el nombre, tipus i qualitat de bases de dades factuais sigui molt més limitat. Amb la notable excepció de les Taules Landolt-Börnstein, i a pesar del cada cop més gran nombre de *handbooks*, que sobre diferents branques de la Química es van publicant, no és fàcil ni immediat disposar d'una col·lecció de valors fiable i prou àmplia de propietats que permeti fer-ne una anàlisi conjunta.

Un cop es tenen triats els valors de la propietat experimental cal dividir el conjunt de dades en dos subconjunts; el primer, que normalment conté el 80% dels valors i s'anomena *training set* o *tset*, s'utilitzarà per establir el model que correlaciona la propietat experimental amb els descriptors. L'altre subconjunt, que conté el 20% restant de valor experimentals, s'anomena *prediction set* o *pset* i s'empra per comprovar la capacitat predictiva del model establert ja que els compostos que el formen no han participat en la seva creació.

2.1.2. Optimització de la geometria

Quan ja es coneix el conjunt de compostos per als quals s'ha de calcular la propietat experimental, cal representar-ne les molècules corresponents. Com que la forma i les possibles conformacions de les molècules juguen un paper molt important, cal utilitzar un programa de dibuix que permeti passar de l'estructura en dues dimensions a la geometria tridimensional el més correcta possible. Hi ha programes comercials que combinen un editor de molècules amb diverses rutines capaces d'optimitzar la geometria, alguns exemples són: PCMODEL,¹ SYBYL² i Spartan.³ En aquest treball, el programa triat per a realitzar aquest procediment ha estat HyperChem.⁴ Amb aquest programari es construeixen els models tridimensionals dels compostos a partir dels dibuixos bidimensionals, i es fa una primera optimització de la geometria mitjançant mecànica molecular.

El següent pas consisteix en optimitzar la geometria dels compostos amb un mètode de càlcul més exacte que la mecànica molecular, el mètode triat en aquest treball sol ser el dels mètodes semiempírics, que donen bons resultats amb temps de càlcul curts. Aquests càlculs semiempírics es realitzen amb el programa MOPAC,^{5,6} gestionat mitjançant el programari AuxQSPR,⁷ desenvolupat en el propi grup d'investigació. Normalment s'utilitza el mètode AM1,⁸ ja que ajusta millor les càrregues parcials dels àtoms en les molècules, factor decisiu en el càlcul d'un gran nombre de descriptors. En cas que els compostos continguin elements no parametritzats en el mètode AM1, es farà servir el mètode PM3,⁹ per exemple, amb compostos organometàl·lics que contenen elements del bloc p com l'antimoni, el bismut i el seleni. Quan no es pot fer servir cap d'aquests dos mètodes semiempírics, perquè hi ha elements no parametritzats en cap dels dos o bé perquè es vulgui fer una optimització de la geometria més ajustada, s'utilitzaran els mètodes DFT que inclou el programa Gaussian.¹⁰ En aquest treball s'ha utilitzat aquest mètode per optimitzar la geometria dels compostos formats en la coordinació dels cations als aminoàcids.

2.1.3. Càlcul de descriptors

Els descriptors moleculars són un dels pilars de la metodologia QSPR. La seva importància ha fet que cada cop se'n desenvolupin més, abastant així cada vegada més característiques estructurals de les molècules. Aquest fet fa necessari escollir amb cura quins es faran servir en cada tipus d'estudi, per aquesta raó, abans de començar s'ha de conèixer quants i de quins tipus de descriptors es disposa. Com ja s'ha comentat anteriorment, els descriptors moleculars són representacions numèriques de les característiques estructurals dels compostos químics que contenen informació sobre els compostos estudiats. És a dir, cada descriptor recull dades sobre una característica diferent de la molècula, assignant-li un valor numèric. Els descriptors moleculars poden ser també valors experimentals d'altres propietats del compost però, en general, és més útil utilitzar descriptors deduïts de l'estructura 2D o 3D ja que aquests permeten estendre la predicció a altres compostos dels quals no es disposa d'informació experimental suficient.

En els últims temps molts grups d'investigació han contribuït a la construcció de la majoria dels descriptors moleculars disponibles, de naturalesa molt diversa. Els grups de Katritzky i Karelson,¹¹⁻¹³ Jurs,¹⁴ Gasteiger,¹⁵ Famini,¹⁶ Todeschini,¹⁷ Balaban,^{18,19} Ivanciuc,²⁰ Kier i Hall,^{21,22} Estrada,^{23,24} etc., han estat dels més importants en aquest aspecte i han aportat un gran nombre de descriptors diferents i de tipologia molt variada, que han resultat veritablement útils en el creixement de la metodologia QSPR. Recentment Livingstone²⁵ ha publicat una revisió en què explica un gran nombre de descriptors, molts dels quals es descriuen detalladament en diversos llibres com els de Karelson²⁶ i Todeschini.²⁷

Aquests descriptors moleculars poden ser calculats per un gran nombre de programes diferents, alguns dels més coneguts són CODESSA (*COmprehensive DEscriptors for Structural and Statistical Analysis*),²⁸ ADAPT (*Automated Data Analysis and Pattern recognition Toolkit*)^{29,30} i Dragon.³¹ En aquest treball s'ha utilitzat principalment el programa CODESSA. Aquest programari, pot calcular fins a mil descriptors diferents. La Taula 2.1, que es troba al final d'aquest capítol, conté una llista amb una bona part dels descriptors calculats per CODESSA.

En general els descriptors moleculars se solen classificar en diferents subconjunts dependent de quin tipus d'informació codifiquen. Així, la classificació més habitual agrupa els descriptors en els següents subgrups:

Descriptors constitucionals

Aquests descriptors són els més fàcils de calcular ja que es deriven directament de la composició molecular, no tenen en compte ni la geometria ni la conformació ni l'estructura electrònica de la molècula i, per aquesta raó, han estat els més àmpliament utilitzats. Alguns exemples poden ser el pes molecular, el nombre total o relatiu dels diferents àtoms, enllaços i anells, etc.

Descriptors topològics³²⁻³⁶

Aquests descriptors descriuen la connectivitat dels àtoms dins de la molècula. El seu càlcul està basat en la teoria de grafs i poden resultar molt útils per modelitzar propietats de molècules que depenen del seu grau de ramificació, com per exemple paràmetres cromatogràfics. Molts dels descriptors que pertanyen a aquesta classe es calculen com el sumatori de la informació continguda en els diferents camins d'enllaç d'ordre n (longitud del camí); de vegades, com en el cas dels índexs Kier & Hall,

aquesta informació pot estar ponderada per la valència dels àtoms implicats. Els índexs de Randic, de Wiener, de Balaban i de Kier & Hall en són alguns exemples.

Descriptors electrostàtics

Aquesta classe de descriptors reflecteix les característiques de la distribució de càrrega de les molècules, és a dir, les càrregues parcials sobre els diferents àtoms. El seu càlcul es realitza a través de l'aproximació empírica de Zefirov,³⁷ basada en l'escala d'electronegativitats de Sanderson, que representa l'electronegativitat molecular com la mitjana geomètrica de les electronegativitats atòmiques.³⁸ Alguns exemples d'aquests descriptors poden ser els índexs topogràfics, les càrregues parcials màximes i mínimes de la molècula i sobre els diferents àtoms, entre d'altres.

Dins d'aquest grup es troba una altra classe específica de descriptors, els anomenats descriptors CPSA³⁹ (*Charged Partial Surface Area*). Aquests es basen en la distribució de càrrega en la superfície molecular i, per tant, permeten caracteritzar la forma i la informació electrònica de la molècula així com la seva capacitat per interaccionar amb altres molècules. Aquest conjunt de descriptors es va ampliar amb la creació d'uns de nous, que tenen la capacitat de contenir informació referent a les interaccions per enllaços d'hidrogen, els anomenats descriptors HB-CPSA⁴⁰ (*Hydrogen-Bond Charged Partial Surface Area*). Dins d'aquests, es consideren grups donadors d'enllaç d'hidrogen els heteroàtoms N, O i S que tenen unit un àtom d'hidrogen. Els grups acceptors d'enllaç d'hidrogen esdevenen aquells grups funcionals amb suficient densitat electrònica per poder participar en un enllaç d'aquest tipus. Per tal de simplificar el càlcul d'aquests descriptors, els halògens, els dobles enllaços i els enllaços aromàtics no es tenen en compte. Aquests descriptors són molt indicats en estudis en què l'enllaç d'hidrogen i les interaccions intermoleculares són rellevants. Alguns exemples de descriptors del tipus CPSA i HB-CPSA poden ser: la superfície total de la molècula, les superfícies totals, parcials i atòmiques carregades positiva o negativament, les diferències entre aquestes superfícies carregades i també els descriptors de tipus HA (*Hydrogen Acceptor*), HD (*Hydrogen Donor*) i HB (*Hydrogen Bonding*).

Descriptors geomètrics

Aquests descriptors es calculen a partir de la posició tridimensional dels nuclis atòmics en l'espai i depenen, habitualment, de la massa i del radi. La bondat d'aquests valors depèn, encara més que per a la resta de descriptors, d'una bona optimització de la geometria molecular. Els moments d'inèrcia, els índexs de projecció d'ombres i el volum molecular en són bons exemples.

Descriptor quàntics⁴¹⁻⁴⁸

Aquest conjunt de descriptors engloba, probablement, el major nombre d'efectius; apart d'això, els descriptors que el componen contenen una gran quantitat d'informació útil que és difícil trobar en els altres tipus. Aquests descriptors se solen extreure directament del fitxer de sortida de l'optimització molecular de MOPAC,^{5,6} encara que alguns d'ells també es poden calcular, per exemple, a partir del fitxers de sortida de Gaussian.¹⁰

i) Relacionats amb la distribució de càrrega

Aquests descriptors representen o depenen directament de la distribució de càrrega calculada amb l'aproximació quàntica de Mulliken⁴⁹ i, per aquesta raó, descriuen les interaccions entre molècules i/o la seva reactivitat química. Aquests descriptors són semblants en interpretació i càlcul als descriptors CPSA i HB-CPSA electrostàtics, encara que se'n troben en major nombre i varietat. Alguns exemples d'aquesta subclasse són: el nombre de nivells electrònics doblement ocupats absolut i relatiu, el moment dipolar i les seves components de càrrega puntual i hibridació, les càrregues màximes i mínimes atòmiques, la població electrònica atòmica màxima, els índexs de reacció electròfila i nucleòfila atòmics.⁵⁰ Alguns descriptors del tipus CPSA també es poden calcular per aquest mètode.

ii) Relacionats amb les valències atòmiques

Aquesta subclasse de descriptors quàntics n'engloba aquells que es relacionen amb les interaccions intramoleculares. Recullen les característiques d'estabilitat de les molècules i la seva flexibilitat conformacional.⁵¹ Alguns exemples clàssics poden ser els ordres d'enllaç màxim i mínim atòmics i per parelles d'àtoms, les contribucions enllaçant i antienllaçant màximes dels orbital moleculars i les valències atòmiques màximes i mínimes.

iii) Relacionats amb l'energia mecanoquàntica

Els descriptors que pertanyen a aquest grup caracteritzen l'energia total de la molècula en diferents escales i la distribució de l'energia intramolecular utilitzant diferents esquemes de partició energètics. Alguns descriptors d'aquesta classe són: l'energia total de la molècula, l'energia electrònica de la molècula, les energies dels orbitals moleculars HOMO, HOMO-1, LUMO i LUMO+1, i les energies de repulsió electrònica i atracció nuclear.

iv) Relacionats amb la solvatació

Aquest grup de descriptors comprèn els models de solvatació de Born per a espècies iòniques, i d'Onsager-Kirkwood per a molècules dipolars.

Descriptors externs

Aquest descriptors no són, de fet, descriptors moleculars extrets directament de l'estructura dels compostos sinó que solen ser propietats experimentals dels compostos químics que es poden trobar en diverses compilacions bibliogràfiques. La selecció dels descriptors d'aquest tipus depèn de l'estudi en particular que es vulgui realitzar.

En el present treball s'han utilitzat aquest tipus de descriptors en dos dels estudis. Així, en els conjunts de compostos inorgànics s'han emprat diversos paràmetres externs per definir de manera més acurada l'efecte de l'àtom metàl·lic en les propietats estudiades. Els descriptors externs són, en aquest cas, propis d'aquest àtom i inclouen el nombre i el pes atòmics, el potencial d'ionització, l'afinitat electrònica, diversos valors d'electronegativitat en diferents escales: Pauling, Mulliken, Sanderson i Allred-Rochow, diferents radis atòmics i iònics i les seves respectives relacions càrrega/radí, la duresa i la tovesa de l'element i dels cations que forma, etc.⁵² En l'apartat 3.3 es mostra la llista completa dels 31 descriptors seleccionats d'aquesta mena, que s'han emprat també en els sistemes multicomponents descrits a l'apartat 4.1.

L'altre estudi on s'han utilitzat descriptors externs és en el càlcul de les constants de dissociació de compostos orgànics en diferents solvents. La caracterització dels solvents s'ha fet només amb descriptors d'aquest tipus, sense calcular-ne cap derivat directament de la seva estructura molecular ja que, en aquest cas, es volia recollir la informació d'aquests compostos com a solvent i no com a molècules individuals. Així, els solvents utilitzats venen descrits per diferents propietats

físicoquímiques: el moment dipolar, l'alfa polaritzabilitat i l'índex de refracció, entre d'altres; i per una gran varietat d'escala de polaritat desenvolupades per Kamlet i Taft, Koppel i Palm, Drago i Catalán, entre d'altres. En total s'han seleccionat 37 descriptors d'aquest tipus, que es comenten en profunditat en l'apartat 4.2.

2.1.4. Establiment dels models

Aquest punt del procés consta de dos passos: primer es fa la selecció de descriptors i després es construeix el model que els relaciona amb els valors de la propietat experimental dels compostos que formen el *training set*. En moltes ocasions aquests dos passos es troben íntimament units i resulta més senzill explicar-los conjuntament. En aquest apartat es farà una descripció dels mètodes utilitzats en aquest treball, que s'ampliarà en l'Annex 2 d'aquest capítol, on també s'expliquen altres procediments i mètodes interessants.

Preselecció de descriptors

Probablement, quan s'arriba a aquest punt es tenen calculats un gran nombre de descriptors diferents; per exemple CODESSA en pot calcular uns 600-700 per a cada compost, ADAPT uns tres-cents i Dragon pot generar prop de dos mil descriptors.

Fins i tot amb el programari i maquinari més moderns, l'elecció dels millors descriptors capaços d'establir una bona correlació amb la propietat experimental és difícil. Per aquesta raó se solen aplicar pretractaments estadístics amb diferents criteris per reduir l'elevat nombre de descriptors a un conjunt més fàcilment manejable. Aquests pretractaments poden ser de dos tipus, l'anomenat objectiu, on només s'avaluen els valors numèrics dels descriptors, i el subjectiu, que ja utilitza la propietat experimental com a element discriminador. En aquest treball, aquest pas es fa utilitzant el mètode heurístic que porta implementat el programari CODESSA, que inclou els dos tipus de pretractament comentats. La reducció de descriptors que fa aquest mètode inclou els criteris estadístics següents: en primer lloc s'eliminen aquells descriptors incomplets, els que es mantenen invariants en el conjunt de compostos analitzat i els que estan correlacionats amb altres descriptors. Després es fan les correlacions individuals de cada descriptor amb la propietat, els resultats que s'obtenen d'aquest pas

permeten eliminar els descriptors que donen un coeficient de determinació insignificant ($R^2 < 0.01$) amb la propietat experimental.

Amb aquest mètode es redueix el nombre de descriptors a uns dos-cents. Un cop fet això es poden establir relacions lineals amb el programa CODESSA o es poden exportar aquests descriptors al altres programes que permetin establir models no lineals. Normalment, les relacions lineals s'estableixen amb el programa CODESSA, mentre que les relacions no lineals es dedueixen amb els programes ADAPT i LibSVM.⁵³ Cal tenir en compte, però, que el programa ADAPT només permet treballar amb un nombre relativament petit de descriptors (menys de 120); així doncs, abans d'establir models amb aquest programa caldrà fer una altra reducció de descriptors amb criteris objectius un mica més estricta que la que fa CODESSA.

Selecció de descriptors i establiment de models

Aquest pas proporciona el model que permet relacionar la propietat estudiada amb els descriptors moleculars deduïts de l'estructura. En aquesta Tesi s'han desenvolupat models de dos tipus: lineals i no lineals. Els models lineals s'estableixen mitjançant el procediment de regressió multilíneal de CODESSA mentre que els models no lineals es construeixen amb altres mètodes, fonamentalment xarxes neuronals, implementades en els programes ADAPT* i LibSVM.

En general, els models no lineals proporcionen uns resultats millors que els obtinguts per regressió lineal. Tot i així, els models lineals tenen un avantatge remarcable sobre els desenvolupats amb xarxes neuronals pel què fa a la seva interpretació, ja que en aquests últims no es disposa de cap tipus d'informació directa que relacioni l'efecte del descriptor en la propietat estudiada.

* S'agraeix el Prof. P. C. Jurs per donar-nos accés al programa ADAPT.

Models lineals

Els models lineals s'estableixen amb el mètode heurístic de CODESSA, a partir del conjunt de descriptors preseleccionats per aquest mateix programa. Els procediments de selecció de descriptors i establiment del model són impossibles de separar ja que aquest mètode avalua alhora l'entrada i sortida dels descriptors i l'efecte que produeixen en el model, seleccionant-ne aquells que proporcionen els millors models, estadísticament parlant. Els paràmetres estadístics per avaluar la bondat del model són calculats per aquest mateix programa: coeficient de determinació (R^2), desviació estàndard (sd) i test- F . Seguint aquest procediment, que es pot considerar un ajust per mínims quadrats múltiple, s'obtenen uns models que es poden representar com a equacions lineals:

$$\text{ propietat} = a_1d_1 + a_2d_2 + a_3d_3 + \text{ terme independent}$$

on d_i són els descriptors inclosos en el model i a_i són els seus coeficients de regressió. Com a norma general, el nombre de descriptors que ha de contenir el model es troba de manera iterativa, començant per un nombre baix de descriptors i augmentant-lo d'un en un fins que l'addició d'un nou descriptor no modifiqui significativament els resultats obtinguts.

El signe dels coeficients de regressió de cada descriptor és indicatiu de la seva influència, positiva o negativa, en el model i sovint ajuda a la seva interpretació fisicoquímica.

Models no lineals

Els models no lineals que s'estableixen habitualment són de dos tipus diferents, el primer, que es pot anomenar no lineal pur, consisteix a realitzar una selecció no lineal dels descriptors, en general exportats des de CODESSA, i a partir d'ells, establir la relació no lineal, pròpiament dita. Aquest pas s'ha dut a terme de dues maneres diferents amb els programes ADAPT i LibSVM. L'altra opció disponible és dotar de no linealitat a un model lineal construït prèviament amb CODESSA; aquest procediment, que s'anomena mètode híbrid lineal-no lineal, utilitza els descriptors del model lineal com a factors d'entrada per als programes que permeten establir models no lineals.

El mètode més habitual d'establir models no lineals dintre de la metodologia QSPR és la utilització de xarxes neuronals. Les xarxes neuronals, desenvolupades en l'àmbit de la intel·ligència artificial, són un conjunt de d'algorismes i mètodes dissenyats per intentar imitar el funcionament del cervell humà.

Una xarxa neuronal consta de, com a mínim tres capes, una d'entrada, una de sortida i una, o més, d'ocultes. La Figura 2.1 mostra una xarxa neuronal completament connectada de tres capes, el nombre de neurones constituents de cada capa conforma l'arquitectura de la xarxa, així doncs, l'arquitectura d'aquesta xarxa neuronal és 5-3-1. La capa d'entrada, conté els descriptors triats per calcular la propietat experimental. La segona capa, que conté les neurones ocultes, és el centre de càlcul de la xarxa neuronal i on s'introdueix la no linealitat; el nombre de neurones que conté i la seva naturalesa varien segons el mètode de càlcul que s'estigui utilitzant. La capa de sortida conté una sola neurona que correspon al valor de la propietat experimental.

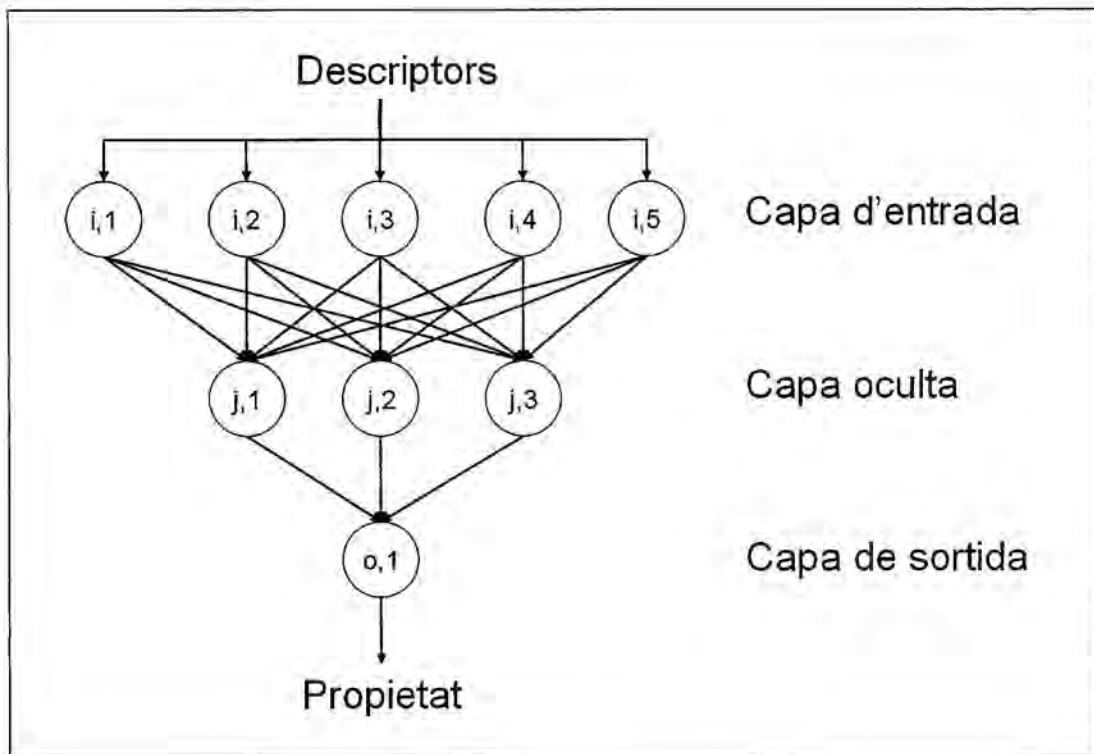


Figura 2.1. Esquema general d'una xarxa neuronal.

Les xarxes neuronals utilitzades en aquest treball són de dos tipus diferents: les xarxes neuronals amb retropropagació de l'error o BPNN (*Back Propagation Neural Network*) i les xarxes neuronals que utilitzen funcions de base radial o RBFNN (*Radial Basis Function Neural Network*), classe a la qual pertanyen les màquines de vectors suport o *Support Vector Machines* (SVM). La diferència entre totes dues classes es deu a la naturalesa de les funcions matemàtiques que s'hi utilitzen i en la manera de seleccionar el nombre de neurones de la capa oculta.

Els resultats obtinguts amb aquest tipus de càlcul solen ser de major qualitat, estadísticament parlant, que els obtinguts amb mètodes lineals. Tot i així, la interpretació fisicoquímica dels descriptors es complica força ja que, en aquest cas, no es disposa d'una equació per al model que permeti comprovar l'efecte individual de cadascun.

i) Xarxes neuronals BPNN

Aquest tipus de models no lineals són la combinació d'una selecció de descriptors amb algorismes genètics i una xarxa neuronal, de tipus BPNN, com a funció avaluadora del model, totes dues implementades pel programa ADAPT. En aquest cas, donat que l'avaluació dels descriptors inclosos en el model es fa amb una xarxa neuronal, es diu que la selecció de descriptors és no lineal.

Per tal d'establir models amb el programa ADAPT és necessari complir certs requisits que no són comuns a la resta de programes emprats. Primerament, és imprescindible fer una altra divisió del conjunt de dades. Un 10% dels compostos del *tset* passa a conformar l'anomenat *cross-validation set* (o *cvset*), aquest nou conjunt és utilitzat per la xarxa neuronal per establir el nombre màxim d'iteracions de càlcul. Aquest nombre és molt important ja que determina la capacitat predictiva de la xarxa neuronal. En cas de no utilitzar-ne la xarxa neuronal pot "memoritzar" el procediment de càlcul i produir models que no tenen cap capacitat predictiva per a compostos externs no inclosos en el *training set* emprat.

El segon requisit per utilitzar aquest programa és l'optimització de l'arquitectura de la xarxa neuronal. Aquesta optimització consisteix en comprovar quantes neurones ocultes es poden incloure en l'estudi. En general, quan es vol desenvolupar un model no lineal d'aquest tipus ja s'ha seleccionat el nombre de descriptors inicials que es

col·locaran en la capa d'entrada de la xarxa neuronal. Així, l'optimització del nombre de neurones ocultes es fa amb un procediment d'addició successiva, començant per un nombre baix, tot comprovant en quin punt una nova addició no fa millorar el model. Per exemple, l'optimització de l'arquitectura d'una xarxa neuronal amb cinc descriptors en la capa d'entrada es comprovaria construint les xarxes d'arquitectura 5-2-1, 5-3-1, 5-4-1, etc. fins al punt en què l'addició d'una neurona oculta més no millorés significativament el model.

Un cop es tenen seleccionats els compostos que conformen el *cvset* i optimitzada l'arquitectura de la xarxa neuronal, cal comprovar la relació entre el nombre de compostos del *tset* i el nombre paràmetres ajustables. És molt important que aquest valor sigui més gran que 2.0 per tal d'evitar el sobreaprenentatge (memorització) de la xarxa neuronal. Els paràmetres ajustables d'una xarxa neuronal es calculen: $PA = [(NE+1) \times NO] + [(NO+1) \times NS]$, on *NE*, *NO* i *NS* corresponen al nombre de neurones en les capes d'entrada, oculta i de sortida, respectivament. Un cop aconseguits aquests requisits s'estableixen els models no lineals adients i s'avaluen els resultats obtinguts.

ii) *Support Vector Machines*

Les màquines de vectors suport o SVM són un mètode, relativament recent, per establir models no lineals utilitzant xarxes neuronals de tipus RBFNN. Els models fruit d'aplicar aquesta metodologia són la combinació d'una selecció no lineal de descriptors amb el mètode heurístic-SVM i de les xarxes neuronals, basades en el mètode de *Support Vector Machines*, implementades pel programa LibSVM, acoblat a funcions de Matlab.⁵⁴ Aquest procediment és, globalment, equivalent al mètode heurístic que s'utilitza per construir els models lineals amb CODESSA. En aquest cas, però, s'ha substituït la regressió multilíneal per una regressió no lineal amb SVM com a funció avaluadora. Aquest procediment de selecció de descriptors és una aportació original d'aquest treball a la metodologia QSPR.

Tant en el procés de selecció de descriptors com en l'establiment definitiu del model, el mètode SVM precisa de l'ajust de tres paràmetres característics γ , ϵ i *C*. El paràmetre γ és l'amplada de les funcions gaussianes que s'utilitza en el procediment de càlcul i és una mesura de la seva capacitat d'interpolació. *C* és l'anomenat paràmetre de regularització i és una mesura de l'equilibri que s'estableix entre l'optimització de les dades amb què es construeix el model i la capacitat predictiva del mateix. Per últim, el

paràmetre ε , anomenat marge de tolerància, és una mesura que determina el nombre de *vectors suport* que s'utilitzaran en el càlcul. Durant el procés de selecció, els tres paràmetres característics es mantenen constants amb els valors per defecte que proporciona LibSVM: $\gamma=0.1$, $\varepsilon=0.1$ i $C=100$, ja que, com s'ha comprovat, la seva variació no afecta de manera significativa l'ordre dels descriptors en el procés de selecció. Un cop s'han seleccionat els descriptors amb els compostos del *tset*, es construeix el model definitiu amb l'optimització dels tres paràmetres característics.

iii) Model híbrid lineal-no lineal

Aquest procediment consisteix en dotar de no linealitat un dels models obtinguts per regressió lineal amb CODESSA. Així, aquest mètode és el més ràpid per obtenir models no lineals ja que no necessita fer el pas de selecció de descriptors. Aquesta no linealitat s'aconsegueix aplicant els descriptors obtinguts amb el model lineal com a dades d'entrada per a les diferents xarxes neuronals disponibles BPNN i SVM.

Així, per trobar els models no lineals definitius amb BPNN només cal optimitzar l'arquitectura de la xarxa neuronal. En el cas que s'utilitzin xarxes neuronals de tipus SVM, l'obtenció del model es produeix per optimització dels tres paràmetres característics γ , ε i C .

En general, els programes utilitzats donen 10 models amb uns paràmetres similars des d'un punt de vista estadístic. D'aquests se selecciona aquell que sembla tenir uns descriptors més capaços d'interpretar la propietat analitzada.

2.1.5. Validacions

La bondat dels models trobats s'avalua pels criteris estadístics més habituals: coeficient de determinació (R^2), el test- F de Snedecor, el test- t de Student i l'error quadràtic mitjà (*rmse*). Altres paràmetres estadístics importants en els models lineals poden ser el coeficient de regressió *leave-one-out* o de validació creuada (R^2_{cv}), la desviació estàndard de la correlació (*sd*), els coeficients de regressió estandarditzats de cada descriptor (β) que indiquen la importància relativa d'un descriptor; i el factor d'inflació de la variància (VIF) que és un diagnòstic de col·linealitat dels descriptors.

Aquests dos darrers paràmetres estadístics es calculen amb el programa SPSS.⁵⁵ Tots aquests paràmetres estadístics i alguns altres es troben detallats en l'Annex 1.

De tots els mètodes de validació dels models que es poden establir amb la metodologia QSPR el més acceptat és la predicció d'un conjunt de valors experimental que no hagi participat en la generació del model. En aquest punt del procés es prediu el valor de la propietat experimental per als compostos del *prediction set* que s'havien apartat prèviament. Els resultats que s'obtenen es caracteritzen pels criteris estadístics que s'han comentat en el punt anterior.

Per a tots els models deduïts en aquesta Tesi s'han representat gràficament els valors experimentals enfront dels calculats. D'aquesta manera es pot comprovar visualment que la relació establerta és lineal i que, per tant, és correcte presentar els coeficients de determinació.

Per tal d'assegurar que els models establerts presenten bona capacitat predictiva, un cop es tenen els resultats, es fa una anàlisi de *outliers*. Un *outlier* és un compost que, per les raons que siguin, a causa de la seva estructura o perquè els valors experimentals no han estat determinats amb prou exactitud, no es troba ben descrit pel model. Aquest procediment consisteix en analitzar aquells compostos que presenten un valor de residual (diferència entre el valor experimental i el calculat) que sigui com a mínim dues vegades més gran que l'error mitjà del model. Normalment, es considera que un compost és un *outlier* si el seu valor de residual és superior a tres vegades aquest valor i, per tant, es podria eliminar del model. En general, les anàlisis de *outliers* s'han realitzat en els models no lineals que presenten millors resultats o que presenten alguna particularitat estructural interessant.

Una altra manera de comprovar el comportament dels models consisteix en analitzar els resultats per subconjunts de compostos anàlegs, de la mateixa família o naturalesa química. Per exemple, en els models desenvolupats per calcular propietats per compostos organometàl·lics s'han analitzat els subconjunts formats pels compostos de cada metall i en els estudis de pK_a s'han analitzat els conjunts de cada solvent. Aquest tipus d'anàlisi serveix per comprovar que el model no afavoreix el càlcul d'un subconjunt de compostos per sobre de la resta i que els resultats que s'obtenen no són

deguts a ajustos locals de les dades estudiades. Normalment, aquest tipus d'anàlisi s'ha realitzat per a models no lineals, per a les propietats que contenen un nombre relativament gran de observacions i per als sistemes multicomponents.

Cal tenir sempre present que els objectius finals de la metodologia QSPR són dos. El primer és establir un model que estimi la propietat experimental a partir, exclusivament, dels descriptors moleculars i que, per tant, pugui ser emprat amb compostos que no hagin estat considerats en la seva deducció. D'altra banda es pretén també que els descriptors que formen el model permetin una interpretació fisicoquímica de la propietat.

Com a resum, a la Figura 2.2 es mostra en detall el diagrama de flux d'un estudi QSPR amb tots els passos realitzats on s'inclouen els mètodes i els programes utilitzats en cadascun d'ells, en color verd i blau, respectivament.

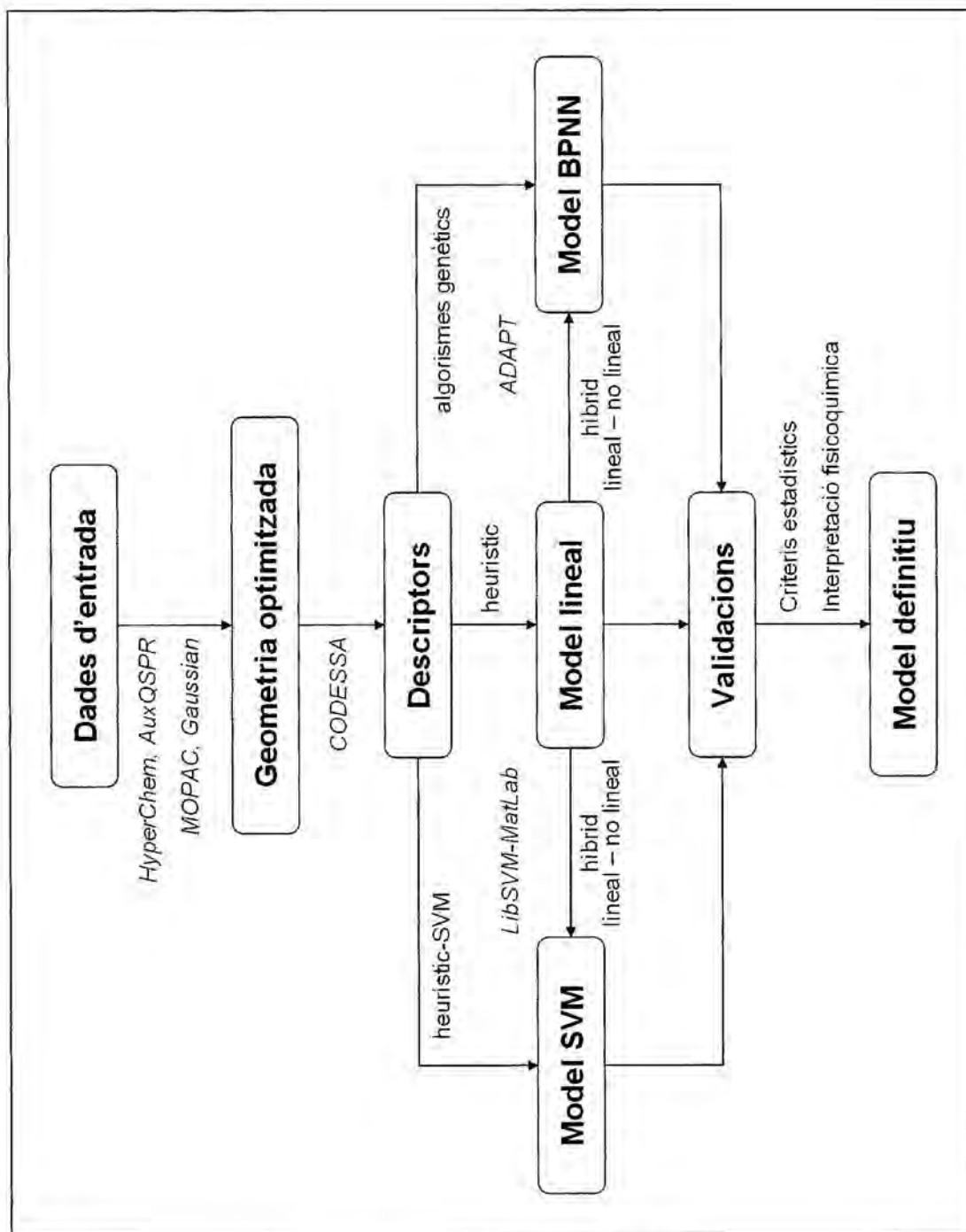


Figura 2.2. Procediment d'un estudi QSPR.

2.2 Annexos

Annex 1. Criteris estadístics

La bondat dels models establerts, ja siguin lineals o no lineals, s'avalua mitjançant diferents criteris estadístics. Hi ha criteris estadístics que són comuns per a tots els models: el coeficient de determinació R^2 , el coeficient de regressió *leave-one-out* o de validació creuada R^2_{cv} , el test-*F* de Snedecor, el test-*t* de Student, la desviació estàndard (*sd*) de la regressió i l'error quadràtic mitjà (*rmse*). Alguns paràmetres estadístics són propis dels mètodes lineals: els coeficients de regressió estandarditzats de cada descriptor (β), que indica la importància relativa d'un descriptor dins del model; i el factor d'inflació de la variància (VIF) que és un diagnòstic de col·linealitat dels descriptors. Donat que els mètodes que permeten obtenir models no lineals no presenten una equació definida hi ha pocs paràmetres estadístics que permetin extreure informació extra del model, potser el més important és el mètode desenvolupat pel grup de P. Jurs,⁷⁸ que permet establir l'ordre d'importància dels descriptors dins de qualsevol model no lineal. En aquesta secció s'enunciaran i s'explicarà com es calculen tots aquests criteris estadístics. Aquest paràmetres estadístics es poden dividir en dos subconjunts, els anomenats criteris estadístics de regressió, que es deriven de la comparació directa de la propietat experimental amb la propietat calculada pel model; i criteris estadístics relacionats amb els descriptors, que proporcionen informació sobre els descriptors inclosos en el model.

A1.1. Criteris estadístics de regressió

Com s'acaba de comentar, aquests paràmetres estadístics s'obtenen de la comparació de la propietat calculada pel model i el seu valor experimental, sense cap influència directa per part dels descriptors que componen el model.

Coefficient de determinació R^2

Aquest paràmetre és un dels més utilitzats per descriure dependències lineals entre dos grups de dades, es deriva del coeficient de correlació lineal de Pearson i és una mesura de la quantitat de variància que el model pot explicar de la variable dependent. Per exemple, un valor de R^2 de 0.94 vol dir que el model pot explicar el 94% de la variància del model. En el tipus de càlculs que es poden trobar en aquest treball (proprietat calculada vs. experimental), el valor de R^2 significa el tant per cent de compostos que el model calcula de manera correcta amb uns límits de confiança fixats (generalment del 95%).

El coeficient R^2 es calcula directament del coeficient de correlació lineal de Pearson, R , el qual es calcula de la següent manera:

$$R = \frac{\sum (y_{\text{exp}} - \bar{y}_{\text{exp}})(y_{\text{calc}} - \bar{y}_{\text{calc}})}{\sqrt{\sum (y_{\text{exp}} - \bar{y}_{\text{exp}})^2 \sum (y_{\text{calc}} - \bar{y}_{\text{calc}})^2}}$$

Coefficient de regressió leave-one-out o de validació creuada R^2_{cv}

Aquest paràmetre és una mesura de l'estabilitat de les correlacions establertes i descriu la sensibilitat del model enfront l'eliminació de qualsevol punt experimental. El coeficient de regressió de validació creuada es troba de la següent manera: per a cada punt experimental es recalcula la regressió multilínia traient aquest punt, la regressió que s'obté s'utilitza per a predir el valor del punt experimental eliminat i finalment es correlaciona la diferència dels punts experimentals predits amb la dels punts experimentals. L'error típic d'aquesta darrera correlació és l'error *leave-one-out* que s'utilitza en diversos mètodes de càlcul.

Desviació estàndard de la regressió (sd)

Aquesta desviació estàndard és, realment, l'error típic que s'obté en correlacionar la propietat experimental amb la calculada pel model. És una mesura de la variabilitat en la distribució dels punts al voltant de la recta de regressió encara que també es pot veure com la magnitud de l'error que es comet en el pronòstic de la propietat calculada. Aquest paràmetre es calcula de la següent manera:

$$sd = \sqrt{\frac{1}{(n-2)} \left[\sum (y_{calc} - \bar{y}_{calc})^2 - \frac{\left[\sum (y_{exp} - \bar{y}_{exp})(y_{calc} - \bar{y}_{calc}) \right]^2}{\sum (y_{exp} - \bar{y}_{exp})^2} \right]}$$

Error quadràtic mitjà (rmse)

Aquest paràmetre és el que donen la majoria dels programes que estableixen models no lineals. Pot semblar equivalent a la desviació estàndard de la regressió, sobretot perquè dona valors molt semblants, però en realitat no ho és. Tot i així, aquest criteri és també un estimador de l'error que es produeix en calcular una propietat amb el model triat. Normalment es presenta en forma de l'arrel quadrada de la mitjana de les diferències entre propietat experimental i calculada:

$$rmse = \sqrt{\frac{\sum (y_{exp} - y_{calc})^2}{n}}$$

Test-F de Snedecor per la regressió

El test- F de Snedecor per a un model és una mesura de la probabilitat que la correlació es doni per atzar; és a dir, que s'obtingui un valor de correlació diferent de zero quan, en realitat, la correlació no existeix. Com més gran sigui el valor de F que s'obté, més gran serà la correlació existent i, per contra, com més baix sigui el valor de F , més probable serà que la correlació es doni per atzar. Per tant, un model que presenti un valor baix per al test- F no es pot fer servir per predir la propietat d'altres compostos. Una manera de calcular el valor de F de Snedecor per un model qualsevol és el següent:

$$F = \frac{\sum (y_{calc} - \bar{y}_{calc})^2}{\sqrt{\frac{\sum ((y_{exp} - y_{calc}) - \bar{y}_{exp})^2}{(n-1)}}$$

A1.2. Criteris estadístics relacionats amb els descriptors

Aquests criteris s'extreuen individualment de cadascun dels descriptors triats pel model i poden representar diferents diagnòstics del mateix. Majoritàriament, els criteris que es presentaran aquí fan referència a dos aspectes dels descriptors: la seva importància dintre del model i la seva col·linealitat.

Test- t de Student

Aquest criteri estadístic serveix per avaluar si un descriptor és significatiu dins d'un model, és a dir, si aquest descriptor està aportant o no capacitat predictiva al model. El test- t per descriptors és semblant al test- F pel model complet però, en aquest cas, ens permet avaluar els descriptors un per un. Així, valors de test- t elevats indiquen que els descriptors són importants dintre del model, per contra, valors petits donen a entendre que els descriptors són poc importants i que, fins i tot, es podrien eliminar. Aquest criteri, que només està disponible per als mètodes lineals, sol aprovar l'entrada de descriptors en el model si el seu test- t dóna un valor superior a quatre, encara que a vegades aquest nombre pot ser flexible.

La manera de calcular el test- t per un descriptor inclòs en un model lineal consisteix en fer el quocient del coeficient del descriptor entre el seu error d'estimació de la regressió.

Coefficients de regressió estandarditzats (β)

Aquests coeficients es calculen seguint el mateix procediment amb què es calculen els valors del test- t , per tant, aquest criteri també serà aplicable només a models desenvolupats amb mètodes lineals. En aquest cas, però, els descriptors s'han de transformar en la seva forma tipificada abans de construir el model. La forma tipificada d'un descriptor s'obté per autoescalat del mateix. L'autoescalat és simplement un procediment en el qual, a cada descriptor, se li resta la seva mitjana i se'l divideix entre la seva desviació estàndard. D'aquesta manera s'obtenen uns nous descriptors més compactes entre sí ja que el valor mitjà per tots ells és zero i la seva dispersió sempre val 1. La Figura 2.3 mostra un esquema del procediment d'autoescalat.

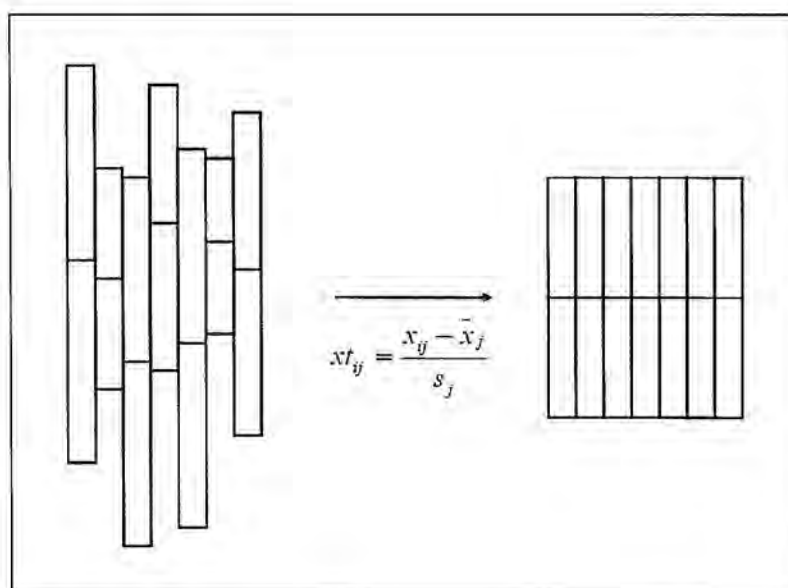


Figura 2.3. Esquema d'autoescalat de descriptors.

Un cop autoescalats els descriptors, es construeix el model; els coeficients de regressió estandarditzats s'obtenen de la mateixa manera que els valors de test- t : dividint el coeficient de regressió individual entre el seu error. Amb aquest mètode de

treball sí que es pot afirmar quin és l'ordre exacte d'importància dels descriptors dins del model, ja que els coeficients de regressió estandarditzats (amb el seu valor absolut) així ho permeten.

Factor d'inflació de la variància (VIF)

Aquest criteri és un diagnòstic de col·linealitat aplicable a cadascun dels descriptors dintre d'un model lineal. Aquest paràmetre, que permet comprovar quin grau d'informació comuna contenen els diferents descriptors del model, es defineix com:

$$VIF_i = \frac{1}{1 - R_i^2}$$

on R_i^2 és el coeficient de correlació del descriptor i amb la resta de descriptors en el model. En cas que no existeixi correlació entre els descriptors del model, el valor de VIF ha de ser molt pròxim a 1; un valor de VIF superior a deu indica un grau de col·linealitat entre els descriptors massa elevat dins del model. Així doncs, en els models lineals, és desitjable obtenir valors el més petits possible de VIF en un rang 1-10.

Mètode de Jurs⁷⁸

Aquest mètode, que ha estat desenvolupat recentment, permet establir una mesura de la importància relativa dels diferents descriptors continguts en qualsevol model no lineal. Aquest és el primer pas que s'ha fet per millorar la interpretació dels models lineals, on no es disposa d'una recta de regressió que permeti una interpretació més directa i intuïtiva dels descriptors continguts en el model.

En aquest apartat s'explicaran els passos a seguir en aquest tipus d'estudi. Primer de tot cal tenir optimitzat el model no lineal que es vulgui tractar, aquest model tindrà un *rmse* característic que s'anomena *rmse* base. El següent pas consisteix en substituir el primer descriptor per una permutació d'ell mateix i, tot seguit, utilitzar el model no lineal per predir el valor de la propietat experimental. És d'esperar que, havent fet aquesta substitució del primer descriptor, la relació estructura-propietat que fa el

model no funcioní tan bé en aquesta ocasió; així doncs, en permutar el descriptor el valor de *rmse* hauria de ser més gran. La diferència entre aquest nou *rmse* i el *rmse* base del model proporciona la mesura de la importància del descriptor dintre del model; com més gran sigui aquesta diferència, major és la influència del descriptor en la capacitat predictiva del model i major és la seva importància. Això vol dir que els descriptors més importants dintre del model donaran variacions en el *rmse* molt més grans que aquells descriptors que no ho són tant. Es repeteix aquest procediment per la resta de descriptors, obtenint per cadascun el seu augment de *rmse*, la qual cosa permet ordenar els descriptors en funció de la seva importància dintre del model.

Annex 2. Mètodes de selecció de descriptors i construcció de models

En aquesta secció s'explicaran, detalladament, els mètodes de pretractament i tria de descriptors i les maneres més habituals d'establir models, tant lineals com no lineals, que s'han esbossat en l'apartat anterior. Com ja s'ha dit, sovint els processos de reducció i selecció de descriptors estan íntimament units i és difícil fixar el punt on acaba un i comença l'altre.

A2.1. Selecció de descriptors

La selecció de descriptors consisteix, normalment, en dos passos conjunts: reducció del nombre de descriptors i tria d'aquells que acabaran constituint el model que permet calcular la propietat estudiada. La reducció de descriptors se sol fer utilitzant criteris estadístics subjectius i objectius. En els del primer tipus s'utilitza el valor de la propietat experimental com a discriminant mentre que els del segon tipus només depenen dels descriptors. Aquests mètodes produeixen l'eliminació dels següents descriptors:

- Pretractament subjectiu: Descriptors que contenen poca o gens d'informació sobre la propietat estudiada; generalment, aquells amb què s'obté un coeficient de correlació (R^2) molt petit amb la propietat experimental.
- Pretractaments objectius: Descriptors que es mantenen invariants o que són incomplets en el conjunt de compostos estudiats. Descriptors que estan molt correlacionats amb altres descriptors i que, normalment, contenen la mateixa informació.

De manera habitual, aquesta reducció de descriptors es realitza utilitzant mètodes de regressió multilíneal; valgui com a exemple el mètode heurístic de CODESSA, que s'explicarà en el següent punt i que és el que s'ha utilitzat de manera general en aquest treball.

El segon pas de la selecció de descriptors consisteix en trobar el millor subconjunt de descriptors que permetin construir el millor model possible que els relacioni amb la propietat experimental estudiada. Hi ha una gran quantitat de mètodes que permeten aquesta selecció final de descriptors: els mètodes de regressió multilíneal clàssics, la millor regressió lineal múltiple, el mètode heurístic, els algorismes genètics, el mètode

GA-PLS, etc. Tot seguit s'explicaran alguns d'aquests mètodes, fent èmfasi en aquells que s'han utilitzat en aquest treball.

A2.2. Mètodes de regressió multilineal clàssics⁵⁶

Mètode “selecció endavant”

Aquest mètode consisteix en anar afegint, d'un en un, descriptors al model. Els descriptors es van afegint consecutivament per ordre de correlació amb la propietat experimental, és a dir, el primer descriptor inclòs és aquell que presenta la major correlació amb la propietat experimental; el segon descriptors és el que presenta la segona major correlació amb la propietat, etc. El procés s'acaba quan el darrer descriptor introduït en el model no el millora significativament o bé quan tots els descriptor disponibles es troben dins del model.

Mètode “selecció enrere”

En aquest cas, tots els descriptors es troben inicialment dins del model. El procediment de selecció consisteix en anar-los eliminant d'un en un per ordre creixent d'error, és a dir, primer s'eliminarà el descriptor que impliqui una menor reducció de l'error del model i es continuarà així, consecutivament, fins que no quedin descriptors poc importants o fins que no en quedi cap.

Mètode “pas a pas”

Aquest mètode de selecció de descriptors és una mescla dels dos anteriors. Essencialment, aquest mètode funciona com el de “selecció endavant” però, en cada etapa, existeix la possibilitat que un descriptor ja inclòs en el model sigui eliminat (com amb el mètode “selecció enrere”). Això implica que un descriptor inclòs en les primeres etapes de selecció pugui ser substituït per un altre en les etapes finals. Els nombre de descriptors que contindrà el model final depèn dels nivells de significació fixats per a la seva entrada i sortida en el model.

Mètode “millor regressió lineal múltiple”²⁷

Aquest mètode aplica el següent procediment per trobar la millor regressió multilíneal possible:

Pas 1: Es busquen les parelles ortogonals de descriptors ($R^2_{desc} < R^2_{min}$) per al conjunt de dades a tractar.

Pas 2: Es correlaciona la propietat experimental amb les parelles de descriptors obtingudes en el pas anterior i es trien les millors (normalment les 400 millors) per als passos següents.

Pas 3: Es troben els descriptors ortogonals als descriptors de la parella triada en el pas anterior i s'afegeixen al model successivament, tot avaluant la seva inclusió mitjançant la regressió multilíneal de tres termes. El tercer descriptor inclòs en el model s'accepta sempre i quan el criteri d'entrada de Fisher que s'obté no sigui menor a aquell que s'ha trobat per a la millor parella de descriptors. Si el descriptor és acceptat dins el model, el procediment continuarà en el següent pas, en cas contrari, cal passar directament el pas 5 del procés.

Pas 4: Es repeteix el pas anterior, afegint cada vegada un descriptor més. Aquest procés s'acabarà quan el criteri d'entrada de descriptors de Fisher així ho indiqui.

Pas 5: S'obtenen els següents resultats corresponents a la millor regressió multilíneal de cada dimensió.

- Els coeficients de regressió R amb els seus errors.
- La desviació estàndard de la regressió (sd).
- El coeficient de determinació R^2 i de validació creuada R^2_{cv} .
- La desviació estàndard ponderada de la regressió.

*Mètode heurístic*²⁷

Aquest mètode combina la capacitat de reduir el nombre de descriptors inicials amb l'establiment del model que els relaciona amb la propietat experimental. Totes les reduccions de descriptors que s'han fet en aquest treball s'han realitzat amb el mètode heurístic implementat en el programa CODESSA.

La reducció de descriptors que fa aquest mètode inclou els següents criteris estadístics. En primer lloc s'eliminen aquells descriptors incomplets o que romanen invariants en el conjunt de compostos analitzat. Després es fan les correlacions individuals de cada descriptor amb la propietat, els resultats que s'obtenen d'aquest pas permeten eliminar certs descriptors: els que presenten un test- F de Snedecor o un test- t de Student menor que 1, els que donen un coeficient de determinació insignificant ($R^2 < 0.01$) enfront la propietat experimental i els que estan correlacionats amb altres descriptors. Aquests últims es poden diferenciar en dues classes:

- Descriptors molt correlacionats ($R > 0.99$), en aquest cas s'elimina el descriptor que presenta un coeficient de correlació pitjor amb la propietat experimental.
- Descriptors correlacionats moderadament ($R > 0.8$), en aquesta ocasió no s'elimina cap dels descriptors però el mètode s'assegura que mai es trobaran alhora dins d'un model.

Un cop reduït el conjunt de descriptors, es pot procedir a establir el model que permet calcular la propietat experimental. Abans de començar, cal ordenar els descriptors en ordre decreixent de coeficient de determinació; un cop fet això, es procedeix de la següent manera:

Pas 1: Aquest pas serveix per trobar les millors parelles inicials de descriptors. Es comença pel millor descriptor de la llista, s'estableixen les correlacions de dos termes amb tota la resta de descriptors i se seleccionen les parelles de descriptors que presenten valors de test- F superiors a un terç del màxim trobat en el conjunt de dades. Després es repeteix aquest procediment amb la resta de descriptors de la llista. Normalment es trien les tres millors parelles de descriptors per continuar.

Pas 2: Es van afegint, d'un en un, la resta de descriptors de la llista, construint d'aquesta manera tots els models possibles de tres descriptors amb regressions

multilineals. Un descriptor és acceptat en el model de $n+1$ descriptors sempre i quan el valor del test- F de Snedecor que s'obté en incloure'l sigui major al valor del test- F del model de n descriptors multiplicat per n i dividit per $(n+1)$, on n és el nombre de descriptors del model inicial.

Pas 3: Un cop acceptat el descriptor, es repeteix el pas anterior, augmentant el nombre de descriptors d'un en un fins que s'arribi al nombre màxim de descriptors (triat per l'usuari) o fins que trenqui l'augment progressiu en el test- F comentat en el punt anterior.

El mètode heurístic dona com a resultat els deu millors models ordenats per coeficient de determinació R^2 i els deu millors models ordenats per test- F , encara que tot sovint aquests models coincideixen.

*Algorismes genètics*⁵⁷

Aquest mètode, tot i que sembli una mica agosarat de dir, intenta imitar la selecció natural que es dona dins de l'evolució biològica. A diferència dels mètodes anteriors, aquest mètode permet la tria de descriptors de manera lineal i no lineal, dependent de quin mètode s'utilitzi en l'avaluació interna dels models obtinguts. En general, en aquest treball, els mètodes emprats en aquest tipus de selecció de descriptors han estat no lineals i, més en concret, amb les xarxes neuronals implementades en el programa ADAPT. L'esquema general d'aquest procediment es pot veure en la Figura 2.4.

En el primer pas, l'algorisme crea aleatòriament un nombre de models diferents (normalment entre 40 i 50) de la dimensió triada, val a dir que aquests models no tenen cap valor per establir correlacions, són només un punt de partida. Tots aquests models es tracten alhora i, mitjançant una funció d'error, s'ordenen de millor a pitjor. El 50% millor passen a la fase següent, on poden romandre igual (reproducció), creuar-se amb altres models (acoblament) o patir canvis a l'atzar en algun dels seus descriptors (mutació). Els processos d'acoblament i mutació, que no semblen immediats, es poden entendre millor a partir de la Figura 2.5 on es presenten exemples d'un acoblament i una mutació en models de cinc descriptors.

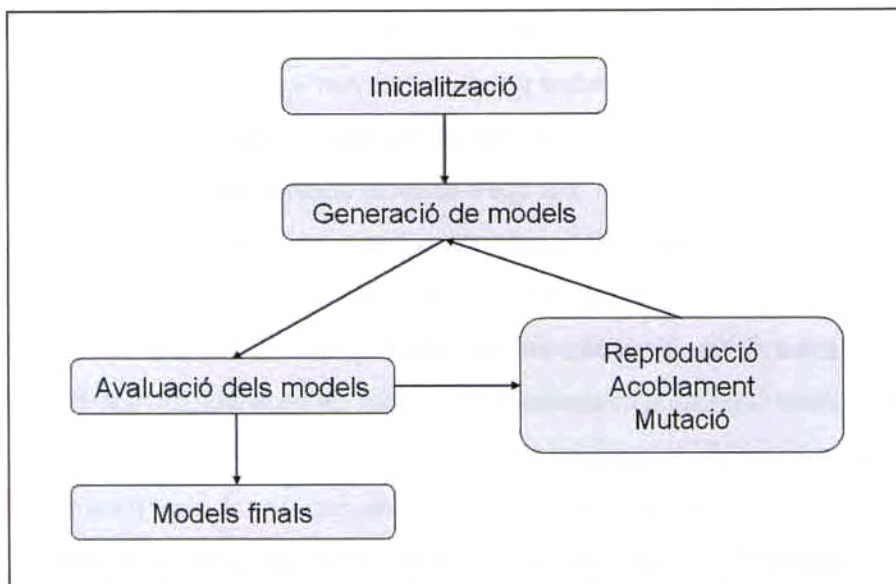


Figura 2.4. Esquema del mètode de selecció de descriptors amb algorismes genètics.

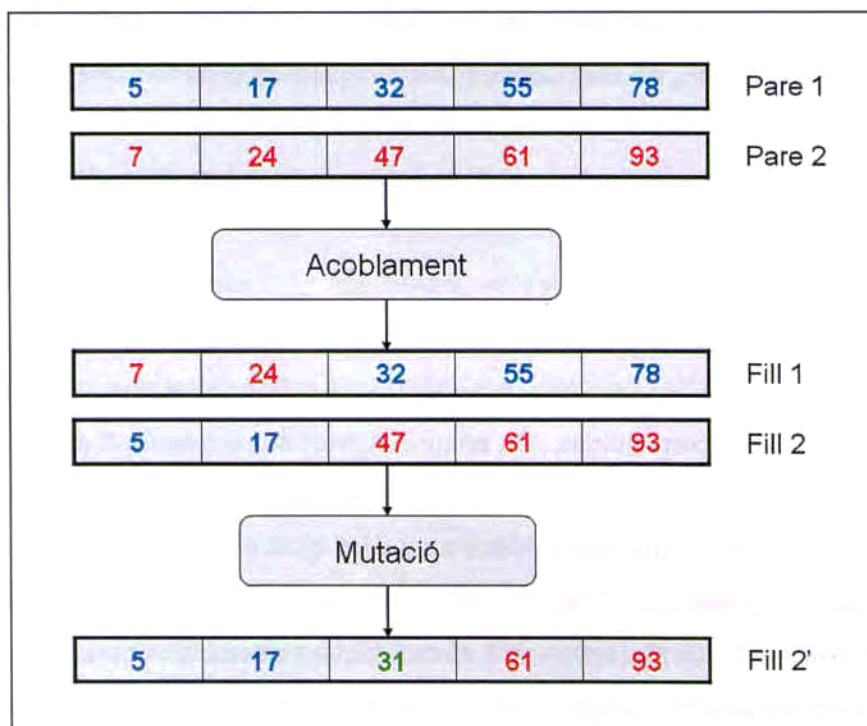


Figura 2.5. Processos d'acoblament i mutació.

L'acoblament es pot entendre com un tall i recombinació dels models inicials (“pares”) per donar lloc a dos models “fills” diferents. El procés de mutació, que es produeix amb relativament poca freqüència, aproximadament un 5% de les vegades, consisteix en la substitució a l'atzar d'un descriptor per un altre i serveix per evitar la convergència prematura del mètode en algun mínim local. Els procediments d'acoblament i mutació se solen realitzar unes mil vegades dins d'un únic procés de selecció de descriptors, assegurant així, que s'obtinguin els millors models possibles.

Selecció amb SVM

Aquest mètode és una nova contribució als mètodes de selecció de descriptors. És, bàsicament, un mètode anàleg a l'heurístic però, en aquest cas, la funció que estableix el model no és la regressió multilíneal sinó una màquina de vectors suport o *Support Vector Machine* (SVM). El mètode de SVM, que s'explica en la següent secció, és relativament recent i, que se sàpiga, és el primer cop que s'utilitza en un procediment de selecció de descriptors. Donat que les correlacions que estableix aquest mètode són no lineals, es pot dir que aquesta selecció de descriptors és també no lineal.

Els passos que s'han de seguir per fer una selecció de descriptors d'aquest tipus són exactament iguals que els del mètode heurístic però, en tots els punts on el mètode heurístic fa servir la regressió multilíneal aquest mètode fa servir la regressió per SVM implementada en el programa LibSVM. Com es veurà més endavant, quan s'expliqui amb més detall en què consisteix el mètode SVM, aquest mètode necessita la fixació per part de l'usuari de tres paràmetres: γ , ϵ i C . En aquest mètode, aquest paràmetres es troben fixats en els valors per defecte del programa ja que, com s'ha comprovat amb diversos exemples, la seva variació no afecta de manera significativa a l'ordre d'importància dels descriptors en el procés de selecció.

Així doncs, aquest mètode, que s'anomena heurístic-SVM, constitueix una aportació nova als mètodes de selecció no lineals i és, a més, el mètode més indicat per triar els descriptors en cas que la correlació que es vulgui establir sigui no lineal i calculada amb SVM.

A2.3. Mètodes per establir models

Un cop coneguts els mètodes per seleccionar els descriptors, el següent pas és establir el model que permet calcular la propietat experimental a partir de l'estructura dels compostos químics. Hi ha diferents mètodes per deduir aquests models, alguns d'ells ja s'han esmentat perquè funcionen paral·lelament i són necessaris en el procés de selecció de descriptors. Tot i així, aquests mètodes són prou importants i estan prou estesos com per merèixer una secció pròpia. Hi ha una gran varietat de mètodes diferents per establir models de correlació; d'entre tots ells es comentaran els que s'han fet servir en aquest treball per construir models lineals: regressió multilineal i PLS; i models no lineals: xarxes neuronals BPNN i SVM. Tots aquests mètodes permeten construir bons models de correlació estructura-propietat però, d'alguna manera, tots presenten pros i contres. Aquests arguments a favor i en contra de cada mètode s'explicaran també en la secció corresponent a cadascun.

Mètodes lineals

Aquests són els mètodes més clàssics i més emprats, gairebé tots es basen en el principi de multilinearitat: una propietat experimental qualsevol pot dependre linealment de diversos paràmetres estructurals alhora i, mitjançant una regressió multiparamètrica i usant el mètode de mínims quadrats es pot trobar la dependència de la propietat en qüestió. Existeix una gran varietat de mètodes per establir correlacions lineals: regressió multilineal (MLR), regressió per mínims quadrats parcials (PLS), anàlisi de components principals (PCA), regressió de components principals (PCR), etc. En aquesta secció es comentaran els dos primers ja que solen ser els més utilitzats.

*Regressió multilineal*⁵⁸

Aquest és el mètode més simple i és el que s'utilitza més habitualment. En general, la regressió multilineal busca establir una relació del tipus:

$$\text{propietat} = a_1d_1 + a_2d_2 + a_3d_3 + \text{terme independent}$$

on d_i són els descriptors inclosos en el model i a_i són els seus coeficients de regressió. El procediment normal en aquest procés és trobar els coeficients de regressió, per tots els descriptors inclosos en la correlació, que minimitza l'error quadràtic del model

global, és a dir, s'ajusten els paràmetres α_{ni} de manera que les diferències entre el valor experimental i calculat de la propietat per cada compost és mínim.

Com ja s'ha dit, aquest mètode és un dels més utilitzats en la metodologia QSPR; això es deu a diversos factors: és un mètode de fàcil accés i de fàcil utilització, permet obtenir estadísticament bons resultats, i, com a punt fort, permet, a través dels signes dels coeficients de regressió, interpretar el sentit de cada descriptor dins del model. En contra d'aquest mètode es pot dir que no funciona excessivament bé quan el conjunt de dades és poc compacte, molt complex o quan el sistema que s'està analitzant és multicomponent.

Regressió PLS⁵⁹

El mètode de regressió PLS (*Partial Least Squares*) és una extensió de la regressió multilineal, en aquest cas la relació que es busca també és del tipus:

$$\text{propietat} = b_1d_1' + b_2d_2' + b_3d_3' + \text{terme independent}$$

on b_i són els seus coeficients de regressió i d_i' són les variables predictives del model. Aquestes variables predictives d_i' no són directament els descriptors sinó que són una combinació lineal dels descriptors estandarditzats. Els resultats que s'obtenen amb el mètode PLS solen ser els mateixos que s'obtenen amb la regressió multilineal però, utilitzant aquest mètode s'amplia la informació que es pot obtenir del model; a més d'interpretar el sentit dels descriptors, es poden associar els descriptors a les diferents parts de les molècules, permetent així, una millor comprensió del mecanisme de la relació estructura-propietat;⁶⁰ a més, els coeficients de regressió estandarditzats donen una idea de la importància relativa de cada descriptor dins del model. De la mateixa manera que la regressió multilineal, el mètode PLS no funciona massa bé amb sistemes complexos o multicomponents.

Mètodes no lineals

Xarxes neuronals^{61,62}

Aquest és un dels mètodes més emprats en la metodologia QSPR per trobar relacions no lineals entre estructura molecular i propietats experimentals. Les xarxes neuronals engloben tot un conjunt de mètodes, models i algorismes dissenyats per mimetitzar el processament de la informació i d'aprenentatge del cervell humà. Així doncs, l'aprenentatge de les xarxes neuronals es fa a través del mateix procediment que el cervell humà: els exemples, proves d'assaig i error, i la seva aplicació en condicions desconegudes però alhora similars.

Les xarxes neuronals més habituals presenten l'esquema mostrat en la Figura 2.1. En aquesta figura es presenta una xarxa neuronal completament connectada de tres capes. L'arquitectura d'una xarxa neuronal es correspon amb el nombre de neurones constituents de cada capa, així doncs, l'arquitectura de la xarxa neuronal d'aquesta figura és 5-3-1. La primera d'aquestes capes, la capa d'entrada, conté els descriptors triats per calcular la propietat experimental; normalment, en aquest punt els descriptors s'escalen per evitar l'efecte dependent de la magnitud dels valors dels descriptors. El nombre de neurones de la capa d'entrada ha de ser forçosament igual al nombre de descriptors triats per establir el model. La tercera capa, que normalment només consta d'una única neurona, és la capa de sortida i és el lloc on s'obté el valor calculat de la propietat experimental. La segona capa, que conté les neurones ocultes, és el centre de càlcul de la xarxa neuronal i és, probablement, la més important de les tres. El nombre de neurones d'aquesta capa i la seva naturalesa varia segons el mètode que es vulgui emprar per establir el model que relaciona estructura i propietat. És en les neurones d'aquesta capa on es dota al sistema de no linealitat i, en cas de necessitat, és possible dotar aquesta capa de diverses subcapes amb diferents nombres de neurones.

Com s'acaba de comentar, la tria del nombre de neurones ocultes i la seva natura depèn del tipus de xarxa neuronal amb la que es vulgui establir el model. D'entre tota la gran varietat de xarxes neuronals existents, els tipus més utilitzats en la metodologia QSPR són dos: les xarxes neuronals amb retropropagació de l'error o BPNN (*Back Propagation Neural Network*) i les xarxes neuronals que utilitzen funcions de base radial; dins d'aquestes últimes es troben les RBFNN (*Radial Basis Function Neural Network*). A continuació s'explicarà, de la manera més planera possible, com es construeixen i com funcionen cadascun d'aquests mètodes de càlcul.

Xarxes neuronals amb retropropagació de l'error o BPNN⁶²

Aquest tipus de xarxes neuronals és, probablement, el més utilitzat en la metodologia QSPR. La seva construcció segueix l'esquema de la Figura 2.1, els descriptors es col·loquen en les neurones de la capa d'entrada; les neurones d'aquesta capa es connecten a les neurones de la capa oculta de manera completa (totes les neurones d'entrada estan connectades a totes les neurones ocultes). Les neurones de la capa oculta funcionen, en aquest cas, com a centres de càlcul independent. Cada neurona oculta rep la informació continguda en els descriptors, modificada amb un pes ajustable i optimitzable, i la processa de la manera que es mostra en la Figura 2.6.

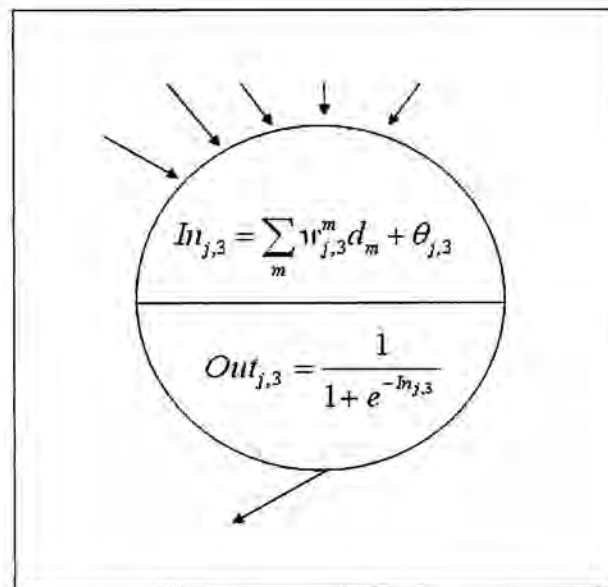


Figura 2.6. Neurona oculta de BPNN.

En aquesta figura es pot veure com la neurona oculta calcula un factor (In) per a cada compost de la següent manera: cada descriptor (d) de la capa d'entrada es corregeix amb el pes de la connexió (w) entre la neurona (m) de la capa d'entrada i la neurona oculta corresponent; se sumen aquests valors dels descriptors "corregits" i se'ls aplica una altra correcció (θ). Un cop es tenen calculats aquests factors In , es procedeix a dotar el sistema de no linealitat; així, cal transformar els aquests factors amb una funció matemàtica extra; en la Figura 2.6 aquesta transformació es realitza amb una funció sigmoide però es poden trobar altres tipus de funcions per realitzar aquest pas:

funcions gaussianes, polinòmiques, tangents i secants hiperbòliques, etc. D'aquesta manera, de cada neurona oculta s'obté un factor *Out* que passa a la neurona de sortida, on serà processat per calcular la propietat experimental. El procediment de càlcul de la propietat en la neurona de sortida és anàleg al de les neurones ocultes però, en aquest cas, no hi ha cap funció transformadora; així, la propietat experimental s'obté amb la següent expressió:

$$Output = \sum_j w'_j Out_j + \theta'$$

on: w'_j és el pes de la connexió entre la neurona oculta i la neurona de sortida, Out_j és el factor de sortida de la neurona oculta corresponent a aquest pes i θ' és una nova correcció.

En aquest tipus de xarxes, l'ensinistrament es realitza ajustant els pesos w de totes les connexions i les correccions θ , comparant, en cada pas, l'error comès per la xarxa respecte el valor experimental. La manera en què s'ajusten aquests paràmetres és la següent: s'assignen uns valors inicials aleatoris per els pesos i les correccions, es calcula la propietat i es compara el seu valor amb l'experimental; l'error comès en el càlcul es retropropaga per les connexions de la xarxa ajustant el valor dels pesos i les connexions per tal de minimitzar aquest error. És a dir, a partir d'un determinat conjunt de pesos i correccions, primer es minimitza l'error ajustant les connexions entre la neurona de sortida i les neurones ocultes i després s'ajusten les connexions entre la capa oculta i la capa d'entrada. L'element avaluador de la bondat de la xarxa neuronal és la funció: $cost = tset + 0.4 * |tset - cvset|$, on $tset$ i $cvset$ són els valors de *rmse* per a aquests dos subconjunts de compostos. Les xarxes neuronals que produeixen errors baixos i de magnituds similars tendeixen a donar bons resultats en la predicció per compostos no inclosos en el desenvolupament del model

Un factor de gran importància en aquest tipus de xarxes és el nombre de neurones ocultes que es fan servir en el càlcul. La manera més habitual de trobar el nombre correcte de neurones ocultes consisteix en fer proves variant el seu nombre en un procediment creixent, és a dir, començant amb un nombre baix de neurones ocultes i anar afegint d'una en una fins que els resultats obtinguts amb una arquitectura determinada no superin els que s'obtenien amb l'arquitectura prèvia. Un altre factor

important en les BPNN és el nombre de paràmetres ajustables (PA). Aquest factor es calcula de la següent manera: $PA = [(NE+1) \times NO] + [(NO+1) \times NS]$, on NE , NO i NS corresponen al nombre de neurones en les capes d'entrada, oculta i de sortida, respectivament. És molt important que el nombre de compostos amb els que s'estableix la xarxa neuronal dividit pel PA sigui més gran que 2.0 ja que, d'aquesta manera, s'evita el sobreaprenentatge. Així doncs, el nombre de neurones ocultes és determinant en aquest punt i cal sempre tenir en compte aquest factor quan s'optimitza l'arquitectura de la xarxa.

Com s'acaba de comentar, un punt a tenir molt en compte quan s'utilitzen aquest tipus de xarxes neuronals és el sobreaprenentatge. Aquest fenomen es dona quan la xarxa neuronal "memoritza" l'ajust dels pesos i les correccions, provocant així que la predicció de compostos que no han estat inclosos en l'ensinistrament de la xarxa neuronal sigui errònia. Donat que la predicció de la propietat de compostos externs és un dels objectius de la metodologia QSPR, cal evitar aquest procés de sobreaprenentatge. Una manera d'evitar-lo és mantenir la relació entre el nombre de compostos en el *tset* amb el nombre de paràmetres ajustables per damunt de 2, una altra manera d'assegurar que no es produirà aquest procés és la utilització d'un subconjunt de compostos amb els quals es controlarà el càlcul. Aquest subconjunt de compostos, que s'anomena *cross-validation set* o *cvset*, s'extreu del conjunt de compostos amb els que s'ha d'establir el model (*tset*) i sol contenir aproximadament un 10% d'aquest. Amb aquest nou subconjunt de compostos s'ajusta el nombre de cicles d'iteració d'optimització de pesos i correccions de la xarxa, mantenint el càlcul en marxa mentre l'error d'aquests compostos disminueixi i aturant el càlcul quan l'error comenci a augmentar.

En general, tots els mètodes que estableixen models no lineals donen millors resultats que els mètodes que empren mètodes lineals. Les BPNN són un cas a remarcar i, probablement, són el mètode que permet obtenir, estadísticament parlant, els millors models tant en correlació com en predicció. D'altra banda, els models que s'estableixen amb aquestes xarxes neuronals són, de vegades, difícils d'interpretar i, a més, la no existència d'una equació que representi el model impedeix conèixer com hi influeix cada descriptor. Un altre desavantatge contrastat d'aquest mètode és l'estancament que pot patir el càlcul en un mínim d'error local, degut principalment a una inicialització de pesos poc afortunada.⁶³

Xarxes neuronals amb funcions de base radial o RBFNN^{64,65}

Aquestes xarxes neuronals són d'aparició més recent que les BPNN. Les diferències entre elles es basen principalment en el tipus de funció de càlcul i transformació que es dona en les neurones ocultes i en la manera de triar el seu nombre, encara que l'esquema general no varia entre unes i altres.

Anàlogament al que s'ha vist per a les BPNN, les neurones ocultes en aquest tipus de xarxes són centres de càlcul independents entre sí; ara bé, un cop aquestes neurones reben la informació continguda en els descriptors, aquesta es processa d'una manera ben diferent. En aquest cas, les neurones ocultes contenen les anomenades "funcions de base", que són les responsables del càlcul intern en cada neurona. Hi ha diversos tipus de funcions de base: gaussianes, multiquadràtiques, cúbiques, etc., d'entre totes elles les que són de major interès en aquest punt són les primeres: les gaussianes (també anomenades funcions de base radial). En la Figura 2.7 es mostra l'esquema de càlcul d'una neurona oculta dins d'una RBFNN.

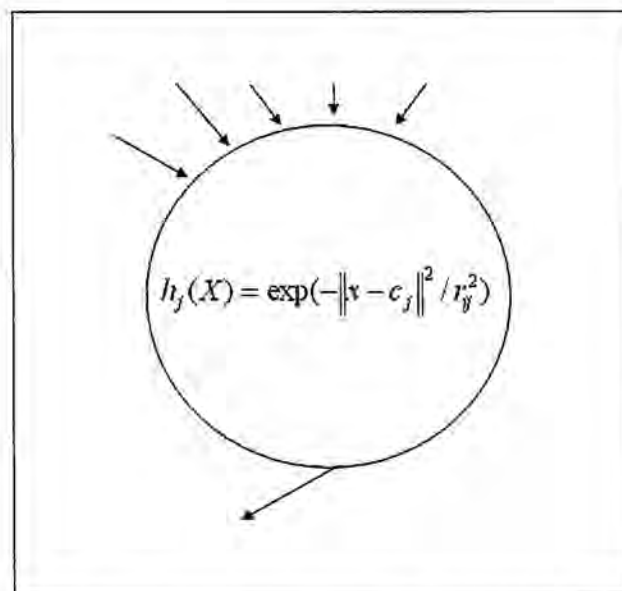


Figura 2.7. Neurona oculta d'una RBFNN.

Els paràmetres que apareixen en la figura són: el factor calculat per la neurona per a un compost qualsevol $h_j(X)$, la distància euclidiana entre el vector de descriptors x

del compost en qüestió i el vector de descriptors c_j d'un compost fixat per la xarxa neuronal per aquesta neurona oculta (també anomenat centre) i l'amplada de la gaussiana continguda en aquesta neurona r_{ij} . Abans de continuar es fa necessari explicar alguns d'aquests paràmetres que s'acaben d'introduir, sobretot el concepte de centre, que és de vital importància per entendre què passa en aquest tipus de xarxes neuronals. Un centre és un compost qualsevol del conjunt del *tset*, és a dir, pot ser qualsevol compost que pertanyi al conjunt d'estructures amb les quals es vol establir el model que relaciona estructura i propietat. Que un compost esdevingui centre en una RBFNN no implica res extra, cal veure-ho només com un artifici que permet calcular la propietat experimental amb la major eficàcia possible. Així doncs, per cada neurona oculta que tingui la RBFNN, es tindrà una funció gaussiana per la qual s'haurà escollit un centre diferent. La manera de triar els centres i el seu nombre es veurà una mica més endavant en aquest mateix apartat.

Un altre paràmetre important que ha aparegut és l'amplada de la gaussiana r_{ij} continguda en cada neurona oculta. Aquest valor, que es correspon a l'amplada de la gaussiana a mitja alçada, és una mesura de les propietats d'interpolació de la funció; com més ampla sigui la gaussiana, més suau serà la interpolació que permetrà.

El procediment pel qual es calcula la propietat en la neurona de sortida és exactament igual al de les BPNN:

$$\text{Output} = \sum_j w_j h_j(X) + \theta'$$

on: w_j és el pes de la connexió entre la neurona oculta i la neurona de sortida, $h_j(X)$ és el factor de sortida de la neurona oculta corresponent a aquest pes i θ' és una correcció.

Així doncs, quan es treballa amb aquest tipus de xarxes neuronals, l'ensinistrament consisteix en les següents etapes:

- 1- Trobar quants i quins centres són necessaris.
- 2- Optimitzar l'amplada de les gaussianes.
- 3- Ajustar els pesos i la correcció de la neurona de sortida.

La determinació dels centres i el seu nombre es pot fer per molts mètodes diferents: aproximacions heurístiques, mètode híbrids, ortonormalització de la funció gaussiana, mínims quadrats ortogonals (RBF-PLS), etc. Potser el millor mètode per fer aquest pas és el mètode *Forward Subset Selection* d'Orr,⁶⁶ aquest mètode va afegint centres d'un en un fins que l'addició d'un centre no comporta una millora en el model. El major avantatge que proporciona aquest mètode és que pot determinar alhora quants centres són necessaris i quins dels compostos ho han de ser. Un cop es té triada l'arquitectura de la xarxa neuronal, el següent pas consisteix en optimitzar l'amplada de les gaussianes contingudes en les neurones ocultes. En general, l'amplada de les gaussianes és igual per a totes les neurones ocultes però, si es vol, és possible optimitzar aquestes amplades individualment. El mètode triat per optimitzar l'amplada de les gaussianes és el de la reducció de l'error *leave-one-out* (veure l'Annex 1), és a dir, es triarà l'amplada que condueixi a la RBFNN a l'error més baix. Quan es tenen tots aquests paràmetres optimitzats, ajustar els pesos i la correcció de la neurona de sortida és relativament senzill mitjançant una regressió multilíneal.

Els resultats que s'obtenen amb aquest tipus de xarxes neuronals són semblants als que es poden obtenir amb les BPNN, normalment, les RBFNN estan millor enfocades en la predicció però això no sempre es compleix. A favor de les RBFNN es pot dir que són capaces d'evitar algun dels problemes irresolubles de les BPNN: no presenten problemes de sobreaprenentatge i, degut al seu mètode de càlcul, les RBFNN no es queden atrapades en mínims d'error local. Tot i així, els desavantatges principals es mantenen: no es disposa d'una equació per al model, no es pot saber com hi influeixen els descriptors i, sovint, la informació que dona el model és complicada d'interpretar.

*SVM (Support Vector Machines)*⁶⁷

Les màquines de vectors suport, SVM, són un mètode no lineal, relativament recent, d'aprenentatge automàtic creat pel grup de Vapnik.⁶⁷ Aquest mètode es va desenvolupar inicialment per resoldre problemes de classificació de compostos químics^{68,69} i després es va estendre per poder solucionar problemes necessitaven aplicacions de regressió ja que la seva capacitat de predicció és molt elevada.⁷⁰⁻⁷² Les SVM consisteixen en construir una espècie de xarxa neuronal on, de la mateixa manera que amb les RBFNN, les neurones ocultes contenen funcions de base que transformen la informació dels descriptors i l'apliquen en la predicció de la propietat experimental.

La diferència entre aquests dos mètodes es troba en el rerefons matemàtic de cadascun d'ells; en el cas de les SVM, les funcions de base projecten la informació continguda en els descriptors en un hiperespai de dimensió superior en el qual les estructures esdevenen més fàcilment separables. El fonament matemàtic d'aquest mètode s'explica en diverses publicacions;⁷³⁻⁷⁵ aquí només s'enunciaran les funcions de base disponibles per a aquest mètode i quins paràmetres cal optimitzar per fer-les servir de manera correcta.

Les funcions de base que es poden emprar quan es construeix una màquina de vectors suport són molt variades: funcions lineals, polinòmiques, sigmoïdes i gaussianes, sent aquestes últimes les més utilitzades; en aquest treball totes les SVM utilitzen funcions de base gaussiana. Quan es tria la funció gaussiana per ser la funció de base de la SVM, el conjunt s'acosta un pas a les RBFNN. Per tal d'entendre una mica millor per què les RBFNN i les SVM són molt semblants caldria explicar què és realment un vector suport. Un vector suport ve constituït pels valors dels descriptors i de la propietat experimental d'un compost contingut en el conjunt de dades amb el qual es vol desenvolupar un model matemàtic per predir una propietat, és a dir, un vector suport en una SVM és equivalent a un centre en una RBFNN. La diferència capital entre aquests dos termes només es deriva de la definició matemàtica triada per escollir-los; en RBFNN el mètode per escollir els centres es basa en triar aquells que minimitzen l'error del conjunt, en SVM el mètode per triar els *vectors suport* és utilitzar un marge de tolerància fix ϵ . Aquest marge de tolerància, que s'explicarà tot seguit, és un dels paràmetres que s'ha d'optimitzar quan es vol utilitzar una SVM en procediments de regressió;⁷⁶ el altres paràmetres a optimitzar són: el paràmetre d'amplada de gaussiana γ i el paràmetre de regularització C .

El paràmetre de regularització C representa el pes dels errors en la predicció i controla l'intercanvi entre maximitzar la predicció i ajustar els valors dels compostos amb els qual s'està construint el model; és a dir, un valor gran per C pot provocar el sobreaprenentatge del model, mentre que un valor petit pot ser causa d'un mal ajust del model. De totes maneres, el valor de C no sembla tenir una influència molt gran en l'error comès en la predicció⁷⁷ i, normalment, un valor al voltant de 100 permet obtenir models estables.

El paràmetre d'amplada de gaussiana és equivalent al r_{ij} que s'ha vist en les RBFNN, el canvi de notació es deu només a un petit canvi en la definició matemàtica:

$$\frac{1}{2r_{ij}^2} = \gamma$$

Per últim, el marge de tolerància ϵ és l'error fix que permet la SVM quan tria els *vectors suport*. Aquest paràmetre, que es pot veure també com l'amplada del tub de radi ϵ al voltant de la recta creada pel model de regressió, és molt important ja que determina quins compostos del conjunt analitzat són *vectors suport*. En la Figura 2.8 es representa esquemàticament aquest paràmetre i la seva influència.

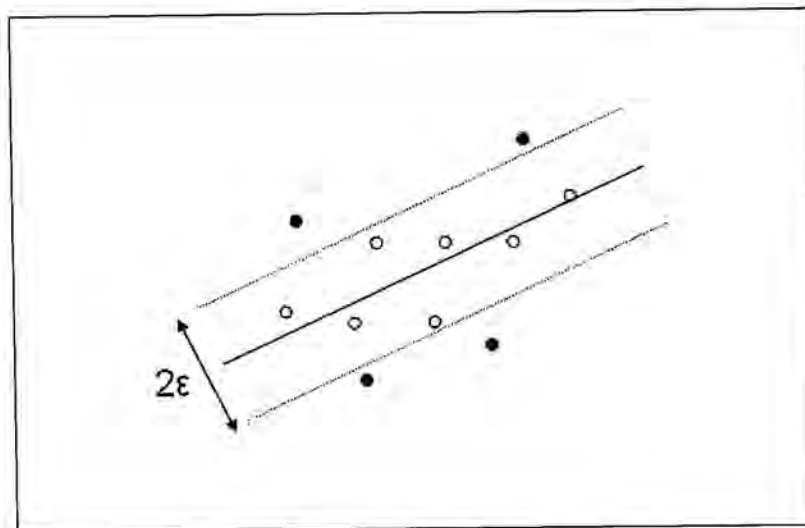


Figura 2.8. Representació del paràmetre ϵ .

En aquesta figura la línia gruixuda representa la recta de regressió, les rectes estretes són els marges del "tub" de radi ϵ i els punts representen els compostos; els de color negre són els *vectors suport* i els altres no ho són. És a dir, només són *vectors suport* aquells compostos que es troben per sobre de l'error marcat per ϵ i, per tant, el nombre de *vectors suport* triats depèn molt directament del valor de ϵ . Cal tenir en compte també que només els vectors suport s'utilitzen a l'hora d'establir el model i, per tant, l'elecció del valor de ϵ és d'una importància crucial. Un valor molt gran d'aquest paràmetre condueix a un nombre petit de *vectors suport* i, probablement, a unes

prediccions molt pobres. L'optimització d'aquests tres paràmetres es fa de manera iterativa utilitzant el mètode de *leave-one-out*, és a dir, es mantenen dos paràmetres fixats i es minimitza l'error del model variant l'altre; normalment s'optimitzen γ , ε i C , per aquest ordre.

Els avantatges i desavantatges derivats de la utilització d'aquest mètode són semblants als de les RBFNN. La capacitat de correlació i predictiva dels models obtinguts és força gran, per contra, es perd la capacitat interpretativa dels descriptors i els seus signes degut a que no es té una equació per al model.

Taula 2.1. Llista de descriptors calculats per CODESSA.

#	Descriptor
1	Nombre d'àtoms
2	Nombre d'àtoms C
3	Nombre relatiu d'àtoms C
4	Nombre d'àtoms H
5	Nombre relatiu d'àtoms H
6	Nombre d'àtoms O
7	Nombre relatiu d'àtoms O
8	Nombre d'àtoms N
9	Nombre relatiu d'àtoms N
10	Nombre d'àtoms S
11	Nombre relatiu d'àtoms S
12	Nombre d'àtoms F
13	Nombre relatiu d'àtoms F
14	Nombre d'àtoms Cl
15	Nombre relatiu d'àtoms Cl
16	Nombre d'àtoms Br
17	Nombre relatiu d'àtoms Br
18	Nombre d'àtoms I
19	Nombre relatiu d'àtoms I
20	Nombre d'àtoms P
21	Nombre relatiu d'àtoms P
22	Nombre d'enllaços
23	Nombre d'enllaços senzills
24	Nombre d'enllaços senzills
25	Nombre d'enllaços dobles
26	Nombre relatiu d'enllaços dobles
27	Nombre d'enllaços triples
28	Nombre relatiu d'enllaços triples
29	Nombre d'enllaços aromàtics
30	Nombre relatiu d'enllaços aromàtics
31	Nombre d'anells
32	Nombre relatiu d'anells
33	Nombre d'anells de benzè
34	Nombre relatiu d'anells de benzè
35	Pes molecular
36	Pes molecular relatiu
37	Índex de gravitació (enllaços)
38	Índex de gravitació (parells)
39	Índex de Wiener
40	Índex de Randić (ordre 0)
41	Índex de Randić (ordre 1)
42	Índex de Randić (ordre 2)
43	Índex de Randić (ordre 3)
44	Índex de Kier & Hall (ordre 0)
45	Índex de Kier & Hall (ordre 1)
46	Índex de Kier & Hall (ordre 2)
47	Índex de Kier & Hall (ordre 3)
48	Índex de forma de Kier (ordre 1)
49	Índex de forma de Kier (ordre 2)
50	Índex de forma de Kier (ordre 3)
51	Índex de flexibilitat de Kier
52	Contingut mitjà d'informació (ordre 0)

#	Descriptor
53	Contingut d'informació (ordre 0)
54	Contingut mitjà d'informació estructural (ordre 0)
55	Contingut d'informació estructural (ordre 0)
56	Contingut mitjà d'informació complementària (ordre 0)
57	Contingut d'informació complementària (ordre 0)
58	Contingut mitjà d'informació d'enllaç (ordre 0)
59	Contingut d'informació d'enllaç (ordre 0)
60	Contingut mitjà d'informació (ordre 1)
61	Contingut d'informació (ordre 1)
62	Contingut mitjà d'informació estructural (ordre 1)
63	Contingut d'informació estructural (ordre 1)
64	Contingut mitjà d'informació complementària (ordre 1)
65	Contingut d'informació complementària (ordre 1)
66	Contingut mitjà d'informació d'enllaç (ordre 1)
67	Contingut d'informació d'enllaç (ordre 1)
68	Contingut mitjà d'informació (ordre 2)
69	Contingut d'informació (ordre 2)
70	Contingut mitjà d'informació estructural (ordre 2)
71	Contingut d'informació estructural (ordre 2)
72	Contingut mitjà d'informació complementària (ordre 2)
73	Contingut d'informació complementària (ordre 2)
74	Contingut mitjà d'informació d'enllaç (ordre 2)
75	Contingut d'informació d'enllaç (ordre 2)
76	Índex de Balaban
77	Moment d'inèrcia A
78	Moment d'inèrcia B
79	Moment d'inèrcia C
80	Ombra XY
81	Ombra XY/rectangle XY
82	Ombra YZ
83	Ombra YZ/rectangle YZ
84	Ombra ZX
85	Ombra ZX/rectangle ZX
86	Volum molecular
87	Volum molecular/caixa XYZ
88	Superfície molecular total
89	Càrrega parcial d'un àtom C
90	Càrrega parcial d'un àtom C
91	Càrrega parcial d'un àtom O
92	Càrrega parcial d'un àtom O
93	Càrrega parcial d'un àtom N
94	Càrrega parcial d'un àtom N
95	Càrrega parcial d'un àtom H
96	Càrrega parcial d'un àtom H
97	Càrrega parcial màxima (Qmax)
98	Càrrega parcial mínima (Qmin)
99	Paràmetre de polaritat (Qmax-Qmin)
100	Paràmetre de polaritat/distància al quadrat
101	Índex de topografia electrònica (parells)
102	Índex de topografia electrònica (enllaços)
103	Àrea superficial molecular total
104	PPSA-1
105	PNSA-1

#	Descriptor
106	DPSA-1
107	FPSA-1
108	FNSA-1
109	WPSA-1
110	WNSA-1
111	PPSA-2
112	PNSA-2
113	DPSA-2
114	FPSA-2
115	FNSA-2
116	WPSA-2
117	WNSA-2
118	PPSA-3
119	PNSA-3
120	DPSA-3
121	FPSA-3
122	FNSA-3
123	WPSA-3
124	WNSA-3
125	RPCG
126	RPCS
127	RNCG
128	RNCS
129	HDSA
130	FHDSA
131	HASA
132	FHASA
133	HBSA
134	FHBSA
135	HDCA
136	FHDCA
137	HACA
138	FHACA
139	HBCA
140	FHBCA
141	Nombre mínim de llocs donadors/acceptors
142	Nombre de llocs donadors d'hidrogen
143	HA dependent HDSA-1
144	HA dependent HDSA-1/TMSA
145	HA dependent HDSA-2
146	HA dependent HDSA-2/TMSA
147	HA dependent HDSA-2/SQRT(TMSA)
148	HA dependent HDCA-1
149	HA dependent HDCA-1/TMSA
150	HA dependent HDCA-2
151	HA dependent HDCA-2/TMSA
152	HA dependent HDCA-2/SQRT(TMSA)
153	HASA-1
154	HASA-1/TMSA
155	HASA-2
156	HASA-2/TMSA
157	HASA-2/SQRT(TMSA)
158	DPSA-1

#	Descriptor
159	HACA-1
160	HACA-1/TMSA
161	HACA-2
162	HACA-2/TMSA
163	HACA-2/SQRT(TMSA)
164	Calor de formació
165	Calor de formació/nombre d'àtoms
166	Nombre d'orbitals moleculars ocupats
167	Nombre d'orbitals moleculars ocupats/nombre d'àtoms
168	Energia HOMO-1
169	Energia HOMO
170	Energia LUMO
171	Energia LUMO+1
172	Diferència d'energia HOMO - LUMO
173	Índex de reactivitat nucleòfila mínim d'un àtom C
174	Índex de reactivitat nucleòfila màxim d'un àtom C
175	Índex de reactivitat nucleòfila mitjà d'un àtom C
176	Índex de reactivitat nucleòfila mínim d'un àtom O
177	Índex de reactivitat nucleòfila màxim d'un àtom O
178	Índex de reactivitat nucleòfila mitjà d'un àtom O
179	Índex de reactivitat nucleòfila mínim d'un àtom N
180	Índex de reactivitat nucleòfila màxim d'un àtom N
181	Índex de reactivitat nucleòfila mitjà d'un àtom N
182	Índex de reactivitat electròfila mínim d'un àtom C
183	Índex de reactivitat electròfila màxim d'un àtom C
184	Índex de reactivitat electròfila mitjà d'un àtom C
185	Índex de reactivitat electròfila mínim d'un àtom O
186	Índex de reactivitat electròfila màxim d'un àtom O
187	Índex de reactivitat electròfila mitjà d'un àtom O
188	Índex de reactivitat electròfila mínim d'un àtom N
189	Índex de reactivitat electròfila màxim d'un àtom N
190	Índex de reactivitat electròfila mitjà d'un àtom N
191	Índex de reactivitat d'un electró mínim d'un àtom C
192	Índex de reactivitat d'un electró màxim d'un àtom C
193	Índex de reactivitat d'un electró mitjà d'un àtom C
194	Índex de reactivitat d'un electró mínim d'un àtom O
195	Índex de reactivitat d'un electró màxim d'un àtom O
196	Índex de reactivitat d'un electró mitjana d'un àtom O
197	Índex de reactivitat d'un electró mínim d'un àtom N
198	Índex de reactivitat d'un electró màxim d'un àtom N
199	Índex de reactivitat d'un electró mitjana d'un àtom N
200	Càrrega neta màxima d'un àtom C
201	Càrrega neta mínima d'un àtom C
202	Càrrega neta màxima d'un àtom O
203	Càrrega neta mínima d'un àtom O
204	Càrrega neta màxima d'un àtom N
205	Càrrega neta mínima d'un àtom N
206	Càrrega neta màxima d'un àtom H
207	Càrrega neta mínima d'un àtom H
208	Càrrega neta màxima
209	Càrrega neta mínima
210	Component direccional carregada del dipol molecular
211	Hibridització total del dipol molecular

#	Descriptor
212	Dipol de la molècula
213	Imatge de l'energia de solvatació d'Onsager-Kirkwood
253	Nombre de llocs acceptors d'hidrogen
275	Població electrònica mínima d'un orbital molecular
276	Població electrònica màxima d'un orbital molecular
277	Ordre d'enllaç sigma-sigma màxim
278	Ordre d'enllaç sigma-pi màxim
279	Ordre d'enllaç pi-pi màxim
280	Màxima contribució enllaçant d'un orbital molecular
281	Màxima contribució antienllaçant d'un orbital molecular
282	València mínima per C
283	València màxima per C
284	València mitjana per C
285	Ordre d'enllaç mínim d'un àtom C
286	Ordre d'enllaç màxim d'un àtom C
287	Ordre d'enllaç mitjà d'un àtom C
288	València mínima d'un àtom O
289	València màxima d'un àtom O
290	València mitjana d'un àtom O
291	Ordre d'enllaç mínim d'un àtom O
292	Ordre d'enllaç màxim d'un àtom O
293	Ordre d'enllaç mitjà d'un àtom O
294	València mínima d'un àtom N
295	València màxima d'un àtom N
296	València mitjana d'un àtom N
297	Ordre d'enllaç mínim d'un àtom N
298	Ordre d'enllaç màxim d'un àtom N
299	Ordre d'enllaç mitjà d'un àtom N
300	València mínima d'un àtom H
301	València màxima d'un àtom H
302	València mitjana d'un àtom H
303	Ordre d'enllaç mínim d'un àtom H
304	Ordre d'enllaç màxim d'un àtom H
305	Ordre d'enllaç mitjà d'un àtom H
306	Repulsió e-e mínima d'un àtom C
307	Repulsió e-e màxima d'un àtom C
308	Atracció e-n mínima d'un àtom C
309	Atracció e-n màxima d'un àtom C
310	Energia atòmica d'estat mínima d'un àtom C
311	Energia atòmica d'estat màxima d'un àtom C
312	Repulsió e-e mínima d'un àtom O
313	Repulsió e-e màxima d'un àtom O
314	Atracció e-n mínima d'un àtom O
315	Atracció e-n màxima d'un àtom O
316	Energia atòmica d'estat mínima d'un àtom O
317	Energia atòmica d'estat màxima d'un àtom O
318	Repulsió e-e mínima d'un àtom N
319	Repulsió e-e màxima d'un àtom N
320	Atracció e-n mínima d'un àtom N
321	Atracció e-n màxima d'un àtom N
322	Energia atòmica d'estat mínima d'un àtom N
323	Energia atòmica d'estat màxima d'un àtom N
324	Repulsió e-e mínima d'un àtom H

#	Descriptor
325	Repulsió e-e màxima d'un àtom H
326	Atracció e-n mínima d'un àtom H
327	Atracció e-n màxima d'un àtom H
328	Energia atòmica d'estat mínima d'un àtom H
329	Energia atòmica d'estat màxima d'un àtom H
330	Energia de ressonància mínima d'un enllaç C-C
331	Energia de ressonància màxima d'un enllaç C-C
332	Energia d'intercanvi mínima d'un enllaç C-C
333	Energia d'intercanvi màxima d'un enllaç C-C
334	Repulsió e-e mínima d'un enllaç C-C
335	Repulsió e-e màxima d'un enllaç C-C
336	Atracció e-n mínima d'un enllaç C-C
337	Atracció e-n màxima d'un enllaç C-C
338	Repulsió n-n mínima d'un enllaç C-C
339	Repulsió n-n màxima d'un enllaç C-C
340	Interacció coulòmbica mínima d'un enllaç C-C
341	Interacció coulòmbica màxima d'un enllaç C-C
342	Interacció total mínima d'un enllaç C-C
343	Interacció total màxima d'un enllaç C-C
344	Energia de ressonància mínima d'un enllaç C-O
345	Energia de ressonància màxima d'un enllaç C-O
346	Energia d'intercanvi mínima d'un enllaç C-O
347	Energia d'intercanvi màxima d'un enllaç C-O
348	Repulsió e-e mínima d'un enllaç C-O
349	Repulsió e-e màxima d'un enllaç C-O
350	Atracció e-n mínima d'un enllaç C-O
351	Atracció e-n màxima d'un enllaç C-O
352	Repulsió n-n mínima d'un enllaç C-O
353	Repulsió n-n màxima d'un enllaç C-O
354	Interacció coulòmbica mínima d'un enllaç C-O
355	Interacció coulòmbica màxima d'un enllaç C-O
356	Interacció total mínima d'un enllaç C-O
357	Interacció total màxima d'un enllaç C-O
358	Energia de ressonància mínima d'un enllaç C-N
359	Energia de ressonància màxima d'un enllaç C-N
360	Energia d'intercanvi mínima d'un enllaç C-N
361	Energia d'intercanvi màxima d'un enllaç C-N
362	Repulsió e-e mínima d'un enllaç C-N
363	Repulsió e-e màxima d'un enllaç C-N
364	Atracció e-n mínima d'un enllaç C-N
365	Atracció e-n màxima d'un enllaç C-N
366	Repulsió n-n mínima d'un enllaç C-N
367	Repulsió n-n màxima d'un enllaç C-N
368	Interacció coulòmbica mínima d'un enllaç C-N
369	Interacció coulòmbica màxima d'un enllaç C-N
370	Interacció total mínima d'un enllaç C-N
371	Interacció total màxima d'un enllaç C-N
372	Energia de ressonància mínima d'un enllaç C-H
373	Energia de ressonància màxima d'un enllaç C-H
374	Energia d'intercanvi mínima d'un enllaç C-H
375	Energia d'intercanvi màxima d'un enllaç C-H
376	Repulsió e-e mínima d'un enllaç C-H
377	Repulsió e-e màxima d'un enllaç C-H

#	Descriptor
378	Atracció e-n mínima d'un enllaç C-H
379	Atracció e-n màxima d'un enllaç C-H
380	Repulsió n-n mínima d'un enllaç C-H
381	Repulsió n-n màxima d'un enllaç C-H
382	Interacció coulòmbica mínima d'un enllaç C-H
383	Interacció coulòmbica màxima d'un enllaç C-H
384	Interacció total mínima d'un enllaç C-H
385	Interacció total màxima d'un enllaç C-H
386	Energia de ressonància mínima d'un enllaç H-O
387	Energia de ressonància màxima d'un enllaç H-O
388	Energia d'intercanvi mínima d'un enllaç H-O
389	Energia d'intercanvi màxima d'un enllaç H-O
390	Repulsió e-e mínima d'un enllaç H-O
391	Repulsió e-e màxima d'un enllaç H-O
392	Atracció e-n mínima d'un enllaç H-O
393	Atracció e-n màxima d'un enllaç H-O
394	Repulsió n-n mínima d'un enllaç H-O
395	Repulsió n-n màxima d'un enllaç H-O
396	Interacció coulòmbica mínima d'un enllaç H-O
397	Interacció coulòmbica màxima d'un enllaç H-O
398	Interacció total mínima d'un enllaç H-O
399	Interacció total màxima d'un enllaç H-O
400	Energia de ressonància mínima d'un enllaç H-N
401	Energia de ressonància màxima d'un enllaç H-N
402	Energia d'intercanvi mínima d'un enllaç H-N
403	Energia d'intercanvi màxima d'un enllaç H-N
404	Repulsió e-e mínima d'un enllaç H-N
405	Repulsió e-e màxima d'un enllaç H-N
406	Atracció e-n mínima d'un enllaç H-N
407	Atracció e-n màxima d'un enllaç H-N
408	Repulsió n-n mínima d'un enllaç H-N
409	Repulsió n-n màxima d'un enllaç H-N
410	Interacció coulòmbica mínima d'un enllaç H-N
411	Interacció coulòmbica màxima d'un enllaç H-N
412	Interacció total mínima d'un enllaç H-N
413	Interacció total màxima d'un enllaç H-N
414	Repulsió E-N molecular d'un centre
415	Repulsió E-N molecular d'un centre/nombre d'àtoms
416	Repulsió E-E molecular d'un centre
417	Repulsió E-E molecular d'un centre/nombre d'àtoms
418	Energia de ressonància molecular
419	Energia de ressonància molecular/nombre d'àtoms
420	Energia d'intercanvi molecular
421	Energia d'intercanvi molecular/nombre d'àtoms
422	Interacció electrostàtica molecular
423	Interacció electrostàtica molecular/nombre d'àtoms
424	Moment principal d'inèrcia A
425	Moment principal d'inèrcia A/nombre d'àtoms
426	Moment principal d'inèrcia B
427	Moment principal d'inèrcia B/nombre d'àtoms
428	Moment principal d'inèrcia C
429	Moment principal d'inèrcia C/nombre d'àtoms
430	α Polaritzabilitat

#	Descriptor
431	β Polaritzabilitat
432	(1/2) β Polaritzabilitat
433	γ Polaritzabilitat
434	(1/6) γ Polaritzabilitat
435	Càrrega parcial d'un àtom S
436	Càrrega parcial d'un àtom S
437	Índex de reactivitat nucleòfila mínim d'un àtom S
438	Índex de reactivitat nucleòfila màxim d'un àtom S
439	Índex de reactivitat nucleòfila mitjà d'un àtom S
440	Índex de reactivitat electròfila mínim d'un àtom S
441	Índex de reactivitat electròfila màxim d'un àtom S
442	Índex de reactivitat electròfila mitjà d'un àtom S
443	Índex de reactivitat d'un electró mínim d'un àtom S
444	Índex de reactivitat d'un electró màxim d'un àtom S
445	Índex de reactivitat d'un electró mitjà d'un àtom S
446	Càrrega neta màxima d'un àtom S
447	Càrrega neta mínima d'un àtom S
448	València mínima d'un àtom S
449	València màxima d'un àtom S
450	València mitjana d'un àtom S
451	Ordre d'enllaç mínim d'un àtom S
452	Ordre d'enllaç màxim d'un àtom S
453	Ordre d'enllaç mitjà d'un àtom S
454	Repulsió e-e mínima d'un àtom S
455	Repulsió e-e màxima d'un àtom S
456	Atracció e-n mínima d'un àtom S
457	Atracció e-n màxima d'un àtom S
458	Energia atòmica d'estat mínima d'un àtom S
459	Energia atòmica d'estat màxima d'un àtom S
460	Energia de ressonància mínima d'un enllaç C-S
461	Energia de ressonància màxima d'un enllaç C-S
462	Energia d'intercanvi mínima d'un enllaç C-S
463	Energia d'intercanvi màxima d'un enllaç C-S
464	Repulsió e-e mínima d'un enllaç C-S
465	Repulsió e-e màxima d'un enllaç C-S
466	Atracció e-n mínima d'un enllaç C-S
467	Atracció e-n màxima d'un enllaç C-S
468	Repulsió n-n mínima d'un enllaç C-S
469	Repulsió n-n màxima d'un enllaç C-S
470	Interacció coulòmbica mínima d'un enllaç C-S
471	Interacció coulòmbica màxima d'un enllaç C-S
472	Interacció total mínima d'un enllaç C-S
473	Interacció total màxima d'un enllaç C-S
474	Energia de ressonància mínima d'un enllaç H-S
475	Energia de ressonància màxima d'un enllaç H-S
476	Energia d'intercanvi mínima d'un enllaç H-S
477	Energia d'intercanvi màxima d'un enllaç H-S
478	Repulsió e-e mínima d'un enllaç H-S
479	Repulsió e-e màxima d'un enllaç H-S
480	Atracció e-n mínima d'un enllaç H-S
481	Atracció e-n màxima d'un enllaç H-S
482	Repulsió n-n mínima d'un enllaç H-S
483	Repulsió n-n màxima d'un enllaç H-S

#	Descriptor
484	Interacció coulòmbica mínima d'un enllaç H-S
485	Interacció coulòmbica màxima d'un enllaç H-S
486	Interacció total mínima d'un enllaç H-S
487	Interacció total màxima d'un enllaç H-S

2.3. Bibliografia

1. PCMODEL, Serena Software, Bloomington, IN, **1996-2005**.
2. SYBYL 7.0, Tripos Inc., 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA.
3. Wavefunction Inc., Spartan 5.0, Irvine, CA, **1995**.
4. HyperChem 7.5, HyperCube Inc., **2002**.
5. Stewart, J. P. P. *MOPAC 6.0 Quantum Chemistry Program Exchange*; QCPE, No. 455, Indiana University, Bloomington, IN, **1989**.
6. Stewart, J. P. P. *MOPAC 7.0 Quantum Chemistry Program Exchange*; QCPE, No. 688, Indiana University, Bloomington, IN, **1991**.
7. Bosque, R. Manual Aux-QSPR versió 0.9, Barcelona, **2006**.
8. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, 107, 3902.
9. Stewart, J. J. P. *J. Comput. Chem.* **1989**, 10, 209.
10. Gaussian 03, Revision 6.0, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Pittsburgh PA, **2003**.
11. Karelson, M.; Maran, U.; Wang, Y.; Katritzky, A. R. *Collect. Czech. Chem. Commun.* **1999**, 64, 1551.
12. Katritzky, A. R.; Maran, U.; Lobanov, V. S.; Karelson, M. *J. Chem. Inf. Comput. Sci.* **2000**, 40, 1.

13. Katritzky, A. R.; Fara, D. C.; Petrukhin, R. O.; Tatham, D. B.; Maran, U.; Lomaka, A.; Karelson, M. *Current Topics in Medicinal Chemistry* **2002**, 2, 1333.
14. Jurs, P. C. in *Handbook of Chemoinformatics – From Data to Knowledge*, Gasteiger, J. A.; Engel, T.; Eds, Wiley-VCH, **2003**.
15. Gasteiger, J. A. in *Handbook of Chemoinformatics – From Data to Knowledge*, Gasteiger, J. A.; Engel, T.; Eds, Wiley-VCH, **2003**.
16. Famini, G. R.; Wilson, L. Y. *Reviews in Computational Chemistry*, Lipkowitz, K. B.; Boyd, D.B.; Eds, vol 18, Wiley-VCH, **2002**.
17. Todeschini, R.; Consonni, V. in *Handbook of Chemoinformatics – From Data to Knowledge*, Gasteiger, J. A.; Engel, T.; Eds, Wiley-VCH, **2003**.
18. Balaban, A. T. *From Chemical Topology to Three-Dimensional Geometry*, Plenum Press: New York, **1996**.
19. Devillers, J.; Balaban, A. T. *Topological Indices and Related Descriptors in QSAR and QSPR*, Gordon & Breach: Amsterdam, **1999**.
20. Ivanciuc, O. in *Handbook of Chemoinformatics – From Data to Knowledge*, Gasteiger, J. A.; Engel, T.; Eds, Wiley-VCH, **2003**.
21. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*, Academic Press: New York, **1976**.
22. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure-Activity Analysis*, John Wiley & Sons: New York, **1986**.
23. Estrada, E.; Uriarte, E. *Curr. Med. Chem.* **2001**, 8, 1699.
24. Estrada, E.; Molina, E. *J. Chem. Inf. Comput. Sci.* **2001**, 41, 791.
25. Livingstone, D. J. *J. Chem. Inf. Comput. Sci.* **2000**, 40, 195.
26. Karelson, M. *Molecular Descriptors in QSAR/QSPR*, John Wiley & Sons: New York, **2000**.
27. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*, Wiley-VCH: Weinheim, **2000**.
28. Katritzky, A. R.; Lovanov V. S.; Karelson, M. *CODESSA, Reference Manual V 2.13*, Semichem and the University of Florida, **1997**.
29. Jurs, P. C.; Chow, J. T.; Yuan, M. in *Computer-Assisted Drug Design*, Olson, E. C.; Christorffersen, R. E.; Eds.; The American Chemical Society: Washington, DC, **1979**.
30. Stuper, A. J.; Brugger, W. E.; Jurs, P. C. *Computer-Assisted Studies of Chemical Structure and Biological Function*, Wiley: New York, **1979**.

31. Todeschini, R.; Consonni, V.; Mauri, A.; Pavan, M. *Dragon Academic Version – Software Version 5.2*, Milano Chemometrics & QSAR Research Group, Milano, **2004**.
32. Balaban, A. T. *J. Chem. Inf. Comput. Sci.* **1997**, 37, 645.
33. Balaban, A. T. *SAR & QSAR Environ. Res.* **1998**, 8, 1.
34. Merrifield, R. E.; Simmons, H. E. *Topological Methods in Chemistry*, John Wiley & Sons: New York, **1989**.
35. Bonchev, D. *Information Theoretic Indices for Characterization of Chemical Structure*, Wiley-Interscience: New York, **1983**.
36. Rouvray, D. H. *Computational Chemical Graph Theory*, Nova Science Publishers: New York, **1990**.
37. Zefirov, N. S.; Kirpichenok, M. A.; Izmailov, F. F.; Trofimov, M. I. *Dokl. Akad. SSSR* **1987**, 296, 883.
38. Sanderson, R. T. en: *Encyclopedia of Chemistry*, 3^a Ed., Hampel, C. A.; Hawley, G. G.; Eds.; Van Nostrand Reinhold: New York, **1973**, 385.
39. Stanton, D. T.; Jurs, P. C. *Anal. Chem.* **1990**, 62, 2323.
40. Stanton, D. T.; Egolf, L. M.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **1992**, 32, 306.
41. Sklenar, H.; Jäger, J. *Int. J. Quant. Chem.* **1979**, 16, 467.
42. Klopman, G.; Iroff, L. D. *J. Comput. Chem* **1981**, 2, 157.
43. Buydens, L.; Massart, D.; Geerlings, P. *Anal. Chem.* **1983**, 55, 738.
44. Franke, R. *Theoretical Drug Design Methods*, Elsevier: Amsterdam, **1984**.
45. Cartier, A.; Rivail, J.-L. *Chemom. & Int. Lab. Systems* **1987**, 1, 335.
46. Bodor, N.; Gabanyi, Z.; Wong, C.-K. *J. Am. Chem. Soc.* **1989**, 111, 3783.
47. Ford, M. G.; Livingstone, D. J. *Quant. Struc.-Act. Relat.* **1990**, 9, 107.
48. Tuppurainen, K.; Löjtönen, S.; Laatikainen, R.; Vartiainen, T.; Maran, U.; Strnadberg, M.; Tamm, T. *Mutation Research* **1991**, 97, 247.
49. Pople, J. A.; Beveridge, D. L. *Approximate Molecular Orbital Theory*, McGraw Hill: New York, **1970**.
50. Fukui, K. *Theory of Orientation and Stereoselection*, Springer-Verlag: Berlin, **1975**.
51. Sannigrahi, A. B. *Adv. Quant. Chem.* **1992**, 23, 301.
52. Emsley, J. *The Elements*, Clarendon Press, **1989**.
53. Chang, C.-C.; Lin, C.-J. *LIBSVM : a library for support vector machines*, **2001**.
Software disponible a <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
54. Matlab, versió 6.5, The MathWorks Inc., **1984-2002**.

55. SPSS 14.0 per Windows, versió 14.0.1, SPSS Inc., **1989-2005**.
56. Xu, L.; Zhang, W. *Anal. Chim. Acta* **2001**, 446, 475.
57. Rogers, D.; Hopfinger, A. J. *J. Chem. Inf. Comput. Sci.* **1994**, 34, 854.
58. Myers, R. H. *Classical and Modern Regression with Applications*, PWS-Kent Publ. Co.: Boston, **1990**.
59. Geladi, P.; Kowalski, B. *Anal. Chim. Acta* **1986**, 185, 1.
60. Stanton, D. T. *J. Chem. Inf. Comput. Sci.* **2003**, 43, 1423.
61. Devillers, J. Ed; *Neural Networks in QSAR and Drug Design*, Academic Press: London, **1996**.
62. Zupan, J.; Gasteiger, J. A. *Neural Networks for Chemistry and Drug Design*, Wiley-VCH: Weinheim, **1997**.
63. Agrafiotis, D. K.; Cedeno, W. *J. Med. Chem.* **2002**, 45, 1098.
64. Orr, M. J. L. *Introduction to Radial Basis Function Networks*, Centre for Cognitive Science, Edinburgh University, **1996**.
65. Walczak, B. Massart, D. L. *Chemom. Int. Lab. Syst.* **2000**, 50, 179.
66. Orr, M. J. L. *Matlab Routines for Subset Selection and Ridge Regression in Linear Neural Networks*, Centre for Cognitive Science, Edinburgh University, **1996**.
67. Vapnik, N. *Statistical Learning Theory*, John Wiley & Sons: New York, **1998**.
68. Hsu, C.-W.; Lin, C.-J. *IEEE Transactions on Neural Networks* **2002**, 12, 181.
69. Belousov, A. I.; Verzakov, S. A.; von Frese, J. *Chemom. Int. Lab. Syst.* **2002**, 64, 15.
70. Xue, C. X.; Liu, H. X.; Yao, X. J.; Liu, M. C.; Hu, Z. D.; Fan, B. T. *J. Chromatogr. A* **2004**, 1048, 233.
71. Xue, C. X.; Zhang, R. S.; Liu, H. X.; Yao, X. J.; Liu, M. C.; Hu, Z. D.; Fan, B. T. *J. Chem. Inf. Comput. Sci.* **2004**, 44, 1267.
72. Zhao, C. Y.; Zhang, H. X.; Zhang, X. Y.; Liu, M. C.; Hu, Z. D.; Fan, B. T. *Toxicology* **2006**, 217, 105.
73. Burbridge, R.; Trotter, M.; Buxton, B.; Holden, S. *Computers and Chemistry* **2001**, 26, 5.
74. Czermínski, R.; Yasri, A.; Hartsough, D. *Quant. Struc.-Act. Relat.* **2001**, 20, 227.
75. Amendolia, S. R.; Cossu, G.; Ganadu, M. L.; Golosio, B.; Masala, G. L.; Mura, G. *M. Chemom. Int. Lab. Syst.* **2003**, 69, 13.
76. Bishop, C. *Neural Networks for Pattern Recognition*, Clarendon Press: Oxford, **1997**.

77. Liu, H. X.; Zhang, R. S.; Yao, X. J.; Liu, M. C.; Hu, Z. D.; Fan, B. T. *J. Chem. Inf. Comput. Sci.* **2003**, 43, 1288.

78. Guha, R.; Jurs, P. C. *J. Chem. Inf. Model.* **2005**, 45, 800.

Capítol 3. Compostos inorgànics

Tal i com s'ha comentat en la introducció d'aquesta Memòria, la metodologia QSPR ha estat aplicada àmpliament i amb èxit en l'estimació d'una quantitat important de propietats físiques i químiques de compostos orgànics. En canvi, el seu ús amb compostos inorgànics es limitava, en la pràctica, a algunes publicacions sobre propietats atòmiques, com els potencials d'ionització¹ i de molècules diatòmiques^{2,3} com la distància d'enllaç o l'energia de dissociació. Recentment s'ha començat a aplicar aquest tipus de metodologia a alguns conjunts de compostos organometàl·lics per tal de predir propietats com la toxicitat en éssers vius: algues, bacteries i peixos,⁴⁻⁸ les constants de complexació de diferents metalls amb transportadors d'ions,⁹ i la capacitat per actuar com a catalitzadors en sistemes homogenis.¹⁰⁻¹⁶

Aquest lent desenvolupament és degut principalment a dues raons: les limitacions en la capacitat computacional i la manca de dades experimentals per a conjunts d'aquest tipus de compostos prou grans. La primera d'aquestes raons està, actualment, minimitzada gràcies al gran desenvolupament dels ordinadors, dels mètodes de càlcul i d'optimització disponibles; però la falta de conjunts prou nombrosos de compostos moleculars inorgànics és de fet intrínseca i quasi característica de la Química Inorgànica. Ja que els mètodes QSPR requereixen disposar d'un conjunt de compostos suficientment gran i d'estructures no excessivament dispars, l'ús de compostos organometàl·lics d'elements del bloc *p* i de compostos de coordinació d'estructures senzilles és el més evident i immediat.

Aquest capítol es divideix en dos blocs, en el primer s'estudien diferents propietats com són la viscositat, η , la tensió superficial, σ , i l'entalpia de formació, $\Delta_f H^p$, de compostos organometàl·lics, i en el segon la basicitat catiònica d'un conjunt de molècules orgàniques \bar{L} front el liti, és a dir, l'energia involucrada en el procés de formació dels compostos de coordinació del tipus $[\text{Li}\bar{L}]^+$. Cal recordar, però, que la definició de compost organometàl·lic s'ha agafat d'una manera àmplia i així doncs, en alguns dels conjunts de dades analitzats s'hi troben compostos derivats d'elements semimetàl·lics i fins i tot no metàl·lics del bloc *p* que, podrien considerar-se derivats orgànics; a més, s'estudien també, alguns derivats que no presenten enllaços directes M-C. Aquests compostos "organometàl·lics" es mantenen en el conjunt estudiat perquè el seu comportament i estructura molecular no difereixen de manera significativa dels de

la resta de substàncies examinades. En l'estudi de la viscositat i la tensió superficial, s'incorpora l'efecte de la temperatura com a descriptor extern als sistemes, fet que permet estimar aquestes propietats a diferents temperatures. Així doncs, en aquest capítol, es presenten els resultats que s'obtenen en aplicar la metodologia QSPR a aquests quatre sistemes, els quals es tracten de manera individual en els següents apartats.

En l'estudi d'aquests conjunts s'ha seguit l'esquema QSPR habitual, és a dir s'ha proposat el model a partir, exclusivament, dels descriptors moleculars dels compostos estudiats. En el cas concret de l'estimació de les entalpies de formació s'han emprat també descriptors dels metalls, com a descriptors externs. Com a norma general, primer s'ha deduït un model lineal i a continuació de no lineals, mitjançant xarxes BPNN i SVM; en algun cas s'ha fet ús també de l'anomenat mètode híbrid.

3.1 Viscositat de compostos organometàl·lics

La viscositat dels líquids és una de les propietats superficials i de transport més importants en molts estudis científics i en aplicacions d'enginyeria, sobretot aquelles que impliquen el moviment de líquids. Altres propietats de transport significatives són la tensió superficial, la conductivitat tèrmica i el coeficient de difusió; la seva magnitud depèn, en tots els casos, de les interaccions intermoleculares que es produeixen en el si del líquid. Aquestes propietats no es consideren propietats termodinàmiques ja que descriuen sistemes que no es troben en equilibri.

La viscositat d'un líquid es defineix com la resistència a fluir que aquest presenta quan se li aplica una força tangencial:

$$\eta = \frac{\tau}{(dv/dx)}$$

on τ és l'esforç tangencial i (dv/dx) és la velocitat relativa del desplaçament produït. Així, els líquids amb viscositat baixa tendeixen a escampar-se, com l'aigua; mentre que els líquids amb elevada viscositat no presenten aquesta tendència.

Aquesta propietat, que és característica de qualsevol fluid líquid o gasós, presenta una marcada dependència de la temperatura sent, en el cas de substàncies líquides, menor en augmentar la temperatura. En general, aquesta dependència de la temperatura té la forma:

$$\eta = A \cdot e^{B/T}$$

$$\ln \eta = \ln A + \frac{B}{T}$$

on A i B són característics de cada compost.

La viscositat dels compostos químics es determina de manera experimental mitjançant l'ús de tres tipus diferents de viscosímetres: de capil·laritat, de cos oscil·lant i de cos caient. El procediment experimental de mesura sol comportar l'ajust de certs paràmetres i el calibratge de l'aparell abans de cada mesura. La importància d'aquesta propietat en aplicacions industrials ve demostrada per la gran quantitat de compendis de valors experimentals que han anat apareixent al llarg del temps.¹⁷⁻²⁵ Per tal d'evitar la

lenta i sovint poc acurada (per la dificultat que comporta) determinació experimental, en els últims anys s'han desenvolupat diferents mètodes que permeten l'estimació d'aquesta propietat.

La metodologia QSPR ha jugat un paper molt important dins d'aquests mètodes d'estimació i es poden trobar diferents estudis, normalment per a conjunts de compostos orgànics, que correlacionen la viscositat amb diferents paràmetres. Els primers estudis empenen els mètodes de contribucions de grups.^{26,27} Els estudis de Suzuki^{28,29} estableixen models per correlacionar la viscositat amb descriptors derivats de l'estructura i amb altres propietats dels compostos analitzats. Altres treballs utilitzen paràmetres deduïts a través de mesures de RMN³⁰ o per teories de grafs³¹ com a descriptors en els seus models per estimar la viscositat de mostres de dièsel i de cicloalcans, respectivament. La resta de treballs trobats mitjançant la recerca bibliogràfica utilitzen descriptors derivats directament de l'estructura molecular que permeten calcular la viscositat de diferents conjunts de compostos orgànics. Ivanciuc i col·laboradors³² van establir, mitjançant regressió multilíneal amb un 20% de validació creuada, un model de cinc descriptors per a un conjunt de 337 compostos diferents. De manera similar, el grup de Katritzky³³ va establir un model lineal de cinc descriptors per a un conjunt de 361 compostos orgànics de naturalesa química molt variada, formats solament per carboni, hidrogen, nitrogen, oxigen, sofre i halògens. En aquest model es va emprar un mètode de validació creuada d'un terç. Més endavant, aquest mateix grup d'investigació va fer una ampliació de l'estudi mitjançant altres mètodes lineals i no lineals.³⁴ Cocchi i col·laboradors³⁵ van desenvolupar un model per a 46 compostos orgànics amb dos, tres i quatre descriptors amb mètodes lineals i de regressió PLS. El grup de Jurs³⁶ va construir models no lineals de vuit i nou descriptors per predir, entre d'altres propietats, la viscositat d'un conjunt de 213 compostos. Més recentment, Zefirov i col·laboradors³⁷ han desenvolupat un mètode no lineal que permet calcular la viscositat de 367 compostos orgànics a partir de descriptors derivats del recompte dels diferents fragments constitucionals que formen les molècules. En general, els resultats que s'obtenen en totes aquestes publicacions són bastant bons, amb uns coeficients de determinació R^2 entre 0.85 i 0.95, i uns errors al voltant de 0.2 unitats logarítmiques.

En aquest capítol s'aplicarà la metodologia QSPR per estimar, per primera vegada, la viscositat en fase líquida d'un conjunt de compostos organometàl·lics amb models lineals i no lineals. En primera instància, s'han proposat models que permeten el càlcul d'aquesta propietat a temperatures al voltant de 20°C; el fet de disposar de valors experimentals per a, pràcticament, els mateixos compostos a diferents temperatures, ha permès ampliar l'estudi establint models que permeten calcular la viscositat a temperatura variable, considerant el seu valor com a un descriptor extern. Aquesta metodologia s'ha trobat documentada en diversos treballs en els quals s'utilitza la temperatura com a descriptor extern en estudis de conjunts de compostos orgànics,³⁸⁻⁴⁰ algunes de les propietats així estudiades són l'índex de refracció, la constant dielèctrica, el moment dipolar i la mateixa viscositat.

En resum, en aquest capítol es presenten els models per a dos conjunts de compostos organometàl·lics, el primer conté valors de viscositat de 205 compostos a temperatures al voltant dels 20°C, i el segon conté 534 valors experimentals, fruit de la combinació dels 205 compostos del primer conjunt determinats a diferents temperatures. Els models deduïts per als dos conjunts són lineals i no lineals, els lineals s'obtenen mitjançant regressió multilinear i els no lineals, que incorporen les seves pròpies seleccions de descriptors, s'estableixen amb BPNN i SVM. Es pretén que els models trobats continguin descriptors que, a més de permetre el càlcul de la propietat, aportin informació fisicoquímica relacionada amb el procés estudiat; d'aquesta manera, és d'esperar que els descriptors triats recullin informació relacionada amb les possibles interaccions intermoleculares que es donen en el líquid i que són la base de què depèn directament el valor de viscositat dels compostos químics.

3.1.1 Viscositat a 20°C

Conjunt de dades

El conjunt de valors de la viscositat en fase líquida, ha estat extret de la base de dades Landolt-Börnstein.²⁵ En total, s'han trobat 205 valors experimentals per a 9 compostos d'arsènic, 14 de germani, 124 de silici, 9 d'estany i 49 de fòsfor. Aquests compostos tenen la fórmula general: MR_nX_m , on R pot ser un grup alquil, alcoxi, aril, ariloxi, amino i alquenil, i, per la seva part, el grup X pot ser F, Cl, Br, I, OH i NCO. La unitat de la viscositat en el sistema internacional és el Pa·s, encara que de manera habitual es dona en mPa·s i en forma logarítmica. Així doncs, la base de dades construïda conté 205 valors de $\log \eta$ que cobreixen un rang de valors que va des de -0.616 per al SiMe_4 fins a 1.221 per al $\text{Sn}(\text{O}(\text{1,1-Me}_2)\text{Pr})_4$, amb un valor mitjà de 0.255; el rang de temperatures d'aquestes mesures es troba entre 15 i 30°C però cal dir que la majoria d'elles, 195, ho són entre 20 i 25°C. Tots els valors experimentals i calculats es recullen en la Taula 3.1 junt amb la temperatura exacta de la determinació experimental per a cada compost.

Els conjunts de dades del *training* i *prediction sets*, els quals s'empraran en l'establiment de models lineals i no lineals amb SVM, contenen 163 i 42 valors experimentals, respectivament. Per tal d'establir els models no lineals amb BPNN és necessari afegir el *cvset*, que conté un 10% dels compostos del *tset*. D'aquesta manera els tres subconjunts utilitzats per construir els models no lineals amb BPNN: *tset*, *cvset* i *pset*, contenen 147, 16 i 42 valors, respectivament.

En aquest cas, el mètode utilitzat per optimitzar la geometria dels compostos ha estat el PM3 ja que la falta de parametrització de diversos elements ha fet impossible utilitzar el mètode AM1.

Model lineal

Els models lineals es dedueixen mitjançant regressió multilíneal amb el mètode heurístic implementat per CODESSA. El nombre de descriptors calculats inicialment és de 968 i amb les rutines de selecció objectiva incorporades per aquest mètode, es redueix fins a 200.

El millor model de set descriptors obtingut amb la regressió multilíneal per als 163 compostos que formen el *tset* es mostra en la Taula 3.2.

Taula 3.2. Model obtingut per regressió lineal.

Descriptor	coeficient	<i>sd</i>	β	test- <i>t</i>	VIF
Terme independent	-0.597	0.084	-	-7.12	-
Nombre d'àtoms d'oxigen	-0.112	0.014	-0.43	-8.11	3.61
Nombre relatiu d'anells de benzè	8.662	1.034	0.36	8.38	2.37
HACA-2/TMSA	126.06	15.05	0.23	8.37	1.01
Àrea superficial positiva fraccionada, FPSA-2	0.398	0.029	0.74	13.51	3.84
Diferència entre superfícies carregades, DPSA-3	0.024	0.004	0.33	6.21	3.67
Àrea superficial negativa ponderada, WNSA-1	0.003	0.001	0.19	3.37	4.12
Càrrega positiva relativa, RPCG	-0.624	0.104	-0.25	-5.99	2.18

La correlació entre els set descriptors es troba en el rang 0.01-0.73, amb un valor mitjà de 0.23. El model presenta un valor VIF mitjà de 2.97, que indica poca col·linealitat entre descriptors. Els descriptors més importants dintre la correlació, segons els valors de β , són FPSA-2, el nombre d'àtoms d'oxigen, el nombre relatiu d'anells de benzè i DPSA-3.

La Taula 3.3 mostra els resultats que s'obtenen amb aquest model per als subconjunts estudiats *tset* i *pset*. El valor de R^2_{cv} trobat per a aquest model és 0.865, molt pròxim al valor del coeficient de determinació dels compostos del *tset* i indicatiu de l'estabilitat del model. Els valors de viscositat calculats amb aquest model es troben en la Taula 3.1 i la seva representació gràfica enfront dels valors experimentals es pot veure en la Figura 3.1.

Taula 3.3. Paràmetres estadístics obtinguts amb el model lineal.

Conjunt	<i>n</i>	R^2	F	<i>rmse</i>
<i>tset</i>	163	0.879	1174.5	0.150
<i>pset</i>	42	0.914	425.9	0.134

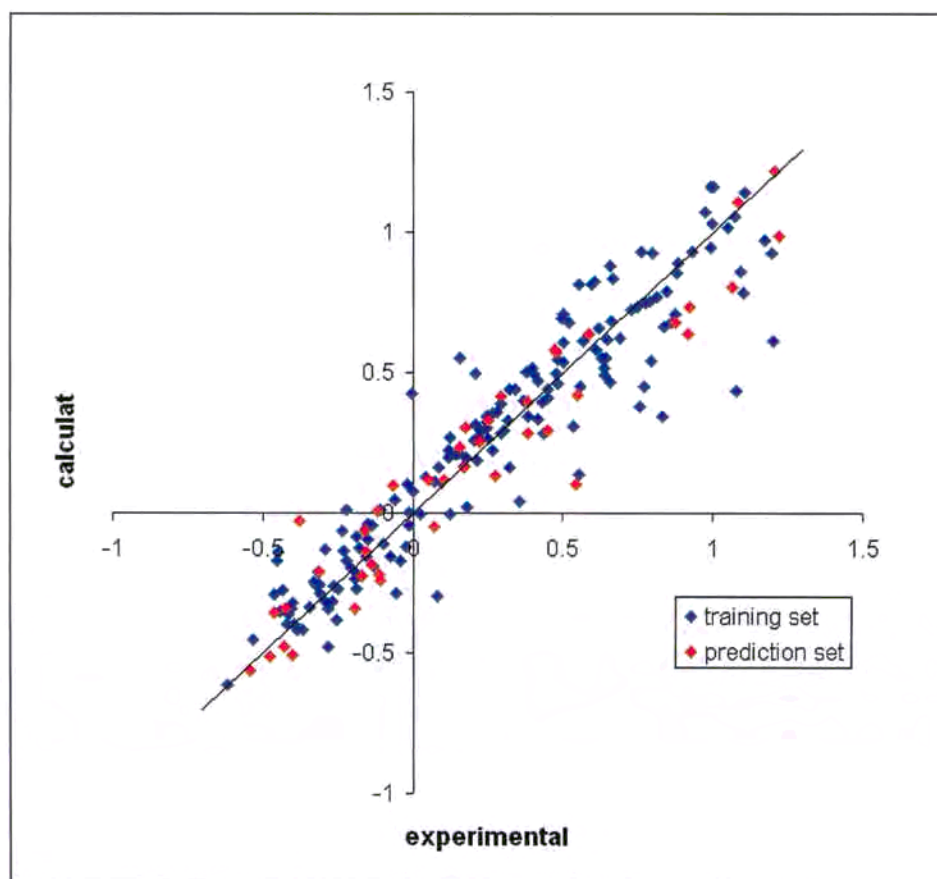


Figura 3.1. Representació gràfica de $\log \eta$ calculat vs. experimental amb el model lineal.

El model conté dos descriptors constitucionals: el nombre d'àtoms d'oxigen i el nombre relatiu d'anells de benzè; i cinc descriptors electrostàtics de tipus CPSA: RPCG, FPSA-2, DPSA-3, WNSA-1 i HACA-2/TMSA. Tots aquests últims descriptors, excepte FPSA-2, han estat calculats emprant les càrregues deduïdes pel mètode de Zefirov,⁴¹ que es basa en les electronegativitats de Sanderson.⁴²

El descriptor càrrega parcial positiva RPCG es calcula de la manera següent:

$$RPCG = \frac{\delta_{\max}}{\sum_A \delta_A}$$

on δ_{\max} és la càrrega atòmica positiva parcial màxima en la molècula, i δ_A és la càrrega atòmica positiva parcial de cada àtom.

El descriptor àrea superficial carregada positivament $FP\text{SA}-2$, es s'obté:

$$FP\text{SA}-2 = \frac{\sum_A q_A \sum_A S_A}{T\text{MSA}}$$

on q_A són les càrregues atòmiques parcials positives, S_A són les superfícies atòmiques accessibles al solvent carregades positivament i $T\text{MSA}$ és la superfície total de la molècula. En aquest cas, les càrregues parcials són calculades directament amb el mètode semiempíric PM3. El descriptor $D\text{PSA}-3$ és la diferència entre les àrees carregades positiva i negativament. Aquestes àrees carregades es calculen:

$$P\text{PSA}-3 = \sum_P q_P \cdot S_P \qquad P\text{NSA}-3 = \sum_N q_N \cdot S_N$$

on q són les càrregues atòmiques parcials, S és la superfície total accessible al solvent carregada, P i N indiquen càrregues parcials positives i negatives, respectivament. El descriptor àrea superficial negativa ponderada $W\text{NSA}-1$ es troba:

$$W\text{NSA}-1 = \frac{\sum_N S_N}{1000} \cdot T\text{MSA}$$

on S_N és la superfície total accessible al solvent carregada negativament i $T\text{MSA}$ és la superfície total de la molècula. El descriptor $H\text{ACA}-2/T\text{MSA}$ mesura la capacitat de les molècules per formar enllaços d'hidrogen per acceptació. Aquest descriptor s'obté a partir de l'àrea acceptora d'hidrogen de la molècula $H\text{ACA}-2$ i la seva superfície total $T\text{MSA}$:

$$H\text{ACA}-2 = \sum_A \frac{q_A \sqrt{S_A}}{\sqrt{S_{TOT}}}$$

on q_A és la càrrega parcial sobre l'àtom acceptor de l'enllaç d'hidrogen, S_A és la superfície d'aquest àtom i S_{TOT} és la superfície molecular accessible al solvent.

Com es pot comprovar, pràcticament tots els descriptors continguts en el model es poden relacionar directament amb les possibles interaccions intermoleculares que es donen en els compostos en estat líquid. Els descriptors CPSA són clarament indicadors d'aquestes interaccions a través de fenòmens electrostàtics: RPCG, FPSA-2, DPSA-3 i WNSA-1, o d'enllaç d'hidrogen: HACA-2/TMSA. Per la seva part, els descriptors constitucionals també poden ser responsables d'interaccions intermoleculares, així, el nombre d'àtoms d'oxigen pot incidir en interaccions de caràcter electrostàtic mentre que el nombre relatiu d'àtoms de benzè recull informació sobre les possibles interaccions que es poden donar per *pi-stacking* entre aquests anells aromàtics.

Els resultats obtinguts amb aquest model lineal són força bons, tal i com era d'esperar els descriptors inclosos contenen informació sobre les interaccions intermoleculares que es donen en els líquids que en determinen, a la seva vegada, el valor de viscositat. Com que aquest estudi és el primer que es fa sobre amb compostos organometàl·lics no es disposa d'altres resultats per poder establir comparacions directes. Els únics resultats que s'han trobat són, els ja mencionats, amb compostos orgànics. Seria d'esperar que els resultats obtinguts fossin inferiors que els publicats amb compostos orgànics ja que els valors experimentals d'aquests són més acurats i també, per la major homogeneïtat química del conjunt. En la pràctica, els resultats aquí obtinguts són molt semblants als dels conjunts de compostos orgànics. En la posterior secció de comparació de resultats es discuteix aquest fet amb més detall.

Model no lineal: Xarxes neuronals BPNN

Per tal d'establir els models no lineals amb BPNN cal utilitzar el programa ADAPT. Les rutines de selecció objectiva de descriptors implementades en aquest programa redueixen el nombre de descriptors inicials dels 200 proporcionats pel mètode heurístic a 105. La xarxa neuronal que s'utilitzarà està formada per tres capes i la seva arquitectura optimitzada és 7-5-1. Segons la fórmula enunciativa anteriorment, a una xarxa neuronal d'aquesta arquitectura li corresponen 46 paràmetres ajustables. El *ratio* d'aquest nombre sobre els compostos del *tset*, 147, és de 3.2, per sobre del 2.0 necessari.

El model no lineal desenvolupat amb ADAPT conté els descriptors que es mostren en la Taula 3.4.

Taula 3.4. Descriptors continguts en el model BPNN.

Descriptor	Classe
Índex de Kier & Hall d'ordre 1	Topològic
Contingut d'informació estructural d'ordre 1	Topològic
Contingut mitjà d'informació d'enllaç d'ordre 1	Topològic
Càrrega parcial màxima d'un àtom d'hidrogen	Electrostàtic
Àrea superficial negativa ponderada, WNSA-1	CPSA
Àrea superficial negativa fraccionada, FNSA-2	CPSA
Càrrega neta atòmica mínima	Quàntic

Com es pot comprovar, el model conté tres descriptors topològics, tres electrostàtics, dos d'ells de tipus CPSA i un descriptor quàntic. Els descriptors topològics donen informació sobre la connectivitat atòmica dins de la molècula i serveixen com a indicadors de la forma i la ramificació d'aquesta. El descriptor: índex de Kier & Hall d'ordre 1, que conté informació sobre els tipus d'àtoms i enllaços en la molècula, es calcula de la següent manera:⁴³

$${}^1\chi = \sum_i^n \sum_{j>1} \left(\frac{1}{\delta_i \delta_j} \right)^{\frac{1}{2}}$$

on n és el nombre total d'àtoms de la molècula (sense comptar hidrògens) i δ_i i δ_j són la connectivitat atòmica d'ordre zero de cada àtom. Els descriptors contingut d'informació estructural d'ordre 1 i contingut mitjà d'informació d'enllaç d'ordre 1, derivats de la teoria d'informació de Shannon,^{44,45} inclouen informació sobre la diversitat atòmica en la molècula i es relacionen amb la seva forma i simetria. El descriptor càrrega parcial màxima d'un àtom d'hidrogen és d'origen electrostàtic i es calcula amb el mètode proposat per Zefirov.⁴¹ Dels dos descriptors de tipus CPSA, un es repeteix respecte el model lineal: WNSA-1, i l'altre descriptor d'aquest tipus hi és molt semblant: àrea superficial negativa fraccionada, FNSA-2. Aquest descriptor és anàleg al FPSA-2, ja comentat en el model lineal, però en aquest cas les càrregues parcials corresponen a les àrees carregades negativament. Les càrregues parcials utilitzades en el càlcul d'aquests dos descriptors es deriven també del model de Zefirov.⁴¹ L'últim descriptor del model, la càrrega neta atòmica mínima s'obté directament del càlcul semiempíric PM3 i correspon al valor més petit de càrrega neta sobre els àtoms que formen la molècula. Els descriptors trobats en aquest cas s'assemblen als que s'han trobat en el model lineal,

potser la diferència més clara es troba en l'aparició dels tres descriptors topològics. Tot i aquesta diferència, en aquest cas els descriptors inclosos en el model també permeten la interpretació fisicoquímica de la propietat estudiada; els descriptors topològics contenen informació sobre les interaccions intermoleculares que es donen degut a la ramificació i hibridació dels àtoms en els substituents orgànics; la resta de descriptors inclouen la informació sobre les interaccions de caràcter electrostàtic i d'enllaç d'hidrogen que s'estableixen entre les molècules del líquid.

Els resultats que s'obtenen amb aquest mètode per als tres subconjunts analitzats *tset*, *cvset* i *pset* es donen en la Taula 3.5. Els valors de viscositat calculats amb aquest model es troben en la Taula 3.1 i la seva representació gràfica enfront dels valors experimentals es pot veure en la Figura 3.2.

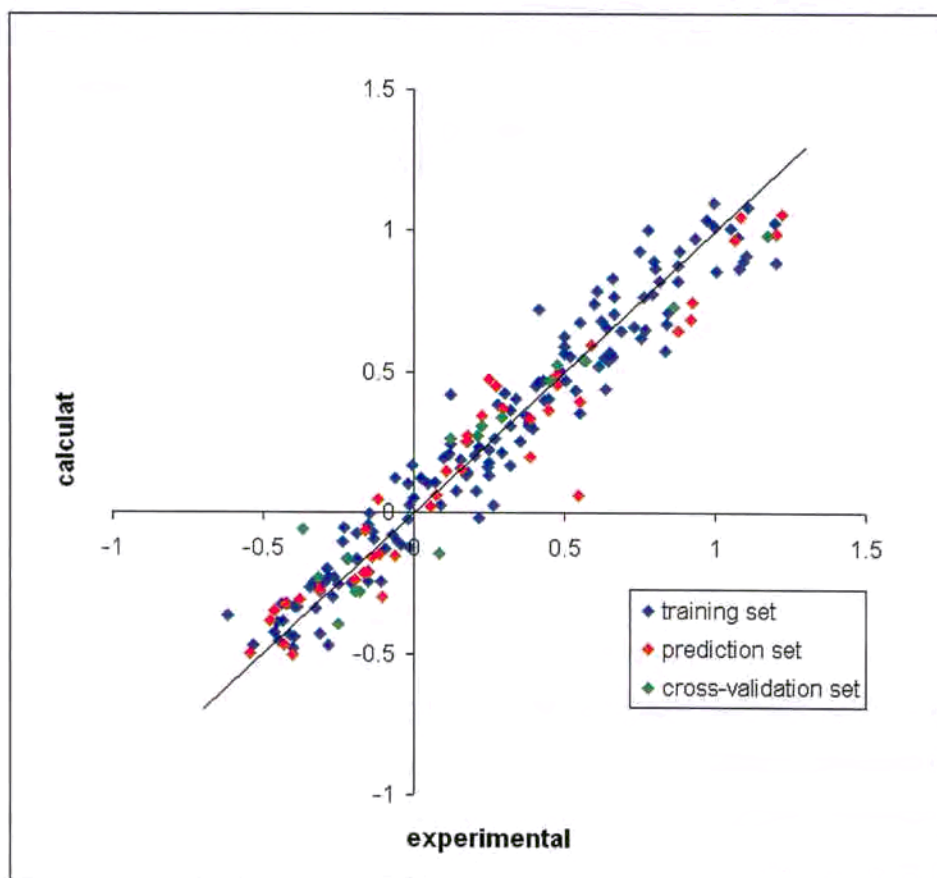


Figura 3.2. Representació gràfica de $\log \eta$ calculat vs. experimental amb el model BPNN.

Taula 3.5. Paràmetres estadístics obtinguts amb el model BPNN.

Conjunt	n	R ²	F	rmse
<i>tset</i>	147	0.943	1647.8	0.105
<i>cvset</i>	16	0.904	132.2	0.132
<i>pset</i>	42	0.929	523.1	0.120

El valor de *rmse* que s'obté amb el conjunt complet estudiat és 0.111. Només hi ha un compost, el Si(2-OMe-Et)₃Cl, que presenta un valor de residual més gran que tres vegades l'error de la correlació i, per tant, es pot considerar *outlier*. Hi ha dotze compostos que tenen un valor de residual per sobre de dues vegades el *rmse* del model (Taula 3.6) tres de fòsfor, vuit de silici i un d'estany. Dins d'aquest conjunt no sembla haver-hi cap relació directa entre l'error i l'estructura ja que els substituents orgànics són variats i els residuals són positius i negatius independentment del metall. Es pot comprovar que aquests resultats són millors als que s'obtenen amb el model lineal i es compararan als que s'obtenen per conjunts de compostos orgànics i als que s'obtenen amb el mètode SVM una mica més endavant.

Taula 3.6. Prediccions amb valors més grans que dues vegades el *rmse*^{*}.

Compost	T (°C)	Exp.	BPNN	residual
SnMe ₄	20	-0.367	-0.061	-0.306
SiPh(CH ₂ -CH=CH ₂) ₃	20	0.417	0.722	-0.305
PO(O'Pr) ₂ F	25	0.122	0.420	-0.298
SiMe ₄	20	-0.616	-0.366	-0.250
POPh(OCH ₂ -CH=CH ₂) ₂	25	0.777	1.001	-0.224
Si(CH=CH ₂)(OMe) ₃	25	0.083	-0.143	0.226
SiMePh ₂ (CH ₂ -CH=CH ₂)	20	0.877	0.644	0.233
SiMePh(OMe) ₂	20	0.216	-0.019	0.235
Si(O(2-OEt-Et)) ₄	20	0.920	0.683	0.237
Si(3-Cl-Pr)Cl ₃	20	0.266	0.025	0.241
Si(2-Et-Hex) ₃ F	25	0.833	0.572	0.261
POMe(Cl)(OBU) ₂	30	1.204	0.888	0.316
Si(2-OMe-Et) ₃ Cl	20	0.547	0.060	0.487

* Els compostos ombrejats tenen residuals més grans que tres vegades el *rmse*

Model no lineal: Support Vector Machines

El mètode triat per construir els models no lineals amb SVM ha estat l'heurístic-SVM, en el qual es fa una selecció no lineal de descriptors. Els descriptors de partida emprats en aquest procediment són els 200 provinents de l'anàlisi preliminar del mètode heurístic. El model trobat conté els set descriptors que es troben en la Taula 3.7.

Taula 3.7. Descriptors continguts en el model SVM.

Descriptor	Classe
Nombre relatiu d'àtoms d'oxigen	Constitucional
Nombre d'enllaços dobles	Constitucional
Contingut mitjà d'informació complementària d'ordre 1	Topològic
Càrrega parcial màxima d'un àtom d'hidrogen	Electroestàtic
Superfície donadora d'enllaç d'hidrogen, HDSA-1	CPSA
Diferència entre superfícies carregades, DPSA-2	CPSA
Àrea superficial parcial negativa, PNSA-1	CPSA

Es pot comprovar que aquest model sembla estar a mig camí entre els dos models trobats anteriorment; conté dos descriptors constitucionals, un descriptor topològic i quatre descriptors electroestàtics, dos d'ells de tipus CPSA. El descriptor topològic, contingut mitjà d'informació complementària d'ordre 1, pertany, a l'igual que els vistos anteriorment, a la teoria d'informació de Shannon.^{44,45} Aquest descriptor conté informació sobre la diversitat atòmica, la forma i la simetria de la molècula. El descriptor HDSA-1 mesura la capacitat de les molècules per establir enllaços d'hidrogen per donació i es calcula:

$$HDSA-1 = \sum_D S_D$$

on S_D és l'àrea superficial dels àtoms d'hidrogen donables. El descriptor DPSA-2 és la diferència entre les àrees carregades positiva i negativament. Aquestes àrees carregades, que s'obtenen de la distribució de càrregues calculada pel mètode de Zefirov,⁴¹ s'obtenen:

$$PPSA-2 = \sum_P q_P \cdot \sum_P S_P \qquad PNSA-2 = \sum_N q_N \cdot \sum_N S_N$$

on q són les càrregues atòmiques parcials, S és la superfície total accessible al solvent carregada, P i N indiquen càrregues parcials positives i negatives, respectivament. El descriptor PNSA-1, que és molt semblant al ja vist, WNSA-1, es calcula:

$$PNSA-1 = \sum_N S_N$$

on S_N és la superfície total accessible al solvent carregada negativament. En aquest cas, la distribució de càrrega també ha estat calculada amb el mètode de Zefirov.⁴¹ Així doncs, es pot dir que els descriptors d'aquest model es poden relacionar directament amb la viscositat ja que contenen informació sobre els diferents tipus d'interaccions intermoleculares que es poden donar en els líquids. Els descriptors constitucionals i els electrostàtics es poden referir a les interaccions de tipus electrostàtic i d'enllaços d'hidrogen mentre que el descriptor topològic conté la informació relacionada amb les interaccions que es donen fruit de la hibridació i ramificació dels substituents orgànics.

L'obtenció del model definitiu, que permet el càlcul de la viscositat, requereix l'optimització dels paràmetres característics γ , ε i C amb el mètode de minimització de l'error de validació creuada, Figura 3.3. L'optimització del paràmetre γ (amb $\varepsilon=0.1$ i $C=100$) mostra que el millor valor trobat per aquest paràmetre és 0.6 ja que en aquest punt el valor de *rmse* és mínim. Un cop optimitzat aquest paràmetre, es passa a optimitzar ε (amb $\gamma=0.6$ i $C=100$), l'optimització d'aquest paràmetre porta a un valor de 0.09. El paràmetre C s'optimitza amb $\gamma=0.6$ i $\varepsilon=0.09$; el valor òptim per aquest paràmetre és 135.

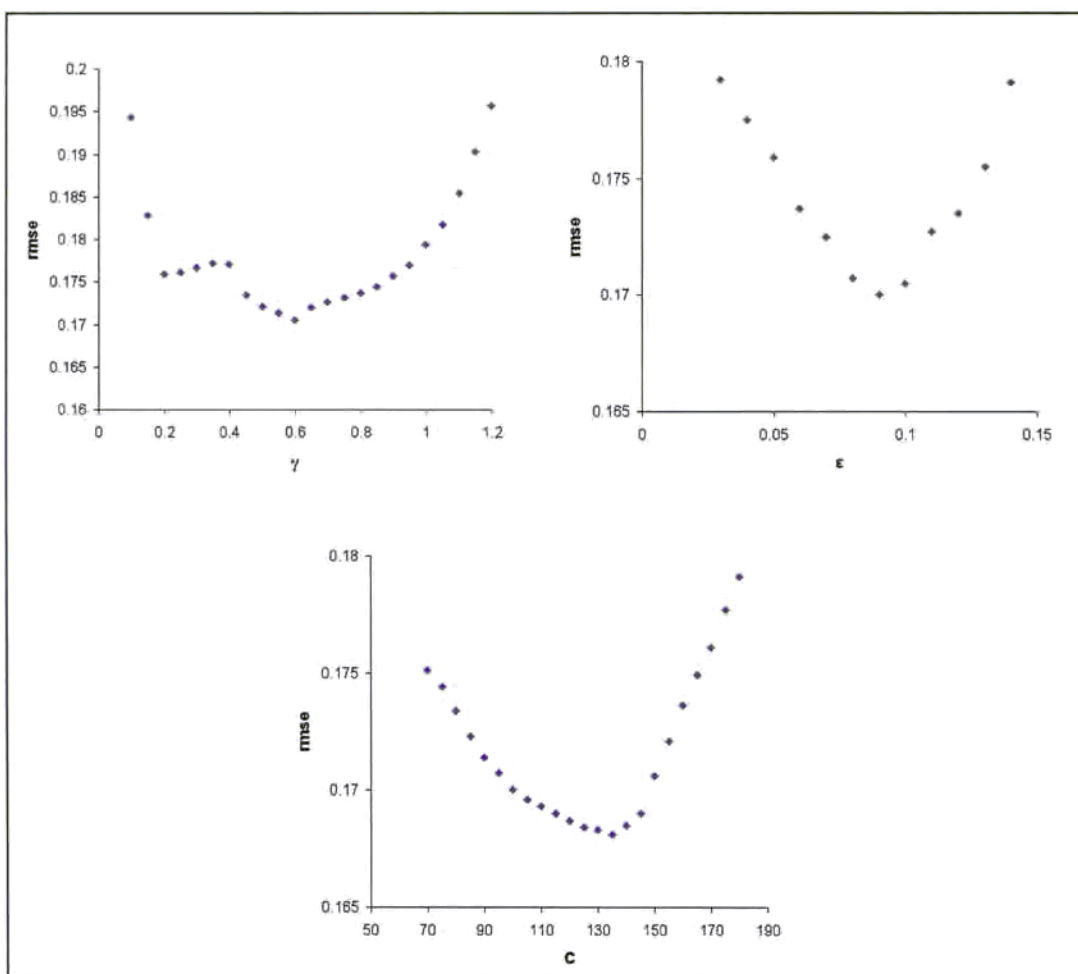


Figura 3.3. Optimització dels paràmetres γ , ϵ i C .

El model construït amb aquest procediment consta de 80 vectors suport, és a dir, s'utilitza aquest nombre de compostos d'un total de 163 del *tset* per establir el model. Els resultats estadístics que s'obtenen amb aquest model són molt bons i semblants als que s'obtenen amb xarxes BPNN, la Taula 3.8 conté els valors del coeficient de determinació R^2 , el test- F i *rmse* per als dos subconjunts de dades: *tset* i *pset*.

Taula 3.8. Paràmetres estadístics obtinguts amb el model SVM.

Conjunt	n	R^2	F	<i>rmse</i>
<i>tset</i>	163	0.940	2532.3	0.107
<i>pset</i>	42	0.942	655.3	0.103

Els valors de viscositat calculats amb aquest model es troben en la Taula 3.1 i la seva representació gràfica enfront dels valors experimentals es pot veure en la Figura 3.4.

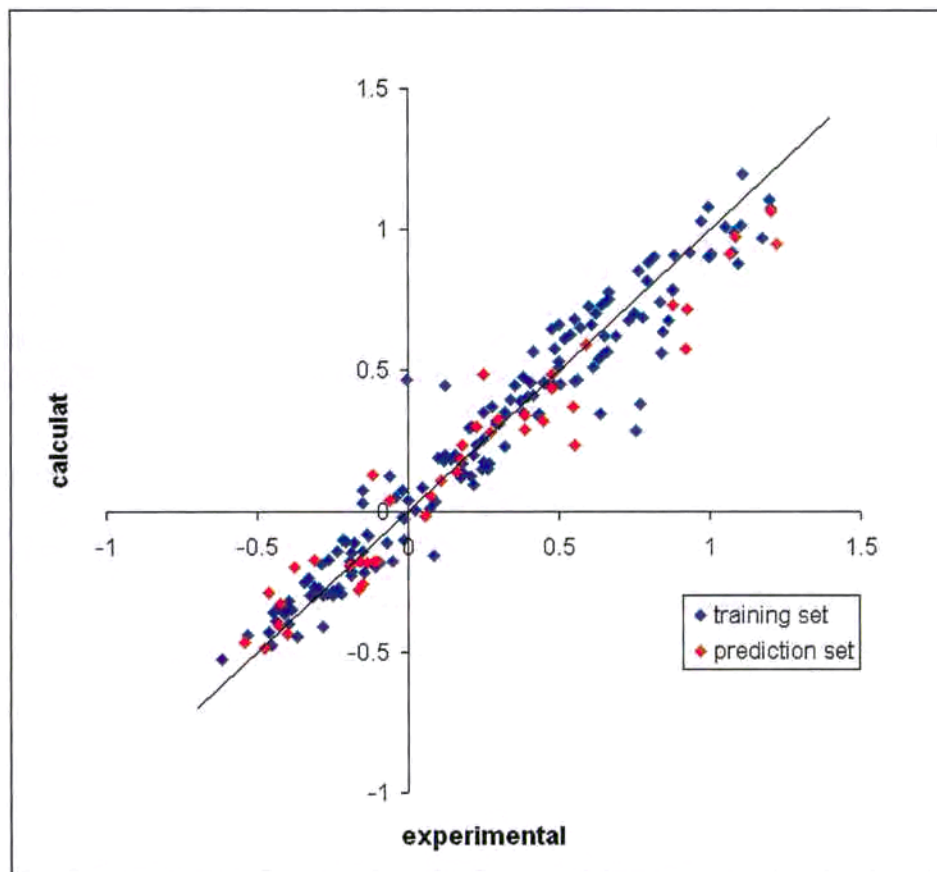


Figura 3.4. Representació gràfica de $\log \eta$ calculat vs. experimental amb el model SVM.

Comparació de resultats

Com ja s'ha comentat anteriorment, aquest és el primer estudi que es fa mitjançant la metodologia QSPR per predir la viscositat d'un conjunt de compostos organometàl·lics. Els únics exemples descrits en la bibliografia de treballs semblants es troben restringits a conjunts de compostos orgànics. Degut a la major facilitat per determinar la propietat estudiada per a aquest últim tipus de compostos, seria d'esperar que els models desenvolupats per ells proporcionessin uns millors resultats que els aquí obtinguts. En la Taula 3.9 es presenten els resultats aconseguits en diferents estudis on s'ha emprat la metodologia QSPR per estimar la viscositat de diferents conjunts de compostos orgànics.

Taula 3.9. Estadística dels diferents estudis pel càlcul de la viscositat.

Referència	Compostos	Model ^a	Descriptors	R ²	rmse
28	116	MLR	5	0.870	0.144
		PLS	5	0.867	0.146
29	361	MLR	9	0.917	0.167
		BPNN	9	0.947	0.132
32	337	MLR	5	0.846	0.160
33	361	MLR	5	0.854	0.220
34	361	MLR	5	0.873	0.206
		CROMRSel	5	0.908	0.175
		MLR	4	0.899	0.174
35	46	PLS	4	0.950	0.148
		MLR	8	0.824	0.258
36	213	BPNN	8	0.944	0.152
		MLR	46	0.959	0.111
37	367	BPNN	46	0.977	0.084

^a: MLR= regressió multilíneal, PLS= regressió per mínims quadrats parcial, BPNN= *Back Propagation Neural Network*, CROMRSel= model no lineal derivat de regressions.

Com es pot comprovar, els valors que apareixen en la Taula 3.9 són del mateix ordre que els resultats que s'obtenen amb els diferents models desenvolupats en aquest treball. Tenint en compte que els resultats de la taula han estat obtinguts per conjunts de compostos orgànics, es pot dir que els models construïts pel càlcul de la viscositat del conjunt de 205 compostos organometàl·lics funcionen molt bé.

Els descriptors que conformen els models d'aquests treballs són força semblants als que s'obtenen aquí. A títol d'exemple, en un dels treballs desenvolupats per Jurs,³⁶ els descriptors triats per calcular la viscositat són del mateix tipus que els que s'han trobat en els models lineals i no lineals construïts: FNSA-2, FNSA-3, WPSA-3 i alguns descriptors relacionats amb la formació d'enllaços d'hidrogen com HACA o HDSA.

Els tres models aquí desenvolupats presenten descriptors del mateix tipus, fins i tot alguns com DPSA-2 i WNSA-1 es repeteixen. Tot i aquestes semblances, es pot observar que els resultats que s'obtenen amb el mètode lineal i els no lineals són diferents, sent molt superiors els del segon tipus. En termes estadístics, els models no lineals proporcionen errors aproximadament un 50% més baixos i valors de coeficient de determinació força més elevats, amb una dispersió en els errors per als valors predits molt menor; fet que estaria d'acord amb la complexitat de les interaccions intermoleculares, en general, i de les responsables de la viscositat, en particular.

Taula 3.1. Valors experimentals i calculats de $\log \eta$ (mPa·s).

#	Compost	T (°C)	Exp.	MLR	BPNN	SVM
1	As(OBu) ₃ ^{a,b}	22	0.248	0.341	0.179	0.257
2	As(O'Bu) ₃ ^{a,b}	22	0.251	0.261	0.130	0.175
3	As(O'Pent) ₃ ^{a,b}	22	0.382	0.499	0.309	0.478
4	As(O'Pr) ₃ ^d	29	0.056	0.117	0.021	-0.016
5	As(OMe) ₃ ^{a,c}	20	-0.178	-0.125	-0.280	-0.114
6	As(OPent) ₃ ^{a,b}	22	0.397	0.514	0.300	0.404
7	As(OPr) ₃ ^{a,b}	28	0.089	0.161	0.027	0.034
8	AsEt ₃ ^d	20	-0.378	-0.029	-0.309	-0.200
9	AsMe ₃ ^{a,b}	20	-0.447	-0.139	-0.455	-0.361
10	Ge(OBu) ₄ ^{a,b}	25	0.340	0.442	0.405	0.394
11	Ge(OEt) ₄ ^{a,b}	20	0.024	-0.003	0.124	0.004
12	Ge(O'Bu) ₄ ^d	20	0.449	0.294	0.361	0.321
13	Ge(O'iPr) ₄ ^{a,b}	20	0.119	0.196	0.207	0.179
14	Ge(OMe) ₄ ^d	20	-0.112	-0.222	-0.152	-0.180
15	Ge(OPr) ₄ ^{a,b}	20	0.179	0.199	0.252	0.172
16	Ge(O ^s Bu) ₄ ^{a,b}	20	0.449	0.408	0.403	0.455
17	Ge(O'Bu) ₄ ^{a,b}	25	0.757	0.379	0.621	0.285
18	GeBu ₄ ^{a,b}	22	0.482	0.459	0.495	0.449
19	GeEt ₄ ^{a,b}	22	-0.109	0.012	-0.247	-0.199
20	GeHex ₄ ^{a,b}	23	0.802	0.924	0.866	0.893
21	GeMe ₄ ^{a,b}	20	-0.454	-0.170	-0.393	-0.473
22	GePent ₄ ^{a,c}	22	0.864	0.677	0.728	0.680
23	GePr ₄ ^d	22	0.160	0.234	0.150	0.138
24	P(NMe ₂) ₃ ^{a,b}	30	0.435	0.284	0.400	0.345
25	P(OMe) ₃ ^{a,b}	25	-0.262	-0.264	-0.230	-0.172
26	P(OPh) ₃ ^d	25	1.205	1.218	0.987	1.065
27	PO(COPh)(OEt) ₂ ^{a,b}	25	0.998	1.031	1.010	0.908
28	PO(NMe ₂) ₃ ^{a,b}	22	0.536	0.309	0.434	0.626
29	PO(O(2-Et-Bu)) ₃ ^{a,b}	20	0.934	0.930	0.969	0.919
30	PO(O(2-Me-Pr)) ₃ ^{a,b}	23	0.645	0.553	0.537	0.560
31	PO(O(2-OBu-Et)) ₃ ^d	25	1.086	1.106	1.050	0.974
32	PO(OBu)(OEt) ₂ ^{a,c}	25	0.294	0.391	0.339	0.322
33	PO(OBu)(OPh) ₂ ^{a,b}	25	1.107	1.144	1.083	1.197
34	PO(OBu) ₂ (O(4-(OEt)-Bu)) ^{a,b}	25	0.661	0.881	0.829	0.751
35	PO(OBu) ₂ (OHex) ^{a,b}	25	0.598	0.811	0.739	0.726
36	PO(OBu) ₃ ^d	20	0.589	0.638	0.596	0.592
37	PO(OEt)(NMe ₂) ₂ ^{a,b}	25	0.293	0.283	0.210	0.312
38	PO(OEt) ₂ (O(4-NO ₂ -Ph)) ^{a,b}	25	1.198	0.925	1.026	1.108
39	PO(OEt) ₂ (OPent) ^{a,b}	25	0.405	0.495	0.448	0.459
40	PO(OEt) ₂ NMe ₂ ^{a,b}	25	0.218	0.267	0.234	0.200
41	PO(OEt) ₃ ^d	20	0.225	0.254	0.342	0.302
42	PO(OHex) ₃ ^{a,b}	20	1.076	1.058	0.977	0.921
43	PO(O'Bu)(OEt) ₂ ^{a,b}	25	0.279	0.360	0.382	0.369
44	PO(O'Pent) ₃	28	0.607	0.825	0.786	0.666
45	PO(O'Pr) ₂ F ^{a,b}	25	0.122	0.225	0.420	0.446
46	PO(O'Pr) ₃ ^d	23	0.385	0.395	0.335	0.291
47	PO(OMe)(OPh) ₂ ^{a,c}	25	1.176	0.971	0.980	0.973
48	PO(OMe) ₃ ^{a,b}	20	0.354	0.038	0.254	0.444
49	PO(OPr) ₃ ^{a,c}	20	0.452	0.439	0.471	0.454
50	POBu(OBu) ₂ ^{a,b}	25	0.501	0.708	0.623	0.662

#	Compost	T (°C)	Exp.	MLR	BPNN	SVM
51	POBu(OEt) ₂ ^d	25	0.297	0.416	0.370	0.325
52	POBu ₂ (OBu) ^{a,b}	25	0.732	0.720	0.660	0.677
53	POBu ₂ (OEt) ^{a,b}	25	0.611	0.581	0.520	0.513
54	POBu ₂ (OMe) ^{a,b}	25	0.638	0.488	0.439	0.345
55	PO(CCl ₃)(OEt) ₂ ^{a,b}	25	0.796	0.542	0.892	0.886
56	POCl(NMe ₂) ₂ ^d	25	0.550	0.418	0.393	0.233
57	POCl ₂ NMe ₂ ^{a,b}	25	0.303	0.294	0.421	0.309
58	POEt(OBu) ₂ ^{a,c}	25	0.504	0.538	0.469	0.450
59	POEt(OEt) ₂ ^{a,b}	25	0.124	0.268	0.264	0.184
60	POHex(OBu) ₂ ^{a,b}	25	0.666	0.834	0.763	0.777
61	POHex(OEt) ₂ ^d	25	0.476	0.583	0.488	0.485
62	POHex ₂ (OEt) ^{a,b}	25	0.879	0.852	0.821	0.789
63	PO(MeCl)(OBu) ₂ ^{a,b}	30	1.204	0.614	0.888	1.074
64	PO(MeCl)(OPr) ₂ ^{a,b}	30	1.079	0.434	0.865	0.989
65	POMe(O(3-Me-Bu)) ₂ ^{a,b}	20	0.651	0.617	0.567	0.623
66	POMe(OEt) ₂ ^d	25	0.173	0.162	0.252	0.183
67	POMe(O'Pr) ₂ ^{a,b}	25	0.267	0.352	0.261	0.168
68	POPh(O(CH ₂ -CH=CH ₂)) ₂ ^{a,b}	25	0.777	0.747	1.001	0.687
69	POPh(OBu) ₂ ^{a,b}	25	0.884	0.890	0.928	0.908
70	POPh(OEt) ₂ ^{a,b}	25	0.770	0.451	0.652	0.382
71	POPh(O'Pr) ₂ ^d	30	0.924	0.733	0.745	0.718
72	POPh(OPent) ₂ ^{a,b}	30	1.053	1.014	1.009	1.011
73	Si(2-CN-Et)(OBu) ₃ ^{a,b}	20	0.486	0.547	0.502	0.576
74	Si(2-Et-Hex) ₃ F ^{a,b}	25	0.833	0.341	0.572	0.743
75	Si(2-OMe-Et) ₃ Cl ^d	20	0.547	0.102	0.060	0.370
76	Si(3-Cl-Pr)(OMe) ₃ ^{a,b}	20	-0.222	0.011	-0.167	-0.293
77	Si(3-Cl-Pr)Cl ₃ ^{a,b}	20	0.266	0.221	0.025	0.150
78	Si(CH=CH ₂)(OMe) ₃ ^{a,c}	25	0.083	-0.296	-0.143	-0.157
79	Si(CH=CH ₂) ₄ ^{a,b}	20	-0.330	-0.248	-0.247	-0.240
80	Si(MeCl)(CH ₂ -CH=CH ₂) ₃ ^d	20	0.274	0.134	0.447	0.282
81	Si(O(1-Me-Bu)) ₄ ^{a,b}	30	0.502	0.604	0.565	0.511
82	Si(O(1-Me-Pr)) ₄ ^{a,b}	20	0.371	0.397	0.343	0.392
83	Si(O(2-Cl-Et)) ₄ ^{a,b}	20	0.993	0.943	1.018	1.083
84	Si(O(2-Me-Pr)) ₄ ^{a,b}	20	0.383	0.345	0.318	0.351
85	Si(O(2-OEt-Et)) ₄ ^d	20	0.920	0.635	0.683	0.577
86	Si(O(2-OMe-Et)) ₄ ^{a,b}	21	0.559	0.451	0.543	0.469
87	Si(O(3-Me-Bu)) ₄ ^{a,b}	20	0.520	0.679	0.556	0.611
88	Si(OBu) ₄ ^{a,b}	20	0.322	0.441	0.365	0.349
89	Si(OEt) ₂ (O(2-Me-Ph)) ₂ ^{a,b}	25	0.663	0.683	0.706	0.753
90	Si(OEt) ₄ ^d	20	-0.159	-0.066	-0.065	-0.180
91	Si(O'Pr) ₂ (O(2-Me-Ph)) ₂ ^{a,b}	25	0.792	0.755	0.774	0.818
92	Si(O'Pr) ₃ (O(2Me-Ph)) ^{a,c}	25	0.477	0.495	0.526	0.647
93	Si(O'Pr) ₄ ^{a,b}	20	0.173	0.167	0.133	0.121
94	Si(OMe) ₂ (O(2-Me-Ph)) ₂ ^{a,b}	25	0.623	0.659	0.680	0.704
95	Si(OMe) ₄ ^d	20	-0.192	-0.344	-0.242	-0.196
96	Si(OPent) ₄ ^{a,b}	20	0.500	0.693	0.589	0.530
97	Si(OPr) ₄ ^{a,b}	25	0.155	0.214	0.185	0.197
98	Si(Pent) ₃ F ^{a,b}	25	0.554	0.137	0.353	0.464
99	Si(CH ₂ -CH=CH ₂)(CH=CH ₂) ₃ ^{a,b}	20	-0.230	-0.136	-0.056	-0.275
100	Si(CH ₂ -CH=CH ₂)(OEt) ₃ ^d	20	-0.117	0.006	0.045	0.127
101	Si(CH ₂ -CH=CH ₂) ₂ (CH=CH ₂) ₂ ^{a,b}	20	-0.147	-0.042	-0.004	-0.218

#	Compost	T (°C)	Exp.	MLR	BPNN	SVM
102	Si(CH ₂ -CH=CH ₂) ₂ (OEt) ₂ ^{a,b}	20	-0.059	0.043	0.120	0.126
103	Si(CH ₂ -CH=CH ₂) ₃ (4Cl-Ph) ^{a,b}	20	0.749	0.733	0.929	0.705
104	Si(CH ₂ -CH=CH ₂) ₄ ^{a,b}	20	0.072	0.110	0.108	0.010
105	SiBu ₃ F ^d	25	0.385	0.284	0.196	0.341
106	SiEt(OBu) ₃ ^{a,c}	20	0.212	0.314	0.274	0.302
107	SiEt(OEt) ₃ ^{a,b}	20	-0.187	-0.084	-0.076	-0.149
108	SiEt(OMe) ₃ ^{a,b}	20	-0.250	-0.275	-0.258	-0.301
109	SiEt(OPr) ₃ ^{a,b}	20	0.045	0.125	0.100	0.084
110	SiEt ₂ (OBu) ₂ ^d	20	0.107	0.116	0.149	0.110
111	SiEt ₂ (OPr) ₂ ^{a,b}	20	-0.013	-0.002	0.027	-0.029
112	SiEt ₂ Cl ₂ ^{a,b}	20	-0.151	-0.096	-0.212	0.027
113	SiEt ₃ Bu ^{a,b}	20	-0.020	-0.122	-0.023	-0.024
114	SiEt ₃ Hex ^{a,b}	20	0.180	0.022	0.142	0.141
115	SiEt ₃ Pent ^d	20	0.074	-0.050	0.060	0.052
116	SiEt ₃ Pr ^{a,b}	20	-0.132	-0.191	-0.093	-0.087
117	SiEt ₄ ^{a,b}	20	-0.188	-0.274	-0.163	-0.212
118	SiEtCl ₃ ^{a,c}	20	-0.218	-0.171	-0.163	-0.101
119	SiH(OEt) ₃ ^{a,b}	15	-0.289	-0.131	-0.201	-0.188
120	SiH ₂ Et ₂ ^d	25	-0.542	-0.568	-0.501	-0.465
121	SiH(CH ₂ -CH=CH ₂) ₃ ^{a,b}	20	-0.095	-0.111	-0.129	-0.184
122	SiHEt ₂ Cl ^{a,b}	20	-0.310	-0.289	-0.430	-0.268
123	SiHEt ₃ ^{a,b}	20	-0.385	-0.419	-0.337	-0.351
124	SiHEtCl ₂ ^{a,b}	20	-0.324	-0.258	-0.344	-0.297
125	SiHMe ₂ (CH ₂ -CH=CH ₂) ^d	20	-0.427	-0.480	-0.474	-0.405
126	SiHMe ₂ Cl ^{a,b}	20	-0.533	-0.456	-0.471	-0.443
127	SiHMe(CH ₂ -CH=CH ₂) ₂ ^{a,b}	20	-0.282	-0.480	-0.474	-0.410
128	SiHMeCl ₂ ^{a,b}	20	-0.463	-0.294	-0.427	-0.429
129	SiHMeEtCl ^{a,b}	20	-0.412	-0.362	-0.439	-0.363
130	SiHMePh ₂ ^d	25	0.477	0.572	0.451	0.438
131	SiHPh(CH ₂ -CH=CH ₂) ₂ ^{a,c}	20	0.228	0.286	0.308	0.237
132	SiMe(2-CF ₃ -Et)Cl ₂ ^{a,b}	20	-0.004	0.424	0.167	0.465
133	SiMe(CH=CH ₂)(OMe) ₂ ^{a,b}	20	-0.345	-0.337	-0.267	-0.255
134	SiMe(CH=CH ₂) ₃ ^{a,b}	20	-0.398	-0.325	-0.338	-0.401
135	SiMe(CH=CH ₂)(CH ₂ -CH=CH ₂) ₂ ^d	20	-0.152	-0.142	-0.217	-0.261
136	SiMe(MeCl)(CH ₂ -CH=CH ₂) ₂ ^{a,b}	20	0.124	-0.003	0.244	0.199
137	SiMe(OEt) ₃ ^{a,b}	20	-0.236	-0.066	-0.105	-0.146
138	SiMe(OMe) ₃ ^{a,b}	20	-0.268	-0.318	-0.300	-0.292
139	SiMe ₂ (CH=CH ₂)Cl ^{a,b}	20	-0.434	-0.276	-0.385	-0.344
140	SiMe ₂ (CH=CH ₂)(OMe) ^d	20	-0.463	-0.359	-0.353	-0.289
141	SiMe ₂ (MeCl)(CH ₂ -CH=CH ₂) ^{a,b}	20	-0.043	-0.173	-0.113	0.047
142	SiMe ₂ (O(1-Me-Heptyl)) ₂ ^{a,b}	20	0.630	0.549	0.548	0.540
143	SiMe ₂ (OBu) ₂ ^{a,b}	20	0.000	0.073	0.052	0.037
144	SiMe ₂ (OEt) ₂ ^{a,c}	20	-0.317	-0.210	-0.232	-0.304
145	SiMe ₂ (OMe) ₂ ^d	20	-0.423	-0.343	-0.325	-0.328
146	SiMe ₂ (OPr) ₂ ^{a,b}	20	-0.137	-0.048	-0.068	-0.084
147	SiMe ₂ (CH ₂ -CH=CH ₂) ₂ ^{a,b}	20	-0.208	-0.198	-0.248	-0.110
148	SiMe ₂ (CH ₂ -CH=CH ₂)Cl ^{a,b}	20	-0.152	-0.156	-0.247	-0.141
149	SiMe ₂ Br ₂ ^{a,b}	25	-0.054	-0.287	-0.103	-0.178
150	SiMe ₂ Cl ₂ ^d	20	-0.310	-0.210	-0.275	-0.171
151	SiMe ₂ Ph ₂ ^{a,b}	25	0.553	0.815	0.675	0.681
152	SiMe ₂ Ph(CH ₂ -CH=CH ₂) ^{a,b}	20	0.143	0.207	0.075	0.185

#	Compost	T (°C)	Exp.	MLR	BPNN	SVM
153	SiMe ₂ PhBr ^{a,b}	25	0.246	0.325	0.160	0.152
154	SiMe ₂ Ph(OBu) ^{a,b}	25	0.248	0.301	0.222	0.352
155	SiMe ₃ (CH=CH ₂) ^d	20	-0.399	-0.508	-0.508	-0.437
156	SiMe ₃ (CH ₂ -CH=CH ₂) ^{a,b}	20	-0.394	-0.401	-0.442	-0.320
157	SiMe ₃ Br ^{a,c}	25	-0.252	-0.383	-0.398	-0.284
158	SiMe ₃ Bu ^{a,b}	20	-0.298	-0.315	-0.251	-0.272
159	SiMe ₃ Cl ^{a,b}	20	-0.403	-0.344	-0.480	-0.322
160	SiMe ₃ Et ^d	20	-0.476	-0.514	-0.387	-0.486
161	SiMe ₃ Hex ^{a,b}	20	-0.072	-0.157	-0.081	-0.113
162	SiMe ₃ (OEt) ^{a,b}	20	-0.440	-0.353	-0.326	-0.389
163	SiMe ₃ (OH) ^{a,b}	20	0.817	0.767	0.820	0.908
164	SiMe ₃ Pent ^{a,b}	20	-0.191	-0.239	-0.170	-0.228
165	SiMe ₃ Ph ^d	25	-0.063	0.097	-0.154	0.040
166	SiMe ₃ Pr ^{a,b}	20	-0.420	-0.401	-0.326	-0.350
167	SiMe ₄ ^{a,b}	20	-0.616	-0.614	-0.366	-0.526
168	SiMe(CH ₂ -CH=CH ₂) ₃ ^{a,b}	20	-0.012	-0.047	-0.124	-0.102
169	SiMeCl ₃ ^{a,b}	20	-0.312	-0.256	-0.281	-0.289
170	SiMeEt ₂ Br ^d	25	-0.105	-0.245	-0.301	-0.177
171	SiMeEt ₃ ^{a,b}	20	-0.281	-0.342	-0.232	-0.299
172	SiMeEtCl ₂ ^{a,c}	20	-0.194	-0.214	-0.282	-0.179
173	SiMePh(OEt) ₂ ^{a,b}	25	0.204	0.256	0.203	0.294
174	SiMePh(OMe) ₂ ^{a,b}	20	0.216	0.185	-0.019	0.094
175	SiMePh ₂ (CH ₂ -CH=CH ₂) ^d	20	0.877	0.678	0.644	0.732
176	SiMePh ₂ Cl ^{a,b}	20	0.846	0.788	0.711	0.638
177	SiMePh ₂ (OEt) ^{a,b}	25	0.690	0.622	0.643	0.623
178	SiMePh(CH ₂ -CH=CH ₂) ₂ ^{a,b}	20	0.320	0.330	0.310	0.343
179	SiMePhCl ₂ ^{a,b}	20	0.209	0.493	0.077	0.122
180	SiOct(OEt) ₃ ^d	25	0.177	0.304	0.273	0.234
181	Si(OEt)(OPh) ₃ ^{a,b}	21	0.974	1.071	1.036	1.033
182	Si(OEt)(CH ₂ -CH=CH ₂) ₃ ^{a,b}	20	-0.017	0.104	0.099	0.073
183	SiPh(CH=CH ₂)(CH ₂ -CH=CH ₂) ₂ ^{a,b}	20	0.431	0.393	0.470	0.341
184	SiPh(OBu) ₃ ^{a,c}	20	0.571	0.613	0.541	0.654
185	SiPh(OEt) ₃ ^d	20	0.250	0.330	0.471	0.489
186	SiPh(OMe) ₃ ^{a,b}	20	0.322	0.160	0.168	0.232
187	SiPh ₂ (OBu) ₂ ^{a,b}	20	1.093	0.861	0.892	0.880
188	SiPh ₂ (OEt) ₂ ^{a,b}	20	0.875	0.705	0.879	0.785
189	SiPh ₂ (OMe) ₂ ^{a,b}	25	0.839	0.660	0.670	0.563
190	SiPh ₂ (OPr) ₂ ^d	20	1.064	0.804	0.965	0.916
191	SiPh ₂ (CH ₂ -CH=CH ₂) ₂ ^{a,b}	20	1.104	0.785	0.910	1.014
192	SiPh ₂ Cl ₂ ^{a,b}	20	0.765	0.930	0.767	0.855
193	SiPh(CH ₂ -CH=CH ₂) ₃ ^{a,b}	20	0.417	0.469	0.722	0.567
194	SiPhCl ₃ ^{a,b}	20	0.158	0.551	0.155	0.197
195	SiPr(OMe) ₃ ^d	20	-0.166	-0.226	-0.215	-0.281
196	SiPrCl ₃ ^{a,b}	20	-0.153	-0.056	-0.043	0.073
197	Sn(O(1,1-Me ₂)Bu) ₄ ^{a,b}	25	0.994	1.163	1.096	0.904
198	Sn(O(1,1-Me ₂)Pr) ₄ ^d	25	1.221	0.985	1.055	0.948
199	SnBrBu ₃ ^{a,b}	20	0.640	0.514	0.657	0.732
200	SnBu ₃ (NCO) ^{a,b}	20	1.005	1.160	0.857	0.915
201	SnBu ₄ ^{a,b}	20	0.417	0.335	0.465	0.410
202	SnClBu ₃ ^{a,b}	25	0.660	0.465	0.555	0.570
203	SnEt ₄ ^d	20	-0.137	-0.189	-0.160	-0.183

#	Compost	T (°C)	Exp.	MLR	BPNN	SVM
204	SnMe ₄ ^{a,c}	20	-0.367	-0.419	-0.061	-0.443
205	SnPr ₄ ^{a,b}	25	0.097	0.107	0.192	0.187

^a *Training set* de MLR i SVM. ^b *Training set* de BPNN.

^c *Cross-validation set* de BPNN. ^d *Prediction set*.

3.1.2. Viscositat a temperatura variable

El valor de viscositat dels compostos químics depèn de manera molt marcada de la temperatura a la qual es fa la determinació experimental, sent més petita quant més elevada és la temperatura. Quan es va construir la base de dades de la viscositat per al conjunt de 205 compostos es va veure que hi havia mesures de molts d'aquests compostos a diferents temperatures, això va plantejar la idea d'establir models amb aquests compostos a diferents temperatures. Aquesta metodologia s'ha trobat documentada en un treball de Zefirov per a aquesta mateixa propietat en un conjunt molt gran de compostos orgànics de naturalesa química variada.⁴⁰

Conjunt de dades

La base de dades utilitzada conté 534 valors experimentals de viscositat dels 205 compostos utilitzats anteriorment. Tots aquests valors han estat extrets també de la base de dades Landolt-Börnstein.²⁵ En aquesta ocasió, els valors de $\log \eta$ cobreixen un rang que va des de -0.762 per al SiMe_4 a 100°C fins a 1.225 per al $\text{PO}(\text{O}(2\text{-Et-Bu}))_3$ a 0°C , amb un valor mitjà de 0.255; el rang de temperatures va des de 0 a 120°C , amb una mitjà de 35°C . En aquest cas, la distribució del nombre de valors experimentals entre els diferents elements queda de la següent manera: 53 d'arsènic, 82 de germani, 153 de fòsfor, 210 de silici i 36 d'estany. Tots els valors experimentals i calculats es troben en la Taula 3.10 així com la temperatura exacta de la determinació experimental per a cada compost.

La divisió del conjunt de dades en *training* i *prediction sets*, que s'utilitzen en l'establiment de models lineals per regressió i no lineals amb SVM, proporciona subconjunts de dades que contenen 427 i 107 valors experimentals de $\log \eta$, respectivament. El *cross-validation set*, necessari per establir els models amb BPNN, ha de contenir un 10% dels valors del *tset*. Així, en els models desenvolupats amb BPNN, el *tset* conté 385 valors de $\log \eta$ mentre que el *cvset* en conté 42.

Model lineal

Els models lineals es construeixen mitjançant regressió multilíneal amb el mètode heurístic implementat per CODESSA. Els descriptors inicials són els mateixos que en el cas anterior, 968, més un. En aquest cas s'ha afegit la temperatura com a descriptor extern ja que aquest paràmetre és el que ha de permetre la distinció entre un mateix compost a diferents temperatures. La temperatura s'importa en forma del seu invers $1/T$, en Kelvin, d'aquesta manera es facilita la construcció i la comprensió dels models que es construiran. El mètode heurístic redueix aquest nombre inicial de descriptors a 215.

El model de set descriptors obtingut amb la regressió multilíneal implementada en el mètode heurístic de CODESSA per als 427 compostos que formen el *tset* es mostra en la Taula 3.11.

Taula 3.11. Model obtingut per regressió lineal.

Descriptor	coeficient	<i>sd</i>	β	test- <i>t</i>	VIF
Terme independent	5.003	0.916	-	5.46	-
Nombre d'àtoms de carboni	0.086	0.002	1.09	35.68	2.75
Índex de topografia electrònica d'enllaços	-0.358	0.026	-0.44	-13.89	2.92
Diferència entre superfícies carregades, DPSA-3	0.021	0.002	0.30	13.14	1.54
HDCA-2/TMSA	759.56	62.67	0.25	12.12	1.27
València mínima d'un àtom de carboni	-1.756	0.243	-0.17	-7.23	1.64
Ordre d'enllaç de l'àtom metàl·lic	-1.039	0.097	-0.30	-10.70	2.26
$1/T$	686.80	39.38	0.35	17.44	1.22

La correlació entre els set descriptors es troba en el rang 0.01-0.72, amb un valor mitjà de 0.20. El model presenta un valor VIF mitjà de 1.94, que indica molt poca col·linealitat entre descriptors. Segons els coeficients de regressió estandarditzats, els descriptors més importants dins la correlació són el nombre d'àtoms de carboni, l'índex de topografia electrònica d'enllaços i la temperatura.

La Taula 3.12 mostra els resultats que s'obtenen amb aquest models per als compostos del subconjunts *tset* i *pset*. L'estabilitat del model és força gran, tal i com demostra el valor del coeficient de correlació de validació creuada obtingut: $R^2_{cv}=0.851$. Els valors de viscositat calculats amb aquest model es troben en la Taula 3.10 i la seva representació gràfica enfront dels valors experimentals es pot veure en la Figura 3.5.

Taula 3.12. Paràmetres estadístics obtinguts amb el model lineal.

Conjunt	n	R ²	F	rmse
tset	427	0.859	2592.8	0.145
pset	107	0.870	700.8	0.135

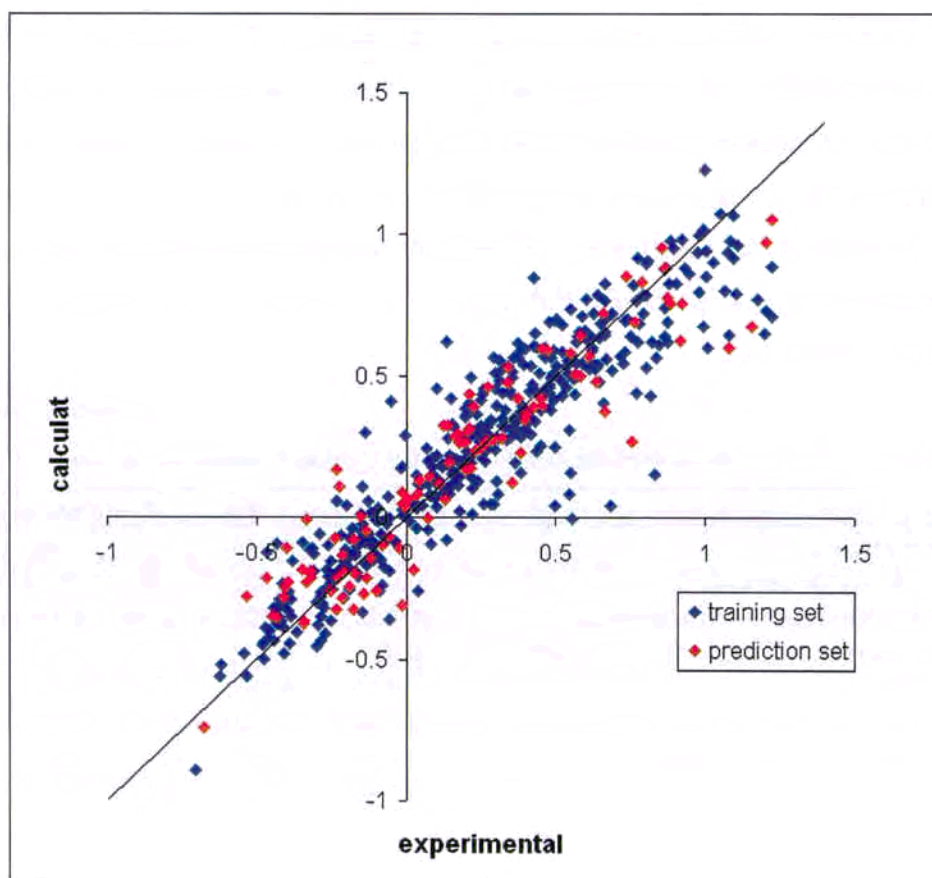


Figura 3.5. Representació gràfica de log η calculat vs. experimental amb el model lineal.

El model conté, a més de la temperatura, un descriptor constitucional: el nombre d'àtoms de carboni; tres descriptors electrostàtics: l'índex de topografia electrònica d'enllaços, DPSA-3 i HDCA-2/TMSA, sent aquests dos últims del tipus CPSA; i dos descriptors quàntics: la valència mínima d'un àtom de carboni i l'ordre d'enllaç de l'àtom metàl·lic. El descriptor índex de topografia electrònica d'enllaços, que s'obté a partir de la distribució de càrregues deduïda pel mètode de Zefirov,⁴¹ es calcula:⁴⁶

$$T_2^E = \sum_{i < j}^{N_R} \frac{|q_i - q_j|}{r_{ij}^2}$$

on N_B és el nombre d'enllaços en la molècula, q_i i q_j són les càrregues parcials els parells d'àtoms enllaçats i r_{ij} és la distància que els separa. Aquest descriptor, que és una combinació de paràmetres electrònics i topològics de la molècula, mesura la capacitat de la molècula d'establir interaccions intermoleculares electrostàtiques. El descriptor DPSA-3 ja s'ha comentat abans perquè apareixia en el model lineal del conjunt de 205 compostos organometàl·lics. El descriptor HDCA-2/TMSA mesura la capacitat de les molècules per formar enllaços d'hidrogen per donació. Aquest descriptor es calcula a partir de l'àrea donadora d'hidrogen de la molècula HDCA-2 i la seva superfície total TMSA:

$$HDCA-2 = \sum_D \frac{q_D \sqrt{S_D}}{\sqrt{S_{TOT}}}$$

on q_D és la càrrega parcial sobre l'àtom acceptor de l'enllaç d'hidrogen, S_D és la superfície d'aquest àtom i S_{TOT} és la superfície molecular accessible al solvent. A més de permetre el càlcul de la viscositat, els descriptors triats en aquest model permeten la interpretació fisicoquímica de la propietat en funció de la informació que contenen. El descriptor nombre d'enllaços de carboni presenta signe positiu en el model, la qual cosa indica que com més gran sigui el nombre de carbonis més viscosos serà el compost. Lògicament, lligands més voluminosos donen lloc a interaccions intermoleculares més grans. Els tres descriptors electrostàtics contenen la informació que fa referència a les interaccions intermoleculares electrostàtiques i d'enllaç d'hidrogen. L'ordre d'enllaç de l'àtom metàl·lic, derivat de la teoria d'ordres d'enllaç de Mulliken,^{47,48} presenta signe negatiu en el model la qual cosa es pot interpretar de la següent manera: els valors més petits d'aquest descriptor es troben en compostos que presenten lligands enllaçats al metall a través d'àtoms de nitrogen i oxigen mentre que els valors més grans es troben en compostos on hi ha enllaç entre l'àtom metàl·lic i àtoms de carboni. Això vol dir que, tenint en compte el signe del descriptor, els compostos on els lligands es troben units al metall per oxigen o nitrogen són més viscosos que els que estan units per àtoms de carboni, aquest comportament es pot relacionar amb la major tendència a establir interaccions electrostàtiques i d'enllaços d'hidrogen que presenten els lligands amb heteroàtoms. La temperatura presenta signe positiu dintre de la correlació, indicant tal i com s'esperava, que a temperatures més altes la viscositat disminueix.

Els resultats obtinguts són estadísticament molt semblants als que s'obtenen amb el model lineal desenvolupat per calcular aquesta mateixa propietat a 20°C. Els descriptors dels dos models lineals són del mateix tipus, encara que només un descriptor es repeteix: DPSA-3. Si es comparen les representacions gràfiques dels valors experimentals i calculats per a tots dos models es pot comprovar que ambdós presenten una tendència a subestimar els valors de viscositat més elevats, com per exemple, per a $\text{Sn}(\text{O}(\text{I},\text{I}-\text{Me}_2)\text{Pr})_4$ a 25 i 30°C, $\text{POMe}(\text{Cl})(\text{OBU})_2$ a 30°C i $\text{PO}(\text{OEt})_2(\text{O}(4-\text{NO}_2-\text{Ph}))$ a 25°C, sent aquests quatre compostos possibles *outliers*.

Model no lineal: Xarxes neuronals BPNN

Els models no lineals BPNN es construeixen amb el programa ADAPT. Amb aquest programa, el nombre de descriptors inicials es redueix de 215 a 114. L'arquitectura de la xarxa neuronal és, en aquest cas, 7-5-1. Així, el nombre de paràmetres ajustables és de 46 i el *ratio* d'aquest nombre sobre els compostos del *tset*, 385, és de 8.37, molt per sobre del 2.0 recomanat.

El model no lineal proposat conté els descriptors que es mostren en la Taula 3.13. Com es pot comprovar, aquest model conté exactament els mateixos tipus i nombre de descriptors que el model lineal desenvolupat per a aquest conjunt de compostos: un de constitucional, tres d'electrostàtics, dos quàntics i, òbviament, la temperatura.

Taula 3.13. Descriptors continguts en el model BPNN.

Descriptor	Classe
Nombre relatiu d'àtoms d'oxigen	Constitucional
Càrrega parcial màxima d'un àtom d'hidrogen	Electrostàtic
Diferència entre superfícies carregades, DPSA-2	CPSA
Càrrega relativa positiva, RPCG	CPSA
Contribució antiennllaçant màxima d'un orbital molecular	Quàntic
Repulsió total molecular electró-electró d'un centre	Quàntic
1/T	-

Els únics descriptors que no han aparegut fins a aquest moment són els dos quàntics, la resta de descriptors ja s'han trobat en els diferents models construïts en aquesta secció i, per aquesta raó no se'ls tornarà a explicar aquí. El descriptor contribució antienllaçant màxima d'un orbital molecular és una mesura de la força de les interaccions moleculars que caracteritza l'estabilitat dels compostos. La repulsió total molecular electró-electró d'un centre es calcula:

$$E_{ee}^{tot} = \sum_A E_{ee}(A)$$

on $E_{ee}(A)$ és l'energia de repulsió interelectrònica per a l'àtom A , aquesta repulsió interelectrònica es mesura amb les integrals de repulsió electrònica adients. L'últim descriptor, la repulsió total molecular electró-electró d'un centre és un dels termes utilitzats per calcular l'energia total de les molècules, descriu les interaccions que es donen a través de repulsions interelectròniques i es pot relacionar amb canvis conformacionals o a la reactivitat atòmica. Els descriptors inclosos en aquest model permeten també fer una interpretació fisicoquímica de la propietat estudiada en funció de la informació que contenen, així, els descriptors fan referència a les interaccions intermoleculars que s'estableixen entre les molècules del líquid per electrostàtica o per formació d'enllaços d'hidrogen. La temperatura també s'inclou en el model, demostrant que la metodologia QSPR és capaç de captar la importància d'aquest paràmetre en el càlcul de la viscositat.

Els resultats que s'obtenen amb aquest model no lineal per als subconjunts de dades: *tset*, *cvset* i *pset* es troben en la Taula 3.14, i els valors de viscositat calculats es recullen en la Taula 3.10 i la seva representació gràfica enfront dels valors experimentals es pot veure en la Figura 3.6.

Taula 3.14. Paràmetres estadístics obtinguts amb el model BPNN.

Conjunt	n	R ²	F	rmse
<i>tset</i>	385	0.947	6875.1	0.093
<i>cvset</i>	42	0.941	635.5	0.095
<i>pset</i>	107	0.947	1879.6	0.093

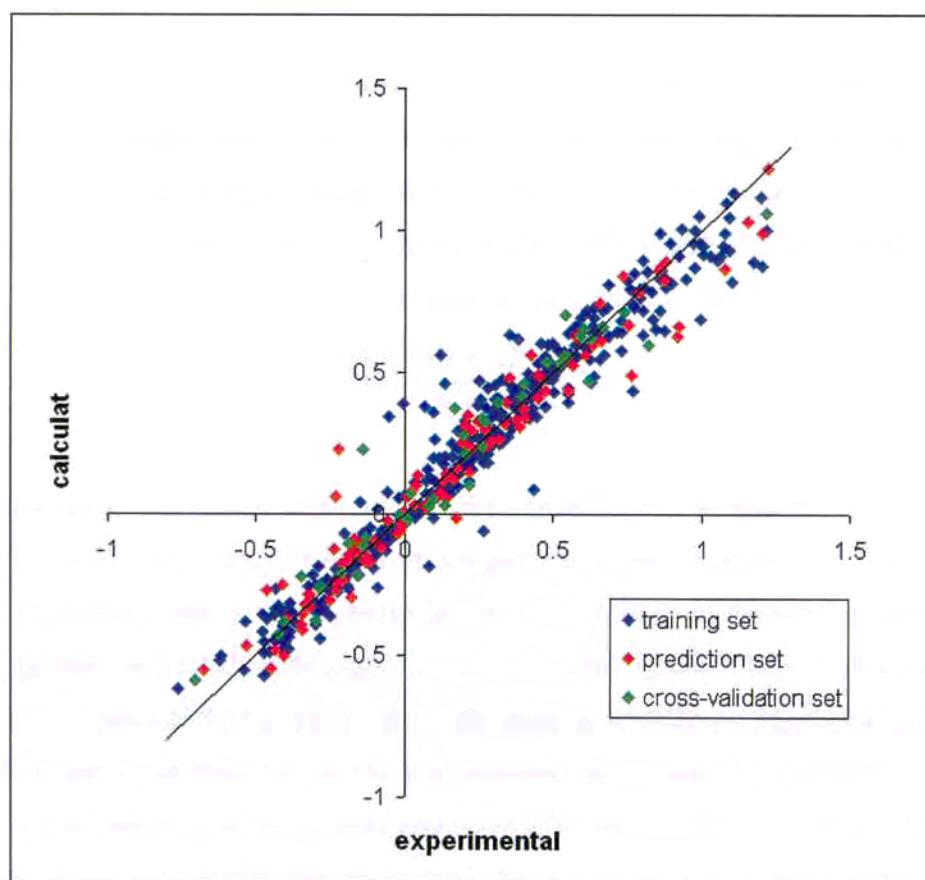


Figura 3.6. Representació gràfica de $\log \eta$ calculat vs. experimental amb el model BPNN.

Es pot comprovar que els resultats que s'obtenen amb els tres subconjunts de dades són molt bons i, a més, són molt semblants entre sí, la qual cosa demostra la gran robustesa del model construït. També es pot veure que aquests resultats són una mica millors als que s'obtenen en el model BPNN del conjunt de 205 compostos organometàl·lics on no es té en compte l'efecte de la temperatura.

El *rmse* del model per a tots els compostos és de 0.093, en total hi ha 33 compostos que presenten valors de residual més grans que dues vegades l'error del model, Taula 3.15. Dintre d'aquests n'hi ha 16 que tenen errors més grans que tres vegades el *rmse* i, per tant, es podrien considerar *outliers*. Tot i així no sembla haver una relació clara entre l'error d'aquests compostos i la seva estructura, metall o temperatura.

Taula 3.15. Prediccions amb valors més grans que dues vegades el *rmse**

Compost	T (°C)	Exp.	BPNN	residual
Si(3Cl-Pr)(OMe) ₃	20	-0.222	0.229	-0.451
PO(O'Pr) ₂ F	25	0.122	0.561	-0.439
SiPh(CH=CH ₂)(CH ₂ -CH=CH ₂) ₂	25	-0.051	0.341	-0.392
Ge(O'Pr) ₄	40	-0.004	0.387	-0.391
SiMe ₂ (OPr) ₂	20	-0.137	0.226	-0.363
GeBu ₄	61	0.134	0.458	-0.324
Si(CH ₂ -CH=CH ₂) ₂ (OEt) ₂	0	0.07	0.38	-0.31
SiPh(CH=CH ₂)(CH ₂ -CH=CH ₂) ₂	60	-0.233	0.06	-0.293
SiPh ₃ Cl	100	0.356	0.634	-0.278
SiMe ₂ PhBr	45	0.099	0.357	-0.258
PO(O'Pr) ₃	23	0.385	0.618	-0.233
SiPh(OEt) ₃	20	0.25	0.469	-0.219
SiMe ₂ PhBr	35	0.168	0.372	-0.204
SiMe ₂ (CH=CH ₂)(OMe)	20	-0.463	-0.271	-0.192
SiEt ₂ Cl ₂	20	-0.151	0.039	-0.19
Si(O'Pr) ₄	20	0.173	-0.018	0.191
SiMePh ₂ Cl	20	0.846	0.652	0.194
P(OPh) ₃	25	1.205	0.989	0.216
PO(MeCl)(OPr) ₂	30	1.079	0.863	0.216
POHex ₃	40	1.221	1.002	0.219
PPh ₃	83	0.822	0.594	0.228
SiMePh ₂ (CH ₂ -CH=CH ₂)	20	0.877	0.647	0.23
Sn(O(1,1-Me ₂)Bu) ₄	30	0.924	0.661	0.263
PO(OEt) ₃	40	0.083	-0.187	0.27
SiPh ₂ Cl ₂	20	0.765	0.491	0.274
SiPh ₂ (CH ₂ -CH=CH ₂) ₂	20	1.104	0.819	0.285
PO(OMe)(OPh) ₂	25	1.176	0.887	0.289
Si(O(2-OEt-Et)) ₄	20	0.92	0.625	0.295
PO(COPh)(OEt) ₂	25	0.998	0.686	0.312
Si(3-Cl-Pr)Cl ₃	20	0.266	-0.059	0.325
PO(MeCl)(OBu) ₂	30	1.204	0.875	0.329
POPh(OEt) ₂	25	0.77	0.433	0.337
P(NMe ₂) ₃	30	0.435	0.086	0.349

* Els compostos ombrejats tenen residuals més grans que tres vegades el *rmse*

Una altra manera de comprovar la robustesa del model obtingut consisteix en analitzar els resultats que s'obtenen amb els diferents subconjunts de compostos classificats pel metall que contenen. La Taula 3.16 mostra els paràmetres estadístics derivats pel model per cadascun d'aquests subconjunts.

Taula 3.16. Estadística per als conjunts de compostos.

Conjunt	n	R ²	F	rmse
As	53	0.979	2358.6	0.041
Ge	82	0.954	1676.3	0.082
P	153	0.911	1551.1	0.097
Si	210	0.932	2837.9	0.102
Sn	36	0.965	945.3	0.080

Com es pot comprovar, els resultats obtinguts per a aquests subconjunts de compostos són del mateix ordre que els que s'obtenen amb el model global i, per tant, es pot dir que el model és prou robust. Clarament, els compostos per als quals s'obtenen els millors resultats són els d'arsènic, tot i ser un dels conjunts amb menys valors. No s'ha trobat cap compost que presenti un valor de residual més de dues vegades el *rmse* del model. En una situació similar es troben els compostos d'estany, que només presenten un compost, el Sn(O(1,1-Me₂)Bu)₄ a 30°C, amb residual superior dos cops a aquest valor. Dels compostos de germani només n'hi ha dos que apareixen en la Taula 3.15: Ge(O'Pr)₄ a 40°C i GeBu₄ a 61°C; això sí, tots dos presenten residuals molt grans, que converteixen a aquests dos compostos en *outliers* i que, probablement empitjoren els resultats obtinguts per a aquest conjunt de compostos. Els compostos que contenen fòsfor, el segon conjunt més gran en nombre, tenen el pitjor valor de R² i presenten un nombre bastant elevat de compostos amb valors de residual grans, sobretot en la banda dels residuals positius. Els resultats obtinguts per als compostos de silici presenten un comportament invers al que s'ha trobat per als de fòsfor, amb la majoria de residuals amb signe negatiu. L'origen de les dades experimentals pot constituir una possible explicació per a aquest fet; hi ha una gran quantitat de determinacions experimentals per a aquests conjunts de compostos que s'han realitzat en diferents grups d'investigació amb mètodes diferents. Aquesta dispersió en les fonts bibliogràfiques comporta una menor cohesió interna del conjunt de valors experimentals determinats per a aquests compostos, augmentant-ne així la dificultat de predicció d'aquests valors.

Model no lineal: Support Vector Machines

En aquest apartat es descriu la deducció dels models amb el mètode de selecció de descriptors no lineal, heurístic-SVM. Els descriptors inicials utilitzats en aquest procediment són els mateixos 215 que s'han emprat en l'establiment del model lineal. El model trobat conté els set descriptors que es troben en la Taula 3.17.

Taula 3.17. Descriptors continguts en el model SVM.

Descriptor	Classe
Nombre relatiu d'àtoms d'oxigen	Constitucional
Índex de Randic d'ordre 3	Topològic
Càrrega parcial màxima d'un àtom d'hidrogen	Electroestàtic
Àrea superficial positiva fraccionada, FPSA-2	CPSA
Àrea superficial parcial negativa, PNSA-3	CPSA
Repulsió total molecular electró-electró d'un centre/nombre d'àtoms	Quàntic
1/T	-

A més de la temperatura, el model conté descriptors que ja han anat sortint en els diversos models desenvolupats anteriorment: el nombre relatiu d'àtoms d'oxigen, la càrrega parcial màxima sobre un àtom d'hidrogen i FPSA-2; aquests descriptors ja han estat explicats i no es tornaran a comentar aquí. El descriptor PNSA-3 ha estat definit abans en l'explicació del descriptor DPSA-3. El descriptor topològic índex de Randic⁴⁹ d'ordre 3 es calcula en base a la diferent coordinació de cada àtom en la molècula i conté informació sobre la grandària i la ramificació de l'estructura. Aquesta informació es pot relacionar directament amb les interaccions intermoleculares que es donen entre les molècules del compost en estat líquid. El descriptor repulsió total molecular electró-electró d'un centre entre el nombre d'àtoms és completament anàleg al que s'ha explicat en l'apartat anterior, amb una única diferència: en aquest cas es troba dividit pel nombre d'àtoms de la molècula. Els descriptors inclosos en aquest model permeten la interpretació fisicoquímica de la viscositat en funció de la informació que conté cadascun d'ells. De la mateixa manera que en els models anteriors, el descriptor topològic conté informació sobre les interaccions intermoleculares que es donen degut a la ramificació i hibridació dels àtoms en els substituents orgànics, la resta de descriptors la informació relacionada amb l'establiment d'interaccions intermoleculares de caràcter electroestàtic i per enllaç d'hidrogen.

El model SVM definitiu s'obté optimitzant els paràmetres característics γ , ϵ i C tal i com es mostra en la Figura 3.7. L'optimització del paràmetre γ (amb $\epsilon=0.1$ i $C=100$) condueix a un valor òptim per aquest paràmetre de 0.7. Un cop optimitzat aquest paràmetre, s'optimitza ϵ (amb $\gamma=0.7$ i $C=100$) i el seu millor valor és 0.07. El paràmetre C s'optimitza amb $\gamma=0.7$ i $\epsilon=0.07$; el valor trobat per aquest paràmetre és 90.

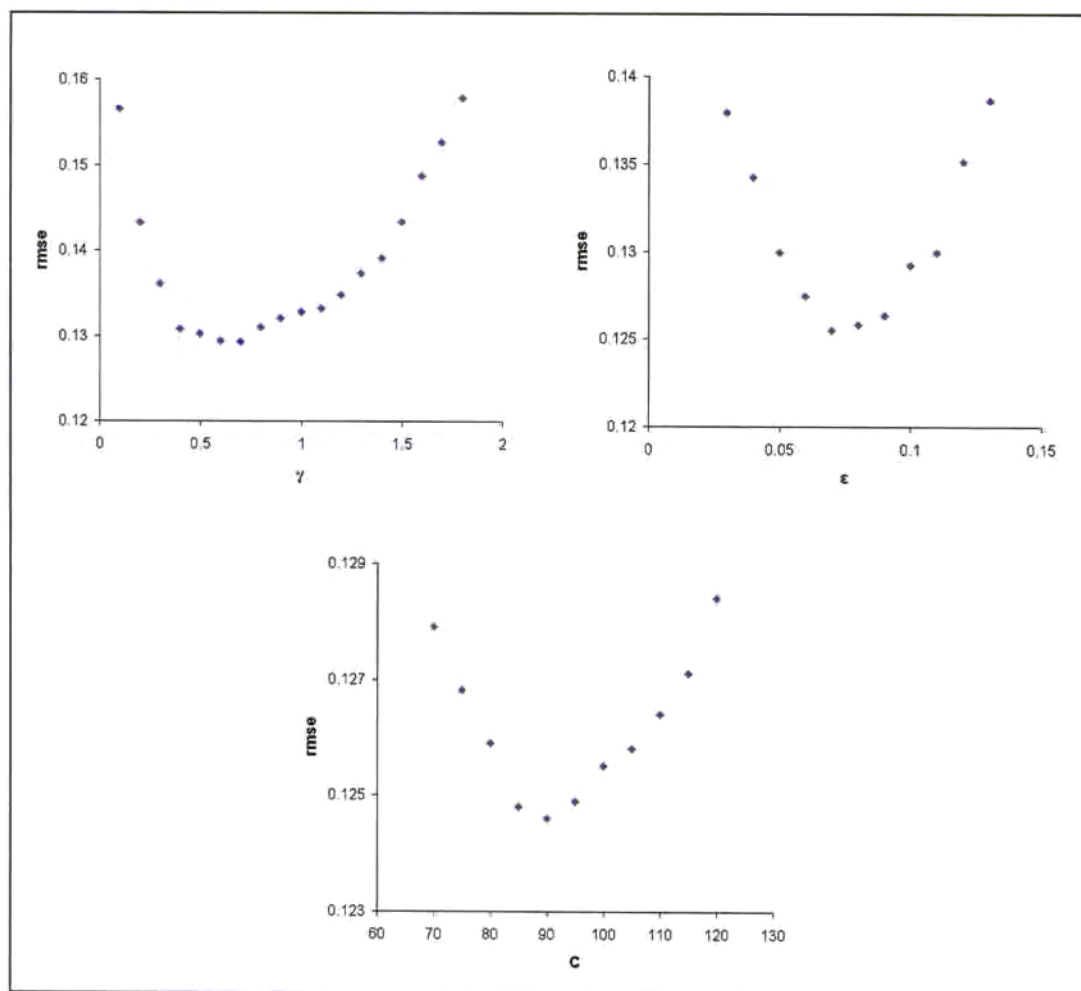


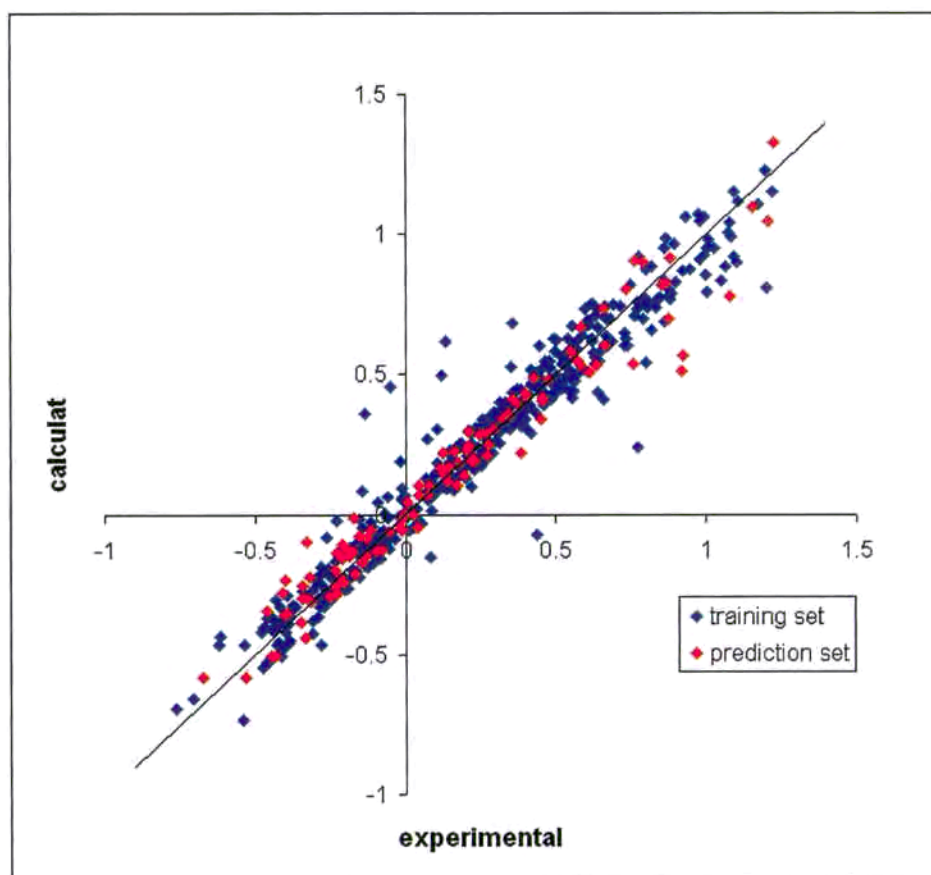
Figura 3.7. Optimització dels paràmetres γ , ϵ i C .

El model desenvolupat consta de 175 vectors suport, és a dir, s'utilitza aquest nombre de compostos d'un total de 427 del *tset* per establir el model. Els resultats estadístics que s'obtenen amb aquest model són molt bons i semblants als del model BPNN, la Taula 3.18 mostra els valors del coeficient de determinació R^2 , el test- F i *rmse* per als dos subconjunts: *tset* i *pset*.

Taula 3.18. Paràmetres estadístics obtinguts amb el model SVM.

Conjunt	n	R^2	F	<i>rmse</i>
<i>tset</i>	427	0.945	7314.0	0.095
<i>pset</i>	107	0.953	2150.0	0.087

Els valors de viscositat calculats amb aquest model es troben en la Taula 3.10 i la seva representació gràfica enfront dels valors experimentals es pot veure en la Figura 3.8.

Figura 3.8. Representació gràfica de $\log \eta$ calculat vs. experimental amb el model SVM.

Els resultats obtinguts analitzant els subconjunts de compostos en funció de l'àtom metàl·lic són pràcticament iguals als que dona el model establert amb BPNN, demostrant que els dos mètodes no lineals es poden utilitzar de manera indistinta per calcular la viscositat. Cal dir que hi ha uns quants compostos que apareixen com a *outliers* molt forts en aquests dos models, la qual cosa es podria relacionar amb errors en la seva determinació experimental: $\text{POPh}(\text{OEt})_2$, $\text{SiPh}(\text{CH}=\text{CH}_2)(\text{CH}_2-\text{CH}=\text{CH}_2)_2$, $\text{P}(\text{NMe}_2)_3$, $\text{Si}(\text{O}(2-\text{OEt}-\text{Et}))_4$, $\text{POMe}(\text{Cl})(\text{OBu})_2$ i SiPh_3Cl en són alguns exemples.

Comparació de resultats

Com ja s'ha comentat, els resultats obtinguts amb aquests compostos organometàl·lics només es poden comparar amb un estudi d'aquest tipus per a un conjunt de més de tres mil quatre-cents compostos orgànics.⁴⁰ En aquest treball s'utilitzen xarxes neuronals de tipus BPNN per establir models no lineals amb 307 descriptors i els resultats que s'obtenen: $R^2=0.989$ i $rmse=0.14$, són semblants als aquí obtinguts.

Els descriptors que contenen els tres models aquí desenvolupats són prou semblants entre ells i són també semblants als que s'obtenen en els models proposats per calcular la viscositat a 20°C. En els tres últims models l'únic descriptor que es repeteix és l'invers de la temperatura. Hi ha dos descriptors que apareixen en els dos models no lineals: el nombre relatiu d'àtoms d'oxigen i la càrrega parcial màxima d'un àtom d'hidrogen. En aquests dos models es pot trobar un altre descriptor gairebé idèntic: la repulsió total molecular electró-electró d'un centre que, en el model BPNN es presenta d'aquesta forma, i en el mode SVM apareix dividida entre el nombre d'àtoms de la molècula. La resta de descriptors continguts en els tres models són, majoritàriament, de tipus CPSA: DPSA-2, DPSA-3, PNSA-3, FPSA-2, RPCG i HDCA-2/TMSA, relacionats amb les interaccions intermoleculares que poden condicionar el valor de la viscositat. Si es comparen els descriptors d'aquests tres models amb els que s'han deduït a 20°C, es pot comprovar que també hi ha alguns dels descriptors CPSA, importants en aquest estudi, que es repeteixen: DPSA-2, DPSA-3, FPSA-2, RPCG. També hi ha altres descriptors que es repeteixen en aquests models: el nombre relatiu d'àtoms d'oxigen i la càrrega parcial màxima d'un àtom d'hidrogen; i d'altres que, tot i no ser idèntics, sí que es poden relacionar entre ells.

En aquest cas, els resultats obtinguts amb els mètodes no lineals són, de nou, superiors als del mètode lineal. Estadísticament parlant, els errors del mètode lineal són gairebé un 50% més elevats que els dels no lineals, 0.15 enfront de 0.1 unitats logarítmiques, aproximadament. Els valors de R^2 i el nombre d'*outliers* i la dispersió en els errors dels valors predits són també bastant més elevats en el cas del model lineal. Val a dir que els dos models no lineals, tot i contenir diferents descriptors, són equivalents a efectes de càlcul.

Taula 3.10. Valors experimentals i calculats de $\log \eta$ (mPa·s) a diferents temperatures.

#	Compost	T (°C)	Exp.	MLR	BPNN	SVM
1	As(OBu) ₃ ^{a,b}	22	0.248	0.414	0.269	0.212
2	As(OBu) ₃ ^d	25	0.228	0.387	0.234	0.194
3	As(OBu) ₃ ^{a,b}	30	0.198	0.353	0.279	0.171
4	As(OBu) ₃ ^{a,b}	40	0.141	0.277	0.115	0.118
5	As(OBu) ₃ ^{a,b}	50	0.087	0.208	0.053	0.068
6	As(OBu) ₃ ^{a,b}	60	0.036	0.146	0.003	0.025
7	As(OBu) ₃ ^d	70	-0.011	0.085	-0.041	-0.016
8	As(OEt) ₃ ^{a,b}	37	-0.125	-0.058	-0.157	-0.189
9	As(OEt) ₃ ^{a,b}	47	-0.167	-0.126	-0.197	-0.216
10	As(OEt) ₃ ^d	57	-0.208	-0.195	-0.215	-0.245
11	As(OEt) ₃ ^{a,b}	67	-0.245	-0.257	-0.259	-0.272
12	As(O'Bu) ₃ ^{a,b}	22	0.251	0.323	0.326	0.175
13	As(O'Bu) ₃ ^{a,b}	25	0.230	0.296	0.293	0.160
14	As(O'Bu) ₃ ^d	30	0.196	0.262	0.254	0.140
15	As(O'Bu) ₃ ^{a,b}	40	0.133	0.186	0.177	0.096
16	As(O'Bu) ₃ ^{a,b}	50	0.074	0.117	0.117	0.054
17	As(O'Bu) ₃ ^{a,c}	60	0.018	0.056	0.069	0.018
18	As(O'Pent) ₃ ^{a,b}	22	0.382	0.592	0.440	0.359
19	As(O'Pent) ₃ ^{a,b}	25	0.359	0.564	0.402	0.335
20	As(O'Pent) ₃ ^{a,b}	30	0.325	0.530	0.357	0.305
21	As(O'Pent) ₃ ^{a,b}	40	0.259	0.454	0.268	0.236
22	As(O'Pent) ₃ ^{a,b}	50	0.198	0.386	0.194	0.173
23	As(O'Pent) ₃ ^d	60	0.140	0.324	0.142	0.118
24	As(O'Pr) ₃ ^{a,b}	29	0.056	0.040	0.033	-0.034
25	As(O'Pr) ₃ ^{a,c}	37	0.013	-0.015	-0.021	-0.057
26	As(O'Pr) ₃ ^{a,b}	47	-0.038	-0.084	-0.079	-0.088
27	As(O'Pr) ₃ ^{a,b}	57	-0.085	-0.152	-0.130	-0.119
28	As(O'Pr) ₃ ^d	67	-0.130	-0.214	-0.170	-0.146
29	As(OMe) ₃ ^{a,b}	20	-0.178	-0.040	-0.171	-0.201
30	As(OMe) ₃ ^{a,b}	30	-0.204	-0.116	-0.235	-0.226
31	As(OMe) ₃ ^{a,b}	40	-0.228	-0.191	-0.279	-0.260
32	As(OMe) ₃ ^d	50	-0.252	-0.260	-0.362	-0.295
33	As(OMe) ₃ ^{a,b}	60	-0.273	-0.322	-0.332	-0.329
34	As(OMe) ₃ ^{a,b}	70	-0.293	-0.383	-0.351	-0.363
35	As(OPent) ₃ ^{a,c}	22	0.397	0.594	0.414	0.370
36	As(OPent) ₃ ^{a,b}	25	0.375	0.567	0.375	0.347
37	As(OPent) ₃ ^{a,b}	30	0.340	0.532	0.330	0.316
38	As(OPent) ₃ ^d	40	0.275	0.457	0.239	0.248
39	As(OPent) ₃ ^{a,b}	50	0.213	0.388	0.168	0.185
40	As(OPent) ₃ ^{a,b}	60	0.155	0.326	0.111	0.132
41	As(OPr) ₃ ^{a,b}	28	0.089	0.176	0.071	0.011
42	As(OPr) ₃ ^{a,b}	37	0.039	0.115	0.016	-0.019
43	As(OPr) ₃ ^d	47	-0.015	0.046	-0.037	-0.055
44	As(OPr) ₃ ^{a,b}	57	-0.064	-0.023	-0.084	-0.092
45	As(OPr) ₃ ^{a,b}	67	-0.111	-0.085	-0.123	-0.124
46	AsEt ₃ ^{a,b}	20	-0.378	-0.156	-0.354	-0.327
47	AsEt ₃ ^d	30	-0.395	-0.232	-0.376	-0.356
48	AsEt ₃ ^{a,b}	40	-0.411	-0.307	-0.401	-0.382
49	AsEt ₃ ^{a,c}	50	-0.426	-0.376	-0.427	-0.400
50	AsEt ₃ ^{a,b}	60	-0.440	-0.438	-0.360	-0.413

#	Compost	T (°C)	Exp.	MLR	BPNN	SVM
51	AsMe ₃ ^d	20	-0.447	-0.349	-0.551	-0.505
52	AsMe ₃ ^{a,b}	30	-0.461	-0.425	-0.565	-0.526
53	AsMe ₃ ^{a,b}	40	-0.475	-0.500	-0.494	-0.545
54	Ge(OBu) ₄ ^d	25	0.340	0.474	0.320	0.350
55	Ge(OBu) ₄ ^{a,b}	40	0.203	0.364	0.185	0.236
56	Ge(OEt) ₄ ^{a,b}	20	0.024	0.050	0.010	-0.047
57	Ge(OEt) ₄ ^d	40	-0.078	-0.101	-0.120	-0.126
58	Ge(O ^t Bu) ₄ ^d	20	0.449	0.420	0.531	0.340
59	Ge(O ^t Bu) ₄ ^{a,b}	25	0.403	0.379	0.351	0.304
60	Ge(O ^t Bu) ₄ ^{a,b}	40	0.276	0.269	0.219	0.206
61	Ge(O ⁱ Pr) ₄ ^{a,b}	20	0.119	0.186	0.197	0.108
62	Ge(O ⁱ Pr) ₄ ^d	40	-0.004	0.035	0.387	0.002
63	Ge(OMe) ₄ ^{a,b}	20	-0.112	-0.104	-0.122	-0.042
64	Ge(OMe) ₄ ^d	25	-0.116	-0.146	-0.174	-0.060
65	Ge(OMe) ₄ ^{a,b}	40	-0.188	-0.255	-0.267	-0.121
66	Ge(OPr) ₄ ^d	20	0.179	0.262	0.295	0.186
67	Ge(OPr) ₄ ^{a,b}	40	0.051	0.111	-0.001	0.071
68	Ge(O ^t Bu) ₄ ^{a,b}	20	0.449	0.421	0.468	0.461
69	Ge(O ^t Bu) ₄ ^{a,c}	25	0.402	0.380	0.462	0.415
70	Ge(O ^t Bu) ₄ ^d	40	0.270	0.270	0.307	0.290
71	Ge(O ^t Bu) ₄ ^d	25	0.757	0.269	0.666	0.539
72	Ge(O ^t Bu) ₄ ^{a,b}	40	0.553	0.159	0.445	0.418
73	GeBu ₄ ^{a,b}	2	0.729	0.542	0.641	0.602
74	GeBu ₄ ^d	9	0.635	0.480	0.596	0.532
75	GeBu ₄ ^{a,b}	15	0.566	0.432	0.547	0.477
76	GeBu ₄ ^{a,b}	22	0.482	0.377	0.479	0.413
77	GeBu ₄ ^{a,b}	27	0.430	0.336	0.420	0.366
78	GeBu ₄ ^{a,b}	30	0.397	0.315	0.388	0.342
79	GeBu ₄ ^{a,b}	35	0.349	0.281	0.333	0.304
80	GeBu ₄ ^{a,b}	41	0.299	0.240	0.263	0.259
81	GeBu ₄ ^{a,b}	45	0.256	0.205	0.204	0.224
82	GeBu ₄ ^{a,b}	47	0.218	0.192	0.181	0.210
83	GeBu ₄ ^d	50	0.217	0.171	0.145	0.189
84	GeBu ₄ ^{a,b}	55	0.180	0.144	0.098	0.163
85	GeBu ₄ ^{a,c}	61	0.134	0.102	0.458	0.126
86	GeBu ₄ ^{a,b}	66	0.099	0.075	-0.014	0.104
87	GeEt ₄ ^{a,b}	5	-0.056	-0.004	-0.099	-0.068
88	GeEt ₄ ^{a,b}	11	-0.086	-0.059	-0.079	-0.110
89	GeEt ₄ ^{a,b}	16	-0.109	-0.100	-0.118	-0.140
90	GeEt ₄ ^{a,b}	22	-0.109	-0.148	-0.094	-0.174
91	GeEt ₄ ^{a,b}	28	-0.165	-0.196	-0.179	-0.206
92	GeEt ₄ ^{a,b}	34	-0.190	-0.244	-0.181	-0.236
93	GeEt ₄ ^{a,c}	40	-0.210	-0.286	-0.213	-0.260
94	GeEt ₄ ^d	46	-0.232	-0.327	-0.245	-0.282
95	GeEt ₄ ^{a,b}	53	-0.251	-0.368	-0.277	-0.302
96	GeEt ₄ ^{a,b}	59	-0.271	-0.409	-0.309	-0.319
97	GeEt ₄ ^{a,b}	64	-0.287	-0.437	-0.330	-0.330
98	GeHex ₄ ^{a,b}	3	1.096	1.067	1.047	1.154
99	GeHex ₄ ^{a,b}	11	0.979	0.998	1.001	1.045
100	GeHex ₄ ^{a,b}	17	0.896	0.950	0.957	0.966
101	GeHex ₄ ^{a,b}	23	0.802	0.895	0.894	0.875

#	Compost	T (°C)	Exp.	MLR	BPNN	SVM
102	GeHex ₄ ^d	29	0.734	0.854	0.837	0.807
103	GeHex ₄ ^{a,b}	40	0.603	0.771	0.700	0.673
104	GeHex ₄ ^{a,b}	46	0.552	0.737	0.635	0.620
105	GeHex ₄ ^{a,c}	53	0.481	0.689	0.537	0.549
106	GeHex ₄ ^{a,b}	59	0.431	0.648	0.450	0.491
107	GeHex ₄ ^{a,b}	65	0.376	0.613	0.332	0.446
108	GeMe ₄ ^{a,b}	4	-0.393	-0.265	-0.477	-0.459
109	GeMe ₄ ^{a,b}	9	-0.414	-0.313	-0.483	-0.484
110	GeMe ₄ ^d	14	-0.433	-0.354	-0.338	-0.504
111	GeMe ₄ ^{a,b}	20	-0.454	-0.402	-0.491	-0.527
112	GeMe ₄ ^{a,b}	26	-0.475	-0.444	-0.574	-0.545
113	GePent ₄ ^{a,b}	3	0.978	0.797	0.870	0.908
114	GePent ₄ ^{a,b}	10	0.864	0.742	0.716	0.833
115	GePent ₄ ^{a,b}	16	0.779	0.687	0.775	0.756
116	GePent ₄ ^{a,b}	22	0.864	0.639	0.830	0.687
117	GePent ₄ ^{a,b}	28	0.623	0.591	0.646	0.619
118	GePent ₄ ^{a,c}	36	0.536	0.536	0.554	0.543
119	GePent ₄ ^{a,b}	41	0.483	0.494	0.479	0.487
120	GePent ₄ ^d	53	0.376	0.419	0.375	0.391
121	GePent ₄ ^{a,b}	59	0.329	0.378	0.250	0.342
122	GePent ₄ ^{a,b}	65	0.282	0.343	0.183	0.305
123	GePr ₄ ^{a,b}	3	0.265	0.274	0.313	0.227
124	GePr ₄ ^d	7	0.230	0.233	0.295	0.190
125	GePr ₄ ^{a,c}	11	0.198	0.205	0.198	0.166
126	GePr ₄ ^{a,b}	16	0.160	0.164	0.203	0.129
127	GePr ₄ ^{a,b}	22	0.160	0.116	0.248	0.087
128	GePr ₄ ^{a,b}	27	0.076	0.075	0.158	0.052
129	GePr ₄ ^{a,b}	32	0.035	0.034	0.109	0.017
130	GePr ₄ ^{a,b}	37	0.003	-0.001	0.065	-0.011
131	GePr ₄ ^d	47	-0.056	-0.069	-0.065	-0.062
132	GePr ₄ ^{a,b}	51	-0.077	-0.097	-0.060	-0.081
133	GePr ₄ ^{a,b}	57	-0.103	-0.131	-0.104	-0.104
134	GePr ₄ ^{a,b}	63	-0.137	-0.166	-0.118	-0.125
135	GePr ₄ ^{a,b}	68	-0.165	-0.193	-0.148	-0.140
136	P(NMe ₂) ₃ ^{a,b}	30	0.435	0.234	0.086	-0.071
137	P(OMe) ₃ ^{a,b}	25	-0.262	-0.100	-0.249	-0.273
138	P(OMe) ₃ ^d	38	-0.324	-0.196	-0.256	-0.305
139	P(OMe) ₃ ^{a,b}	52	-0.384	-0.292	-0.316	-0.346
140	P(OMe) ₃ ^{a,b}	64	-0.433	-0.375	-0.487	-0.386
141	P(OMe) ₃ ^{a,b}	78	-0.480	-0.450	-0.357	-0.421
142	P(OMe) ₃ ^{a,b}	99	-0.535	-0.560	-0.383	-0.465
143	P(OPh) ₃ ^d	25	1.205	0.972	0.989	1.049
144	P(OPh) ₃ ^{a,b}	35	1.029	0.903	0.908	0.948
145	P(OPh) ₃ ^{a,b}	45	0.902	0.828	0.814	0.832
146	P(OPh) ₃ ^{a,b}	55	0.771	0.766	0.738	0.735
147	P(OPh) ₃ ^{a,c}	65	0.670	0.704	0.664	0.640
148	P(OPh) ₃ ^{a,b}	75	0.575	0.642	0.595	0.549
149	P(OPh) ₃ ^{a,b}	85	0.491	0.587	0.538	0.474
150	PO(COPh)(OEt) ₂ ^{a,b}	25	0.998	1.226	0.686	0.855
151	PO(NMe ₂) ₃ ^{a,b}	12	0.633	0.539	0.662	0.563
152	PO(NMe ₂) ₃ ^d	17	0.583	0.498	0.619	0.531

#	Compost	T (°C)	Exp.	MLR	BPNN	SVM
153	PO(NMe ₂) ₃ ^{a,b}	22	0.536	0.456	0.571	0.497
154	PO(NMe ₂) ₃ ^{a,b}	27	0.492	0.415	0.520	0.461
155	PO(NMe ₂) ₃ ^{a,b}	32	0.451	0.381	0.476	0.430
156	PO(NMe ₂) ₃ ^{a,b}	37	0.411	0.340	0.421	0.393
157	PO(NMe ₂) ₃ ^{a,b}	42	0.374	0.305	0.375	0.361
158	PO(NMe ₂) ₃ ^{a,b}	47	0.337	0.271	0.330	0.329
159	PO(O(2-Et-Bu)) ₃ ^d	0	1.225	1.051	1.217	1.330
160	PO(O(2-Et-Bu)) ₃ ^{a,b}	20	0.934	0.879	1.004	1.060
161	PO(O(2-Et-Bu)) ₃ ^{a,b}	60	0.462	0.597	0.603	0.526
162	PO(O(2-Me-Pr)) ₃ ^{a,b}	23	0.645	0.481	0.616	0.548
163	PO(O(2-Me-Pr)) ₃ ^{a,c}	31	0.522	0.426	0.530	0.469
164	PO(O(2-Me-Pr)) ₃ ^d	35	0.457	0.392	0.484	0.418
165	PO(O(2-Me-Pr)) ₃ ^{a,b}	40	0.391	0.357	0.439	0.366
166	PO(O(2-Me-Pr)) ₃ ^{a,b}	50	0.276	0.289	0.372	0.262
167	PO(O(2-Me-Pr)) ₃ ^{a,b}	59	0.204	0.234	0.326	0.181
168	PO(O(2-OBu-Et)) ₃ ^{a,b}	25	1.086	0.975	1.096	0.993
169	PO(OBu)(OEt) ₂ ^d	25	0.294	0.271	0.256	0.304
170	PO(OBu)(OEt) ₂ ^{a,b}	35	0.207	0.195	0.179	0.209
171	PO(OBu)(OPh) ₂ ^{a,b}	25	1.107	0.963	1.133	1.119
172	PO(OBu)(OPh) ₂ ^{a,b}	38	0.863	0.874	0.989	0.950
173	PO(OBu)(OPh) ₂ ^{a,b}	99	0.305	0.510	0.454	0.289
174	PO(OBu) ₂ (O(4-OEt-Bu)) ^d	25	0.661	0.718	0.742	0.735
175	PO(OBu) ₂ (OHex) ^{a,b}	25	0.598	0.659	0.693	0.733
176	PO(OBu) ₃ ^{a,c}	20	0.589	0.544	0.627	0.637
177	PO(OBu) ₃ ^{a,b}	30	0.479	0.468	0.507	0.526
178	PO(OBu) ₃ ^{a,b}	40	0.387	0.393	0.412	0.409
179	PO(OBu) ₃ ^{a,b}	50	0.320	0.324	0.247	0.301
180	PO(OBu) ₃ ^d	60	0.271	0.262	0.297	0.207
181	PO(OEt)(NMe ₂) ₂ ^{a,b}	25	0.293	0.562	0.238	0.356
182	PO(OEt)(NMe ₂) ₂ ^{a,b}	35	0.216	0.493	0.185	0.298
183	PO(OEt)(NMe ₂) ₂ ^{a,b}	45	0.150	0.418	0.115	0.229
184	PO(OEt) ₂ (O(4-NO ₂ -Ph)) ^{a,b}	25	1.198	0.652	1.117	1.229
185	PO(OEt) ₂ (OPent) ^d	25	0.405	0.364	0.339	0.419
186	PO(OEt) ₂ NMe ₂ ^{a,b}	25	0.218	0.285	0.116	0.257
187	PO(OEt) ₂ NMe ₂ ^{a,b}	35	0.144	0.217	0.071	0.186
188	PO(OEt) ₂ NMe ₂ ^d	45	0.077	0.141	0.036	0.102
189	PO(OEt) ₃ ^{a,b}	20	0.225	0.226	0.224	0.295
190	PO(OEt) ₃ ^{a,b}	25	0.190	0.184	0.150	0.248
191	PO(OEt) ₃ ^{a,b}	30	0.152	0.150	0.101	0.208
192	PO(OEt) ₃ ^d	35	0.114	0.116	0.062	0.166
193	PO(OEt) ₃ ^{a,c}	40	0.083	0.074	-0.187	0.114
194	PO(OHex) ₃ ^{a,b}	20	1.076	0.939	0.945	1.006
195	PO(OHex) ₃ ^{a,b}	30	0.947	0.863	0.828	0.876
196	PO(OHex) ₃ ^{a,b}	40	0.805	0.788	0.718	0.740
197	PO(OHex) ₃ ^{a,b}	50	0.688	0.719	0.627	0.617
198	PO(OHex) ₃ ^{a,b}	60	0.579	0.657	0.554	0.509
199	PO(OHex) ₃ ^d	70	0.453	0.595	0.489	0.408
200	PO(OHex) ₃ ^{a,b}	80	0.391	0.540	0.442	0.326
201	PO(OHex) ₃ ^{a,c}	90	0.313	0.485	0.395	0.253
202	PO(O ⁱ Bu)(OEt) ₂ ^{a,b}	25	0.279	0.275	0.292	0.333
203	PO(O ⁱ Pent) ₃ ^{a,c}	28	0.607	0.652	0.650	0.692

#	Compost	T (°C)	Exp.	MLR	BPNN	SVM
204	PO(O'Pent) ₃ ^d	30	0.583	0.638	0.631	0.670
205	PO(O'Pent) ₃ ^{a,b}	40	0.465	0.563	0.533	0.547
206	PO(O'Pent) ₃ ^{a,b}	50	0.356	0.494	0.460	0.436
207	PO(O'Pent) ₃ ^{a,b}	60	0.255	0.432	0.404	0.339
208	PO(O'Pr) ₂ F ^{a,b}	25	0.122	0.179	0.561	0.498
209	PO(O'Pr) ₃ ^{a,b}	23	0.385	0.281	0.618	0.315
210	PO(OMe)(OPh) ₂ ^{a,b}	25	1.176	0.772	0.887	1.106
211	PO(OMe) ₃ ^d	20	0.354	0.121	0.480	0.411
212	PO(OMe) ₃ ^{a,b}	25	0.307	0.079	0.406	0.368
213	PO(OMe) ₃ ^{a,c}	30	0.260	0.045	0.336	0.330
214	PO(OMe) ₃ ^{a,b}	35	0.220	0.011	0.263	0.290
215	PO(OMe) ₃ ^{a,b}	40	0.182	-0.031	0.174	0.241
216	PO(OPh) ₃ ^{a,b}	50	1.002	0.942	0.950	0.949
217	PO(OPh) ₃ ^d	60	0.866	0.881	0.873	0.826
218	PO(OPh) ₃ ^{a,b}	70	0.761	0.819	0.795	0.708
219	PO(OPh) ₃ ^{a,b}	80	0.664	0.764	0.726	0.608
220	PO(OPh) ₃ ^{a,b}	90	0.587	0.709	0.661	0.517
221	PO(OPh) ₃ ^{a,b}	20	0.452	0.386	0.437	0.486
222	PO(OPr) ₃ ^d	25	0.400	0.345	0.368	0.432
223	PO(OPr) ₃ ^{a,b}	30	0.351	0.310	0.319	0.385
224	PO(OPr) ₃ ^{a,b}	35	0.307	0.276	0.278	0.337
225	PO(OPr) ₃ ^{a,c}	40	0.264	0.235	0.237	0.278
226	POBu(OBu) ₂ ^{a,b}	15	0.640	0.783	0.725	0.733
227	POBu(OBu) ₂ ^{a,b}	25	0.501	0.701	0.599	0.621
228	POBu(OBu) ₂ ^{a,b}	45	0.314	0.557	0.411	0.403
229	POBu(OBu) ₂ ^d	65	0.210	0.433	0.292	0.218
230	POBu(OEt) ₂ ^{a,b}	25	0.297	0.453	0.271	0.339
231	POBu ₂ (OBu) ^{a,b}	25	0.732	0.697	0.682	0.649
232	POBu ₂ (OEt) ^d	25	0.611	0.572	0.562	0.506
233	POBu ₂ (OMe) ^{a,b}	25	0.638	0.527	0.485	0.437
234	POBu ₃ ^d	60	0.665	0.372	0.613	0.604
235	POBu ₃ ^{a,b}	70	0.545	0.310	0.539	0.515
236	POBu ₃ ^{a,b}	80	0.449	0.255	0.411	0.441
237	PO(CCl ₃)(OEt) ₂ ^{a,b}	25	0.796	0.787	0.791	0.772
238	POCl(NMe ₂) ₂ ^{a,c}	18	0.621	0.633	0.471	0.628
239	POCl(NMe ₂) ₂ ^d	25	0.550	0.579	0.433	0.581
240	POCl(NMe ₂) ₂ ^{a,b}	40	0.422	0.469	0.340	0.472
241	POCl(NMe ₂) ₂ ^{a,b}	60	0.283	0.338	0.213	0.328
242	POCl ₂ NMe ₂ ^{a,b}	18	0.344	0.447	0.326	0.395
243	POCl ₂ NMe ₂ ^{a,b}	25	0.303	0.392	0.297	0.352
244	POCl ₂ NMe ₂ ^{a,b}	40	0.185	0.282	0.226	0.255
245	POCl ₂ NMe ₂ ^{a,b}	60	0.086	0.151	0.127	0.129
246	POCl ₂ NMe ₂ ^d	75	0.009	0.062	0.061	0.045
247	POEt(OBu) ₂ ^{a,b}	25	0.504	0.536	0.427	0.477
248	POEt(OEt) ₂ ^{a,b}	15	0.211	0.405	0.181	0.300
249	POEt(OEt) ₂ ^d	25	0.124	0.323	0.086	0.220
250	POEt(OEt) ₂ ^{a,b}	45	-0.014	0.179	-0.034	0.056
251	POEt(OEt) ₂ ^{a,c}	65	-0.129	0.055	-0.084	-0.087
252	POEt ₃ ^{a,b}	50	0.292	0.047	0.337	0.222
253	POEt ₃ ^{a,b}	70	0.137	-0.077	0.202	0.096
254	POEt ₃ ^d	90	0.027	-0.187	0.082	-0.002

#	Compost	T (°C)	Exp.	MLR	BPNN	SVM
255	POHex(OBu) ₂ ^{a,b}	25	0.666	0.825	0.734	0.742
256	POHex(OEt) ₂ ^d	25	0.476	0.591	0.437	0.480
257	POHex ₂ (OEt) ^{a,b}	25	0.879	0.824	0.796	0.769
258	POHex ₃ ^{a,b}	40	1.221	0.885	1.002	1.151
259	POHex ₃ ^d	60	0.883	0.754	0.890	0.914
260	PO(MeCl)(OBu) ₂ ^{a,b}	30	1.204	0.729	0.875	0.811
261	PO(MeCl)(OBu) ₂ ^{a,b}	41	0.667	0.654	0.719	0.701
262	PO(MeCl)(OBu) ₂ ^{a,b}	100	0.255	0.304	0.234	0.185
263	PO(MeCl)(OPr) ₂ ^d	30	1.079	0.598	0.863	0.782
264	PO(MeCl)(OPr) ₂ ^{a,b}	41	0.613	0.523	0.719	0.678
265	PO(MeCl)(OPr) ₂ ^{a,c}	100	0.204	0.173	0.212	0.175
266	POMe(O(3-Me-Bu)) ₂ ^{a,b}	20	0.651	0.687	0.612	0.581
267	POMe(OEt) ₂ ^d	25	0.173	0.292	0.133	0.106
268	POMe(O'Pr) ₂ ^{a,b}	25	0.267	0.318	0.184	0.204
269	POMe(O'Pr) ₂ ^{a,b}	40	0.161	0.208	0.094	0.091
270	POMe(O'Pr) ₂ ^d	60	0.037	0.078	0.032	-0.049
271	POMe ₂ Et ^{a,b}	90	0.041	-0.263	-0.013	-0.029
272	POMe ₂ Et ^d	100	-0.013	-0.311	-0.022	-0.060
273	POMe ₂ Et ^{a,b}	110	-0.056	-0.359	-0.025	-0.087
274	POPh(O(CH ₂ -CH=CH ₂) ₂) ^{a,b}	25	0.777	0.917	0.850	0.922
275	POPh(OBu) ₂ ^{a,b}	25	0.884	0.975	0.822	0.788
276	POPh(OEt) ₂ ^{a,b}	25	0.770	0.438	0.433	0.242
277	POPh(O'Pr) ₂ ^d	30	0.924	0.756	0.907	0.569
278	POPh(OPent) ₂ ^{a,b}	30	1.053	1.073	0.895	0.836
279	POPr ₃ ^{a,b}	50	0.497	0.248	0.502	0.493
280	POPr ₃ ^{a,b}	70	0.301	0.125	0.367	0.341
281	PPh ₃ ^{a,c}	83	0.822	0.621	0.594	0.657
282	PPh ₃ ^{a,b}	90	0.733	0.587	0.576	0.626
283	PPh ₃ ^{a,b}	95	0.672	0.559	0.560	0.602
284	PPh ₃ ^{a,b}	100	0.624	0.532	0.543	0.579
285	PPh ₃ ^d	106	0.565	0.505	0.526	0.558
286	PPh ₃ ^{a,b}	111	0.520	0.477	0.575	0.537
287	PPh ₃ ^{a,b}	116	0.480	0.456	0.494	0.523
288	PPh ₃ ^{a,b}	120	0.439	0.436	0.479	0.509
289	Si(2-CN-Et)(OBu) ₃ ^{a,b}	20	0.486	0.565	0.597	0.424
290	Si(2-Et-Hex) ₃ F ^{a,b}	25	0.833	0.153	0.782	0.740
291	Si(2-Et-Hex) ₃ F ^{a,b}	40	0.684	0.043	0.726	0.696
292	Si(2-OMe-Et) ₃ Cl ^{a,c}	20	0.547	0.083	0.557	0.443
293	Si(2-OMe-Et) ₃ Cl ^{a,b}	25	0.500	0.042	0.519	0.430
294	Si(3Cl-Pr)(OMe) ₃ ^d	20	-0.222	0.108	0.229	-0.222
295	Si(3-Cl-Pr)Cl ₃ ^{a,b}	20	0.266	0.195	-0.059	0.196
296	Si(CH=CH ₂)(OMe) ₃ ^{a,b}	25	0.083	-0.093	0.026	-0.156
297	Si(CH=CH ₂) ₄ ^d	20	-0.330	-0.009	-0.318	-0.099
298	Si(MeCl)(CH ₂ -CH=CH ₂) ₃ ^{a,b}	20	0.274	0.243	0.259	0.265
299	Si(O(1-Me-Bu)) ₄ ^{a,b}	30	0.502	0.501	0.490	0.498
300	Si(O(1-Me-Bu)) ₄ ^{a,b}	38	0.414	0.446	0.421	0.425
301	Si(O(1-Me-Bu)) ₄ ^{a,b}	80	0.199	0.178	0.166	0.129
302	Si(O(1-Me-Bu)) ₄ ^d	99	0.042	0.082	0.100	0.067
303	Si(O(1-Me-Pr)) ₄ ^{a,c}	0	0.541	0.508	0.700	0.629
304	Si(O(1-Me-Pr)) ₄ ^{a,b}	20	0.371	0.336	0.423	0.447
305	Si(O(2-Cl-Et)) ₄ ^{a,b}	20	0.993	0.675	1.049	1.063

#	Compost	T (°C)	Exp.	MLR	BPNN	SVM
306	Si(O(2-Me-Pr)) ₄ ^{a,b}	0	0.570	0.543	0.643	0.500
307	Si(O(2-Me-Pr)) ₄ ^{a,b}	20	0.383	0.371	0.396	0.348
308	Si(O(2-OEt-Et)) ₄ ^d	20	0.920	0.627	0.625	0.512
309	Si(O(2-OEt-Et)) ₄ ^{a,b}	25	0.524	0.586	0.520	0.454
310	Si(O(2-OMe-Et)) ₄ ^{a,b}	21	0.559	0.487	0.564	0.442
311	Si(O(2-OMe-Et)) ₄ ^{a,b}	25	0.459	0.446	0.469	0.391
312	Si(O(3-Me-Bu)) ₄ ^{a,b}	20	0.520	0.679	0.508	0.563
313	Si(O(3-Me-Bu)) ₄ ^{a,b}	30	0.425	0.603	0.559	0.462
314	Si(O(3-Me-Bu)) ₄ ^d	40	0.342	0.528	0.391	0.361
315	Si(O(3-Me-Bu)) ₄ ^{a,c}	50	0.272	0.459	0.323	0.272
316	Si(OBu) ₄ ^{a,b}	20	0.322	0.436	0.285	0.392
317	Si(OBu) ₄ ^{a,b}	30	0.235	0.361	0.197	0.308
318	Si(OBu) ₄ ^d	40	0.163	0.285	0.120	0.223
319	Si(OBu) ₄ ^{a,b}	50	0.102	0.217	0.059	0.149
320	Si(OEt) ₂ (O(2-Me-Ph)) ₂ ^{a,b}	25	0.663	0.645	0.654	0.707
321	Si(OEt) ₂ (O(2-Me-Ph)) ₂ ^{a,b}	38	0.542	0.556	0.540	0.585
322	Si(OEt) ₄ ^{a,b}	10	-0.101	-0.010	-0.121	-0.072
323	Si(OEt) ₄ ^d	15	-0.131	-0.052	-0.151	-0.090
324	Si(OEt) ₄ ^{a,b}	20	-0.159	-0.093	-0.216	-0.109
325	Si(OEt) ₄ ^{a,b}	25	-0.183	-0.134	-0.202	-0.131
326	Si(OEt) ₄ ^{a,b}	30	-0.210	-0.168	-0.220	-0.150
327	Si(OEt) ₄ ^{a,b}	35	-0.236	-0.203	-0.128	-0.169
328	Si(OEt) ₄ ^{a,b}	40	-0.259	-0.244	-0.256	-0.193
329	Si(OEt) ₄ ^{a,c}	45	-0.283	-0.278	-0.255	-0.213
330	Si(O ⁱ Pr) ₂ (O(2-Me-Ph)) ₂ ^{a,b}	25	0.792	0.722	0.761	0.700
331	Si(O ⁱ Pr) ₂ (O(2-Me-Ph)) ₂ ^{a,b}	38	0.646	0.633	0.635	0.576
332	Si(O ⁱ Pr) ₃ (O(2Me-Ph)) ^{a,b}	25	0.477	0.440	0.558	0.422
333	Si(O ⁱ Pr) ₃ (O(2Me-Ph)) ^{a,b}	38	0.360	0.350	0.454	0.321
334	Si(O ⁱ Pr) ₄ ^{a,b}	20	0.173	0.112	-0.018	0.136
335	Si(O ⁱ Pr) ₄ ^{a,c}	30	0.100	0.037	0.052	0.078
336	Si(O ⁱ Pr) ₄ ^{a,b}	40	0.036	-0.039	-0.016	0.018
337	Si(O ⁱ Pr) ₄ ^d	50	-0.023	-0.108	-0.070	-0.036
338	Si(OMe) ₂ (O(2-Me-Ph)) ₂ ^{a,b}	25	0.623	0.601	0.584	0.754
339	Si(OMe) ₂ (O(2-Me-Ph)) ₂ ^{a,b}	38	0.491	0.511	0.479	0.630
340	Si(OMe) ₄ ^{a,b}	20	-0.192	-0.176	-0.258	-0.122
341	Si(OMe) ₄ ^d	25	-0.229	-0.217	-0.274	-0.140
342	Si(OPent) ₄ ^{a,b}	20	0.500	0.698	0.475	0.584
343	Si(OPr) ₄ ^{a,b}	25	0.155	0.116	0.147	0.141
344	Si(OPr) ₄ ^{a,b}	30	0.146	0.082	0.117	0.111
345	Si(OPr) ₄ ^{a,b}	80	-0.081	-0.241	-0.093	-0.157
346	Si(OPr) ₄ ^d	85	-0.143	-0.269	-0.107	-0.173
347	Si(Pent) ₃ F ^{a,b}	25	0.554	-0.005	0.422	0.484
348	Si(CH ₂ -CH=CH ₂)(CH=CH ₂) ₃ ^{a,b}	20	-0.230	0.055	-0.198	-0.022
349	Si(CH ₂ -CH=CH ₂)(OEt) ₃ ^{a,b}	20	-0.117	0.017	-0.052	-0.026
350	Si(CH ₂ -CH=CH ₂)(OEt) ₃ ^{a,b}	25	-0.145	-0.024	-0.090	-0.047
351	Si(CH ₂ -CH=CH ₂)(OEt) ₃ ^d	40	-0.213	-0.134	-0.176	-0.110
352	Si(CH ₂ -CH=CH ₂)(OEt) ₃ ^{a,c}	60	-0.283	-0.265	-0.270	-0.188
353	Si(CH ₂ -CH=CH ₂) ₂ (CH=CH ₂) ₂ ^{a,b}	20	-0.147	0.106	-0.099	0.085
354	Si(CH ₂ -CH=CH ₂) ₂ (OEt) ₂ ^{a,b}	0	0.070	0.242	0.380	0.131
355	Si(CH ₂ -CH=CH ₂) ₂ (OEt) ₂ ^{a,b}	20	-0.059	0.071	0.077	0.066
356	Si(CH ₂ -CH=CH ₂) ₂ (OEt) ₂ ^{a,b}	25	-0.093	0.030	0.017	0.045

#	Compost	T (°C)	Exp.	MLR	BPNN	SVM
357	Si(CH ₂ -CH=CH ₂) ₂ (OEt) ₂ ^d	40	-0.173	-0.080	-0.116	-0.014
358	Si(CH ₂ -CH=CH ₂) ₂ (OEt) ₂ ^{a,b}	60	-0.262	-0.211	-0.234	-0.085
359	Si(CH ₂ -CH=CH ₂) ₃ (4-Cl-Ph) ^{a,b}	20	0.749	0.638	0.825	0.819
360	Si(CH ₂ -CH=CH ₂) ₄ ^{a,b}	20	0.072	0.195	0.118	0.269
361	SiBu ₃ F ^{a,b}	25	0.385	0.157	0.410	0.448
362	SiBu ₃ F ^{a,b}	40	0.281	0.047	0.446	0.351
363	SiEt(OBu) ₃ ^d	20	0.212	0.314	0.350	0.295
364	SiEt(OEt) ₃ ^{a,c}	20	-0.187	-0.075	-0.104	-0.117
365	SiEt(OMe) ₃ ^{a,b}	20	-0.250	-0.153	-0.243	-0.226
366	SiEt(OPr) ₃ ^d	20	0.045	0.125	0.135	0.106
367	SiEt ₂ (OBu) ₂ ^{a,b}	20	0.107	0.130	0.111	0.182
368	SiEt ₂ (OPr) ₂ ^{a,b}	20	-0.013	0.005	-0.058	0.051
369	SiEt ₂ Cl ₂ ^{a,b}	20	-0.151	-0.047	0.039	-0.087
370	SiEt ₃ Bu ^d	0	0.131	0.061	0.117	0.141
371	SiEt ₃ Bu ^{a,c}	20	-0.020	-0.111	-0.004	-0.007
372	SiEt ₃ Hex ^{a,b}	0	0.364	0.192	0.304	0.334
373	SiEt ₃ Hex ^{a,b}	20	0.180	0.020	0.137	0.168
374	SiEt ₃ Hex ^{a,b}	60	-0.090	-0.262	-0.261	-0.066
375	SiEt ₃ Pent ^{a,b}	0	0.241	0.126	0.215	0.242
376	SiEt ₃ Pent ^{a,b}	20	0.074	-0.046	0.069	0.085
377	SiEt ₃ Pent ^d	60	-0.176	-0.328	-0.297	-0.131
378	SiEt ₃ Pr ^{a,b}	0	-0.001	0.000	0.004	0.042
379	SiEt ₃ Pr ^{a,b}	20	-0.132	-0.172	-0.077	-0.103
380	SiEt ₄ ^{a,b}	0	-0.068	-0.073	-0.093	0.003
381	SiEt ₄ ^d	20	-0.188	-0.245	-0.147	-0.148
382	SiEtCl ₃ ^{a,b}	20	-0.218	-0.066	-0.259	-0.130
383	SiH(OEt) ₃ ^{a,b}	15	-0.289	-0.058	-0.188	-0.299
384	SiH ₂ Et ₂ ^{a,b}	25	-0.542	-0.484	-0.555	-0.732
385	SiH(CH ₂ -CH=CH ₂) ₃ ^{a,b}	20	-0.095	-0.012	-0.147	-0.091
386	SiHEt ₂ Cl ^{a,b}	20	-0.310	-0.190	-0.310	-0.424
387	SiHEt ₂ Cl ^d	25	-0.333	-0.231	-0.321	-0.439
388	SiHEt ₃ ^{a,b}	20	-0.385	-0.349	-0.371	-0.452
389	SiHEt ₃ ^{a,c}	25	-0.407	-0.391	-0.440	-0.477
390	SiHEtCl ₂ ^{a,b}	20	-0.324	-0.117	-0.286	-0.300
391	SiHMe ₂ (CH ₂ -CH=CH ₂) ^{a,b}	20	-0.427	-0.350	-0.444	-0.467
392	SiHMe ₂ Cl ^{a,b}	20	-0.533	-0.282	-0.469	-0.583
393	SiHMe(CH ₂ -CH=CH ₂) ₂ ^d	20	-0.282	-0.350	-0.444	-0.467
394	SiHMeCl ₂ ^{a,b}	20	-0.463	-0.125	-0.374	-0.372
395	SiHMeEtCl ^{a,b}	20	-0.412	-0.236	-0.354	-0.506
396	SiHMePh ₂ ^{a,b}	25	0.477	0.384	0.378	0.416
397	SiHPh(CH ₂ -CH=CH ₂) ₂ ^{a,b}	20	0.228	0.302	0.158	0.204
398	SiMe(2-CF ₃ -Et)Cl ₂ ^{a,b}	20	-0.004	0.287	0.005	0.066
399	SiMe(CH=CH ₂)(OMe) ₂ ^{a,c}	20	-0.345	-0.092	-0.224	-0.275
400	SiMe(CH=CH ₂) ₃ ^d	20	-0.398	-0.112	-0.379	-0.233
401	SiMe(CH=CH ₂)(CH ₂ -CH=CH ₂) ₂ ^{a,b}	20	-0.152	0.011	-0.218	-0.017
402	SiMe(MeCl)(CH ₂ -CH=CH ₂) ₂ ^{a,b}	20	0.124	0.088	0.078	0.097
403	SiMe(OEt) ₃ ^d	20	-0.236	-0.075	-0.237	-0.201
404	SiMe(OMe) ₃ ^{a,b}	20	-0.268	-0.185	-0.343	-0.291
405	SiMe ₂ (CH=CH ₂)Cl ^{a,b}	20	-0.434	-0.094	-0.269	-0.304
406	SiMe ₂ (CH=CH ₂)(OMe) ^d	20	-0.463	-0.216	-0.271	-0.347
407	SiMe ₂ (MeCl)(CH ₂ -CH=CH ₂) ^{a,b}	20	-0.043	-0.085	-0.064	-0.062

#	Compost	T (°C)	Exp.	MLR	BPNN	SVM
408	SiMe ₂ (O(1-Me-Hept)) ₂ ^{a,b}	20	0.630	0.523	0.463	0.538
409	SiMe ₂ (OBu) ₂ ^{a,b}	20	0.000	0.033	-0.033	0.014
410	SiMe ₂ (OEt) ₂ ^d	20	-0.317	-0.214	-0.335	-0.307
411	SiMe ₂ (OMe) ₂ ^{a,b}	20	-0.423	-0.290	-0.331	-0.454
412	SiMe ₂ (OPr) ₂ ^{a,c}	20	-0.137	-0.088	0.226	-0.141
413	SiMe ₂ (CH ₂ -CH=CH ₂) ₂ ^{a,b}	20	-0.208	-0.153	-0.231	-0.137
414	SiMe ₂ (CH ₂ -CH=CH ₂)Cl ^{a,b}	20	-0.152	-0.076	-0.154	-0.222
415	SiMe ₂ Br ₂ ^{a,b}	25	-0.054	-0.192	-0.047	-0.124
416	SiMe ₂ Br ₂ ^d	35	-0.099	-0.261	-0.097	-0.132
417	SiMe ₂ Br ₂ ^{a,b}	45	-0.146	-0.336	-0.126	-0.144
418	SiMe ₂ Cl ₂ ^{a,b}	0	-0.201	0.040	-0.069	-0.271
419	SiMe ₂ Cl ₂ ^{a,b}	20	-0.310	-0.132	-0.219	-0.268
420	SiMe ₂ Cl ₂ ^d	40	-0.409	-0.283	-0.254	-0.279
421	SiMe ₂ Ph ₂ ^{a,b}	25	0.553	0.577	0.395	0.672
422	SiMe ₂ Ph ₂ ^{a,b}	35	0.444	0.508	0.344	0.604
423	SiMe ₂ Ph ₂ ^{a,b}	45	0.349	0.432	0.275	0.530
424	SiMe ₂ Ph(CH ₂ -CH=CH ₂) ^d	20	0.143	0.191	0.057	0.169
425	SiMe ₂ PhBr ^{a,b}	25	0.246	0.269	0.386	0.212
426	SiMe ₂ PhBr ^{a,c}	35	0.168	0.200	0.372	0.180
427	SiMe ₂ PhBr ^{a,b}	45	0.099	0.124	0.357	0.140
428	SiMe ₂ Ph(OBu) ^d	25	0.248	0.244	0.328	0.288
429	SiMe ₃ (CH=CH ₂) ^{a,b}	20	-0.399	-0.325	-0.456	-0.422
430	SiMe ₃ (CH ₂ -CH=CH ₂) ^{a,b}	20	-0.394	-0.332	-0.346	-0.289
431	SiMe ₃ Br ^{a,b}	25	-0.252	-0.232	-0.309	-0.210
432	SiMe ₃ Br ^{a,b}	35	-0.301	-0.300	-0.394	-0.231
433	SiMe ₃ Br ^d	45	-0.342	-0.376	-0.415	-0.256
434	SiMe ₃ Bu ^{a,b}	0	-0.182	-0.130	-0.184	-0.131
435	SiMe ₃ Bu ^{a,b}	20	-0.298	-0.302	-0.251	-0.241
436	SiMe ₃ Cl ^d	20	-0.403	-0.251	-0.503	-0.361
437	SiMe ₃ Et ^{a,b}	20	-0.476	-0.440	-0.417	-0.407
438	SiMe ₃ Hex ^d	0	0.074	-0.004	0.018	0.069
439	SiMe ₃ Hex ^{a,b}	20	-0.072	-0.176	-0.094	-0.055
440	SiMe ₃ Hex ^{a,c}	60	-0.299	-0.457	-0.377	-0.232
441	SiMe ₃ (OEt) ^{a,b}	20	-0.440	-0.321	-0.451	-0.370
442	SiMe ₃ (OH) ^{a,b}	20	0.817	0.427	0.851	0.888
443	SiMe ₃ Pent ^{a,b}	0	-0.060	-0.072	-0.076	-0.014
444	SiMe ₃ Pent ^d	20	-0.191	-0.244	-0.169	-0.132
445	SiMe ₃ Ph ^{a,b}	25	-0.063	-0.004	-0.132	-0.076
446	SiMe ₃ Pr ^d	0	-0.320	-0.192	-0.288	-0.221
447	SiMe ₃ Pr ^{a,b}	20	-0.420	-0.364	-0.337	-0.330
448	SiMe ₄ ^{a,b}	20	-0.616	-0.522	-0.507	-0.435
449	SiMe ₄ ^{a,b}	25	-0.622	-0.563	-0.519	-0.465
450	SiMe ₄ ^d	50	-0.674	-0.742	-0.558	-0.582
451	SiMe ₄ ^{a,c}	75	-0.706	-0.893	-0.590	-0.655
452	SiMe ₄ ^{a,b}	100	-0.762	-1.024	-0.622	-0.692
453	SiMe(CH ₂ -CH=CH ₂) ₃ ^{a,b}	20	-0.012	0.025	-0.043	0.092
454	SiMeCl ₃ ^{a,b}	20	-0.312	-0.108	-0.330	-0.382
455	SiMeCl ₃ ^d	30	-0.348	-0.183	-0.395	-0.383
456	SiMeCl ₃ ^{a,b}	40	-0.393	-0.259	-0.437	-0.388
457	SiMeCl ₃ ^{a,b}	50	-0.434	-0.327	-0.455	-0.395
458	SiMeCl ₃ ^{a,b}	60	-0.472	-0.389	-0.455	-0.402

#	Compost	T (°C)	Exp.	MLR	BPNN	SVM
459	SiMeEt ₂ Br ^{a,b}	25	-0.105	-0.140	-0.192	-0.085
460	SiMeEt ₂ Br ^{a,c}	35	-0.159	-0.209	-0.178	-0.119
461	SiMeEt ₂ Br ^d	45	-0.213	-0.284	-0.242	-0.157
462	SiMeEt ₃ ^{a,b}	0	-0.168	-0.133	-0.201	-0.110
463	SiMeEt ₃ ^{a,b}	20	-0.281	-0.305	-0.247	-0.250
464	SiMeEtCl ₂ ^{a,b}	20	-0.194	-0.136	-0.218	-0.264
465	SiMePh(OEt) ₂ ^d	25	0.204	0.165	0.311	0.232
466	SiMePh(OMe) ₂ ^{a,b}	20	0.216	0.143	0.172	0.217
467	SiMePh ₂ (CH ₂ -CH=CH ₂) ^{a,b}	20	0.877	0.621	0.647	0.782
468	SiMePh ₂ Cl ^{a,b}	20	0.846	0.609	0.652	0.776
469	SiMePh ₂ Cl ^{a,b}	25	0.716	0.568	0.642	0.743
470	SiMePh ₂ (OEt) ^{a,b}	25	0.690	0.520	0.810	0.747
471	SiMePh(CH ₂ -CH=CH ₂) ₂ ^{a,b}	20	0.320	0.347	0.345	0.378
472	SiMePhCl ₂ ^d	20	0.209	0.302	0.242	0.245
473	SiOct(OEt) ₃ ^{a,b}	25	0.177	0.263	0.152	0.238
474	Si(OEt)(OPh) ₃ ^{a,b}	21	0.974	0.935	0.955	1.073
475	Si(OEt)(OPh) ₃ ^{a,b}	29	0.868	0.880	0.859	0.984
476	Si(OEt)(OPh) ₃ ^d	35	0.789	0.832	0.778	0.902
477	Si(OEt)(OPh) ₃ ^{a,c}	50	0.628	0.722	0.617	0.711
478	Si(OEt)(OPh) ₃ ^{a,b}	70	0.436	0.598	0.479	0.506
479	Si(OEt)(CH ₂ -CH=CH ₂) ₃ ^{a,b}	20	-0.017	0.171	0.058	0.187
480	SiPh(CH=CH ₂)(CH ₂ -CH=CH ₂) ₂ ^{a,b}	0	0.134	0.621	0.030	0.619
481	SiPh(CH=CH ₂)(CH ₂ -CH=CH ₂) ₂ ^{a,b}	20	0.431	0.450	0.375	0.490
482	SiPh(CH=CH ₂)(CH ₂ -CH=CH ₂) ₂ ^{a,b}	25	-0.051	0.409	0.341	0.456
483	SiPh(CH=CH ₂)(CH ₂ -CH=CH ₂) ₂ ^{a,c}	40	-0.137	0.299	-0.070	0.363
484	SiPh(CH=CH ₂)(CH ₂ -CH=CH ₂) ₂ ^d	60	-0.233	0.168	0.060	-0.258
485	SiPh(OBu) ₃ ^{a,b}	20	0.571	0.619	0.652	0.592
486	SiPh(OEt) ₃ ^{a,b}	20	0.250	0.297	0.469	0.320
487	SiPh(OMe) ₃ ^d	20	0.322	0.278	0.263	0.340
488	SiPh ₂ (OBu) ₂ ^{a,b}	20	1.093	0.918	0.929	0.919
489	SiPh ₂ (OEt) ₂ ^{a,b}	20	0.875	0.688	0.833	0.805
490	SiPh ₂ (OMe) ₂ ^{a,b}	25	0.839	0.565	0.688	0.769
491	SiPh ₂ (OPr) ₂ ^{a,b}	20	1.064	0.800	0.910	0.887
492	SiPh ₂ (CH ₂ -CH=CH ₂) ₂ ^{a,b}	20	1.104	0.792	0.819	0.899
493	SiPh ₂ Cl ₂ ^d	20	0.765	0.690	0.491	0.905
494	SiPh ₃ Cl ^{a,b}	100	0.356	0.514	0.634	0.685
495	SiPh(CH ₂ -CH=CH ₂) ₃ ^{a,b}	20	0.417	0.533	0.473	0.517
496	SiPhCl ₃ ^{a,b}	20	0.158	0.316	0.070	0.228
497	SiPr(OMe) ₃ ^{a,b}	20	-0.166	-0.105	-0.198	-0.205
498	SiPrCl ₃ ^d	20	-0.153	0.000	-0.145	-0.076
499	Sn(O(1,1-Me ₂)Bu) ₄ ^{a,b}	25	0.994	1.018	0.955	0.924
500	Sn(O(1,1-Me ₂)Bu) ₄ ^{a,b}	30	0.924	0.984	0.661	0.872
501	Sn(O(1,1-Me ₂)Bu) ₄ ^d	35	0.855	0.950	0.858	0.820
502	Sn(O(1,1-Me ₂)Bu) ₄ ^{a,b}	40	0.809	0.908	0.800	0.759
503	Sn(O(1,1-Me ₂)Pr) ₄ ^{a,c}	25	1.221	0.712	1.062	1.151
504	Sn(O(1,1-Me ₂)Pr) ₄ ^d	30	1.155	0.678	1.028	1.097
505	Sn(O(1,1-Me ₂)Pr) ₄ ^{a,b}	35	1.081	0.643	0.991	1.043
506	Sn(O(1,1-Me ₂)Pr) ₄ ^{a,b}	40	1.009	0.602	0.944	0.979
507	SnBrBu ₃ ^{a,b}	10	0.773	0.621	0.729	0.759
508	SnBrBu ₃ ^{a,b}	20	0.640	0.538	0.690	0.670
509	SnBrBu ₃ ^{a,b}	25	0.575	0.497	0.662	0.622

#	Compost	T (°C)	Exp.	MLR	BPNN	SVM
510	SnBrBu ₃ ^d	40	0.425	0.387	0.477	0.486
511	SnBrBu ₃ ^{a,b}	50	0.330	0.319	0.474	0.400
512	SnBrBu ₃ ^{a,b}	70	0.179	0.195	0.162	0.249
513	SnBu ₃ (NCO) ^{a,b}	20	1.005	0.850	0.916	0.797
514	SnBu ₃ (NCO) ^d	30	0.876	0.775	0.825	0.700
515	SnBu ₃ (NCO) ^{a,c}	40	0.741	0.699	0.711	0.602
516	SnBu ₃ (NCO) ^{a,b}	50	0.603	0.631	0.592	0.513
517	SnBu ₃ (NCO) ^{a,b}	60	0.461	0.569	0.477	0.437
518	SnBu ₃ (NCO) ^{a,b}	70	0.297	0.507	0.362	0.366
519	SnBu ₃ (NCO) ^{a,b}	80	0.104	0.452	0.261	0.307
520	SnBu ₄ ^{a,b}	20	0.417	0.355	0.496	0.461
521	SnBu ₄ ^{a,b}	25	0.362	0.314	0.440	0.412
522	SnClBu ₃ ^{a,b}	10	0.803	0.529	0.633	0.541
523	SnClBu ₃ ^{a,b}	25	0.660	0.405	0.544	0.412
524	SnClBu ₃ ^{a,b}	40	0.417	0.295	0.411	0.292
525	SnClBu ₃ ^d	50	0.386	0.227	0.307	0.219
526	SnClBu ₃ ^{a,c}	70	0.220	0.103	0.102	0.097
527	SnClPh ₃ ^{a,b}	113	0.425	0.848	0.407	0.495
528	SnEt ₄ ^{a,b}	20	-0.137	-0.158	-0.147	-0.181
529	SnEt ₄ ^d	25	-0.167	-0.199	-0.090	-0.213
530	SnMe ₄ ^d	15	-0.342	-0.364	-0.358	-0.301
531	SnMe ₄ ^{a,b}	20	-0.367	-0.405	-0.391	-0.330
532	SnMe ₄ ^{a,b}	25	-0.387	-0.447	-0.419	-0.359
533	SnMe ₄ ^{a,b}	30	-0.407	-0.481	-0.384	-0.382
534	SnPr ₄ ^{a,b}	25	0.097	0.068	0.196	0.074

^a Training set de MLR i SVM. ^b Training set de BPNN.

^c Cross-validation set de BPNN. ^d Prediction set.

3.2. Tensió superficial de compostos organometàl·lics

La tensió superficial dels líquids, a l'igual que la viscositat, és una propietat superficial i de transport de gran importància en molts estudis científics i en aplicacions d'enginyeria. La magnitud d'aquesta propietat se sol relacionar amb les interaccions intermoleculares^{50,51} que es produeixen en el si del líquid de caràcter polar, electrostàtic i d'enllaç d'hidrogen, i també amb el pes molecular.⁵² S'anomena tensió superficial al fenomen físic pel qual la superfície dels líquids tendeix a comportar-se com si fos una fina pel·lícula elàstica; per definició, la tensió superficial és la quantitat d'energia lliure de Gibbs, a temperatura, pressió i composició constants, que cal aplicar per produir un canvi en la superfície d'un sistema qualsevol, en aquest cas es tracta de la interfície entre els estats líquid i gasós d'un compost.

Aquesta propietat presenta un comportament invers respecte la temperatura, és a dir, la tensió superficial, σ , dels compostos en estat líquid disminueix en augmentar la temperatura. Aquest comportament es troba recollit en la relació d'Eötvös, que utilitza la següent equació:

$$\sigma V^{\frac{2}{3}} = k(T_{crit} - T_E - T)$$

on V és el volum molar del líquid, k és una constant empírica, T_{crit} és la temperatura crítica on la tensió entre les fases líquida i gasosa és zero i T_E és la temperatura específica, característica del compost.

La determinació experimental d'aquesta propietat es pot realitzar de diverses maneres.⁵³ El mètode més habitual és el de l'ascens o descens per capil·laritat, aquest mètode és completament estàtic i permet fer mesures en un rang de temperatures molt ampli. Altres mètodes utilitzats són el de la bombolla de gas, el del pes o volum de les gotes, el del despreniment d'una anella o planxa, el de la pressió màxima en una gota i el de dispersió superficial de radiació làser. Els valors d'aquesta propietat es troben recollits en moltes bases de dades,^{6,54-59} demostrant així la importància que té. Tot i que amb els mètodes experimentals enunciats es poden obtenir valors molt acurats per a la majoria de compostos químics, aquesta determinació és, sovint, complicada i lenta; per

aquesta raó s'han anat desenvolupant alguns mètodes que en permeten l'estimació, com per exemple el mètode de Mcleod i Sudgen, que utilitza la següent relació empírica:

$$\sigma^{1/4} = P(\rho_L - \rho_V) / M$$

on ρ_L i ρ_V són les densitats del compost estudiat en estat líquid i vapor, respectivament, M és el pes molecular del compost i P és el paràmetre anomenat *paracor*. Aquest paràmetre, que es considera independent de la temperatura, es pot definir com una relació entre el volum dels dos compostos en contacte on es genera la tensió superficial.

Com ja s'ha comentat, és conegut que la tensió superficial depèn de diferents factors estructurals de les molècules, així doncs, no és sorprenent que la metodologia QSPR hagi jugat un paper important dins d'aquests mètodes d'estimació. S'han publicat diversos treballs, tots ells per a conjunts de compostos orgànics, que correlacionen el valor de la tensió superficial amb diferents paràmetres derivats de l'estructura molecular. Els primers estudis, desenvolupats en el grup de Jurs,^{60,61} estableixen correlacions lineals que permeten el càlcul de la tensió superficial de diferents famílies de compostos orgànics: alcans, èsters i alcohols. Més tard, aquest mateix grup va desenvolupar models no lineals que permeten calcular-la per a un conjunt de compostos orgànics bastant més ampli.²⁰ Altres estudis se centren en el càlcul de la tensió superficial d'un conjunt d'alcans amb descriptors topològics i electrotopològics.⁶²⁻⁶⁴ Molt recentment s'ha publicat un estudi que utilitza models lineals i no lineals, desenvolupats amb SVM, que permeten estimar el valor de tensió superficial d'un conjunt de 196 compostos orgànics de naturalesa molt variada.⁶⁵

Com en el cas de la viscositat, tots els treballs realitzats fins a aquest moment són per a conjunts de compostos orgànics degut, probablement, a les dues mateixes raons: la major facilitat de treballar amb compostos d'aquest tipus i la menor fiabilitat dels valors determinats experimentalment per a compostos organometàl·lics. Així doncs, en aquest capítol s'aplicarà la metodologia QSPR per establir models lineals i no lineals que permetin l'estimació de la tensió superficial en fase líquida per a un conjunt de compostos organometàl·lics. El tractament que es fa en aquest cas és idèntic al que s'ha descrit amb la viscositat: en primer lloc es deduiran models lineals i no lineals que permetin l'estimació de la tensió superficial al voltant de 20°C; en segon lloc, i donat

que la recerca bibliogràfica de valors experimentals així ho permet, s'establiran els models lineals i no lineals que permetin calcular aquesta propietat a temperatura variable. Aquesta segona part de l'estudi permetrà comprovar si la metodologia QSPR és capaç d'incorporar l'efecte de la temperatura en el valor de la tensió superficial.

En la pràctica, en aquest capítol es presenten els resultats obtinguts d'aplicar la metodologia QSPR a dos conjunts de compostos organometàl·lics, el primer consta de 131 compostos a temperatures al voltant dels 20°C; el segon conjunt conté 468 valors de tensió superficial, que s'obtenen de la combinació dels 131 compostos del primer conjunt a diverses temperatures. Els models que s'establiran per a tots dos conjunts de dades seran del mateix tipus dels ja vistos en el càlcul de la viscositat. Els models lineals es desenvoluparan mitjançant regressió multilíneal mentre que els no lineals s'establiran a través de la utilització de BPNN i SVM. De nou, s'espera que els models trobats continguin descriptors que, a més de permetre el càlcul de la tensió superficial, aportin la informació fisicoquímica necessària per interpretar aquesta propietat, és a dir, que els descriptors triats recullin informació relacionada amb les possibles interaccions intermoleculares que es donen en el líquid.

3.2.1. Tensió superficial a 20°C

Conjunt de dades

En total es disposa de 131 valors, trobats en la compilació Landolt-Börnstein,⁵⁹ de tensió superficial de compostos organometàl·lics en fase líquida. Aquest conjunt de valors conté 5 compostos d'arsènic, 15 de germani, 62 de fòsfor, 33 de silici i 16 d'estany. Com en els casos anteriors, s'han mantingut els compostos de fòsfor dintre de la base de dades perquè la seva estructura i comportament són semblants a la de la resta de compostos estudiats, tot i no ser, de fet, compostos organometàl·lics. La fórmula general de tots els compostos és MR_nX_m , on R són grups alquil, alcoxi, aril, ariloxi, amino i alquenil, per la seva part, el grup X pot ser F, Cl, Br, OH i NCO. Les unitats més habituals en què es dona el valor de tensió superficial són les del sistema internacional: J/m^2 o N/m , encara que de vegades, i per conveniència, s'utilitza el submúltiple mN/m que coincideix amb dyn/cm i erg/cm^2 , unitats molt usades en la bibliografia més antiga. En aquest treball tots els valors de tensió superficial es donen en mN/m . La base de dades construïda conté 131 valors de tensió superficial, σ , que abasten un rang que va des de 12.85 mN/m per al $SiMe_4$ fins a 44.32 mN/m per al $PO(OMe)(OPh)_2$, amb un valor mitjà de 27.18 mN/m . El rang de temperatures en el qual s'han realitzat aquestes determinacions experimentals va des de 15 a 30°C; tot i així, la majoria dels valors, 111, han estat mesurats en l'interval 20-25°C. Els valors experimentals i calculats es troben en la Taula 3.19 on es dona també la temperatura exacta de la determinació experimental per a cada compost.

Els *training* i *prediction sets* provinents de la primera divisió del conjunt de dades s'empraran en l'establiment de models lineals per regressió i models no lineals amb SVM. Aquests dos subconjunts contenen 105 i 26 valors experimentals, respectivament. El *cross-validation set*, necessari per establir els models no lineals amb les xarxes neuronals d'ADAPT, conté 11 valors experimentals, deixant el *tset* per a aquest tipus de models amb 94 valors experimentals.

El mètode escollit per optimitzar la geometria d'aquests compostos ha estat el PM3 ja que alguns d'ells presenten àtoms no parametritzats pel mètode AM1.

Model lineal

El mètode triat per establir els models lineals és l'heurístic de CODESSA, amb aquest mètode es redueix el nombre inicial de descriptors calculats des de 885 a 152.

El model de set descriptors obtingut amb el mètode heurístic amb els 105 compostos que formen el *tset* es mostra en la Taula 3.20. La correlació entre els set descriptors es troba en el rang 0.03-0.75, amb un valor mitjà de 0.23. El model presenta un valor VIF mitjà de 9.13, que indica una col·linealitat acceptable entre descriptors. Els descriptors més importants dintre la correlació, segons els valors de β , són l'energia molecular d'intercanvi/nombre d'àtoms, l'energia molecular de ressonància/nombre d'àtoms, FNSA-3 i FPSA-3.

Taula 3.20. Model obtingut per regressió lineal.

Descriptor	coef.	sd	β	test-t	VIF
Terme independent	-19.75	11.48	-	-1.72	-
Àrea superficial negativa fraccionada, FNSA-3	-116.41	6.85	-1.78	-16.98	7.54
Àrea superficial positiva fraccionada, FPSA-3	974.16	73.05	1.38	13.34	7.36
Àrea superficial relativa carregada positivament, RPCS	-0.94	0.12	-0.48	-7.68	2.74
Àrea carregada donadora d'hidrogen, HDCA	-0.74	0.12	-0.30	-6.35	1.48
Repulsió electró-electró màxima d'un enllaç C-H	-1.01	0.29	-0.15	-3.46	1.22
Energia molecular d'intercanvi/nombre d'àtoms	-31.94	2.32	-2.48	-13.75	21.12
Energia molecular de ressonància/nombre d'àtoms	9.80	0.94	1.82	10.40	22.46

La Taula 3.21 mostra els resultats que s'obtenen per als dos subconjunts de dades estudiats, *tset* i *pset*. El valor de R^2_{cv} obtingut és 0.822, indicatiu de l'estabilitat del model construït. Els valors de tensió superficial calculats amb aquest model es troben en la Taula 3.18 i la seva representació gràfica enfront dels valors experimentals es pot veure en la Figura 3.9.

Taula 3.21. Paràmetres estadístics obtinguts amb el model lineal.

Conjunt	n	R^2	F	rmse
<i>tset</i>	105	0.859	628.9	1.64
<i>pset</i>	26	0.858	144.7	2.05

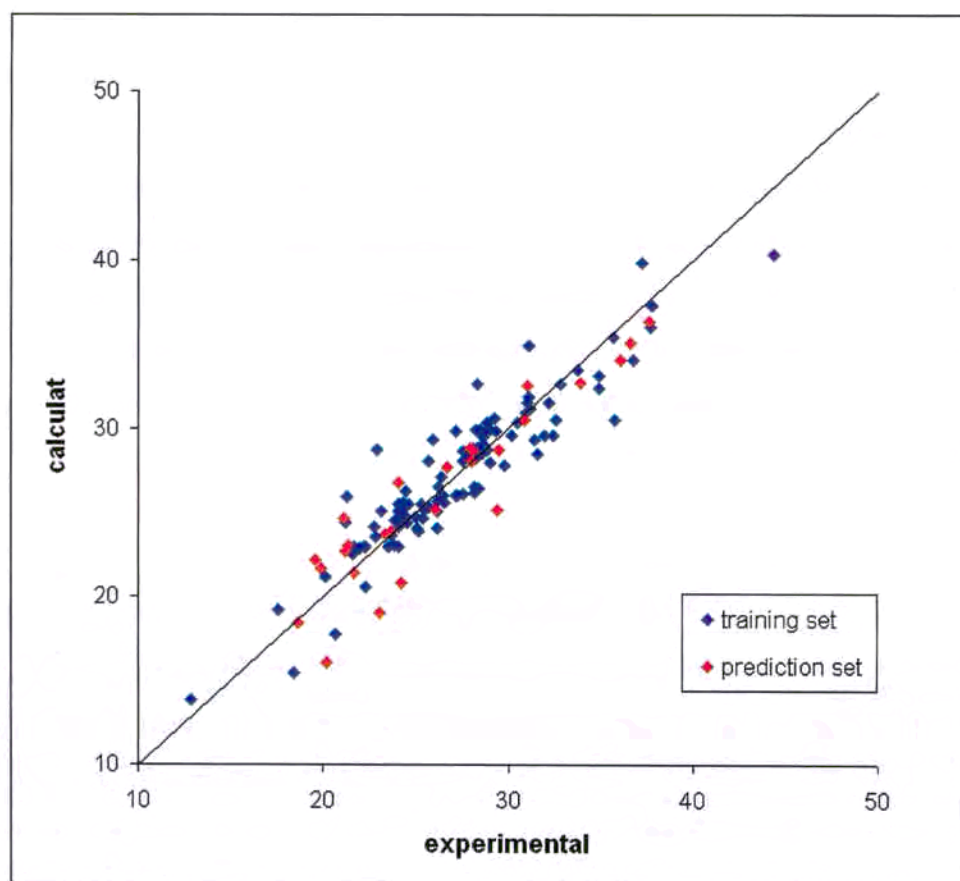


Figura 3.9. Representació gràfica de σ calculada vs. experimental amb el model lineal.

El model construït conté quatre descriptors electrostàtics de tipus CPSA: FNSA-3, FPSA-3, RPCS i HDCA; i tres quàntics relacionats amb característiques energètiques de les molècules. Els descriptors FNSA-3 i FPSA-3, de tipus CPSA, es calculen de la següent manera:

$$FXSA - 3 = \frac{\sum_A q_A \cdot S_A}{TMSA}$$

on q_A són les càrregues atòmiques parcials negatives i positives, respectivament; S_A és la superfície total accessible al solvent carregada amb el signe corresponent i $TMSA$ és la superfície total de la molècula. Les càrregues parcials per computar FNSA-3 s'obtenen directament del càlcul de MOPAC mentre que les que s'utilitzen per calcular FPSA-3 es troben amb el mètode de Zefirov.⁴¹

El descriptor RPCS, que s'obté també a partir de la distribució de càrregues calculades pel mètode de Zefirov,⁴¹ es calcula segons:

$$RPCS = \frac{S_{max}}{\sum_A S_A}$$

on S_{max} és l'àrea superficial atòmica màxima amb càrrega positiva en la molècula, i S_A és l'àrea superficial atòmica amb càrrega positiva de cada àtom. El descriptor HDCA és una mesura del poder donador d'enllaç d'hidrogen que presenta la molècula, aquest descriptor s'obté:

$$HDCA = \sum_D S_D$$

on S_D és l'àrea superficial donadora d'enllaç d'hidrogen. Les càrregues necessàries per poder calcular aquest descriptor es dedueixen també mitjançant el mètode de Zefirov.⁴¹ Els dos descriptors quàntics d'energia molecular d'intercanvi i ressonància es calculen com la suma de totes les interaccions diatòmiques d'aquest tipus dintre de la molècula.

Com es pot comprovar, pràcticament tots els descriptors continguts en aquest model es poden relacionar directament amb les possibles interaccions intermoleculares que es donen en els compostos en estat líquid. Els descriptors CPSA són clarament indicadors d'aquestes interaccions a través de fenòmens electrostàtics: FNSA-3, FPSA-3 i RPCS; o d'enllaç d'hidrogen: HDCA.

Els resultats obtinguts amb aquest model lineal són prou bons i, tal i com s'esperava, els descriptors inclosos contenen informació sobre les interaccions intermoleculares que es donen en els líquids que, alhora, determinen el valor de tensió superficial de cada compost. De manera semblant a la viscositat, aquest estudi és el primer que es fa sobre aquesta propietat amb aquest tipus de compostos i, per tant, no es disposa d'altres resultats per poder establir comparacions directes. La comparació només és possible amb altres estudis on es calcula la tensió superficial de compostos orgànics, tal i com es mostra al final d'aquest apartat.

Model no lineal: Xarxes neuronals BPNN

Com ja s'ha indicat, aquest tipus de models es dedueixen amb el programa ADAPT. Les rutines de selecció objectiva de descriptors implementades en aquest programa redueixen el nombre de descriptors inicials de 152 a 100. L'arquitectura idònia de la xarxa neuronal és 7-5-1, la qual cosa suposa un nombre de paràmetres ajustables de 46 i un *ratio* del nombre de compostos del *tset*, 94, sobre aquest nombre de 2.04, just per sobre del 2.0 recomanat.

El model no lineal desenvolupat amb les xarxes neuronals BPNN conté els descriptors que es mostren en la Taula 3.22, dels quals, tres són descriptors electrostàtics de tipus CPSA: FNSA-3, WNSA-2 i RPCG; i quatre de tipus quàntic. El descriptor FNSA-3 ja ha aparegut en el model lineal. RPCG és anàleg a RPCS, per calcular-lo només cal substituir la superfícies carregades positivament per la càrrega total d'aquestes superfícies, derivada del mètode de Zefirov.⁴¹

Taula 3.22. Descriptors continguts en el model BPNN.

Descriptor	Classe
Àrea superficial negativa fraccionada, FNSA-3	CPSA
Àrea superficial negativa ponderada, WNSA-2	CPSA
Càrrega relativa positiva, RPCG	CPSA
Diferència d'energia HOMO-LUMO	Quàntic
Índex màxim de reacció nucleòfila d'un àtom de carboni	Quàntic
Interacció coulòmbica màxima d'un enllaç C-H	Quàntic
Ordre d'enllaç de l'àtom metàl·lic	Quàntic

El descriptor àrea superficial negativa ponderada WNSA-2 es calcula:

$$WNSA-2 = \frac{\sum_N q_N \sum_N S_N}{1000} \cdot TMSA$$

on q_N és la càrrega atòmica parcial negativa, S_N és la superfície total accessible al solvent carregada negativament i $TMSA$ és la superfície total de la molècula. La diferència d'energies entre els orbitals moleculars HOMO i LUMO és considerada, sovint, com a una mesura de la polaritzabilitat de les molècules. D'aquesta manera, aquest descriptor recolliria informació relacionada amb interaccions polars entre les

molècules del líquid. Ja que els descriptors CPSA es relacionen amb les interaccions intermoleculars, el model proposat s'adiu, una vegada més, a les causes que determinen el valor de la propietat.

Els resultats d'aquest model per als tres subconjunts de compostos estudiats: *tset*, *cvset* i *pset*, són bastant millors que els obtinguts amb el lineal. Els coeficients estadístics per a aquests subconjunts de compostos es troben en la Taula 3.23, els valors de tensió superficial calculats amb aquest model es troben en la Taula 3.19 i la seva representació gràfica enfront dels valors experimentals es troba en la Figura 3.10.

Taula 3.23. Paràmetres estadístics obtinguts amb el model BPNN.

Conjunt	n	R ²	F	rmse
<i>tset</i>	94	0.968	2766.8	0.81
<i>cvset</i>	11	0.981	472.8	0.80
<i>pset</i>	26	0.957	530.7	1.17

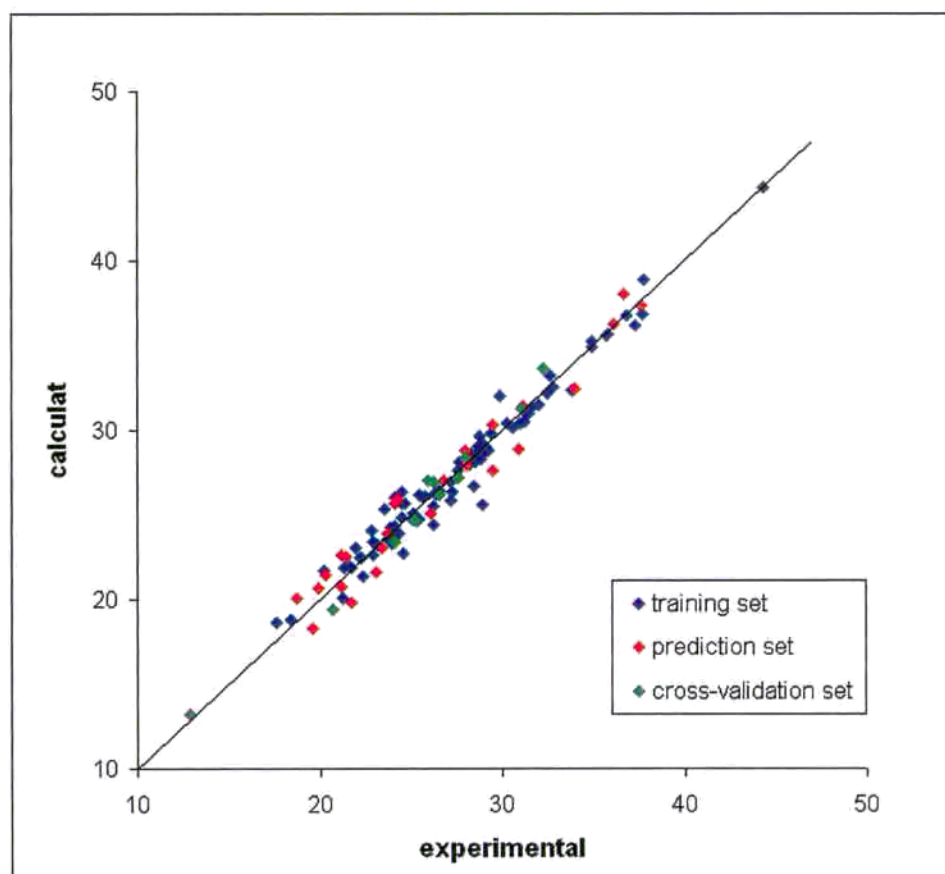


Figura 3.10. Representació gràfica de σ calculada vs. experimental amb el model BPNN.

El valor de *rmse* obtingut amb el conjunt complet de compostos és de 0.89. Hi ha vuit compostos que presenten residuals dues vegades més grans que aquest valor i només un que el té tres cops més gran, el SnMe₃Ph, Taula 3.24. D'aquests vuit compostos, n'hi ha un d'arsènic, un de germani, un de fòsfor, quatre de silici i un d'estany, amb residuals positius i negatius independentment del metall, la qual cosa fa pensar que no hi ha cap tendència estructural que expliqui aquests errors.

 Taula 3.24. Prediccions amb valors més grans que dues vegades el *rmse**.

Compost	T (°C)	Exp.	BPNN	residual
Si(OEt) ₃ (3-OPh-Pr)	20	29.83	32.00	-2.17
GePr ₄	24	24.08	25.98	-1.90
Sn ⁿ Bu ₄	25	24.50	26.34	-1.84
SiMeEt ₂ Br	25	23.55	25.34	-1.79
As(OPr) ₃	27	29.41	27.62	1.79
SiMe ₂ Br ₂	25	24.57	22.71	1.86
Si(OEt) ₄	20	21.67	19.79	1.88
PHO(OEt) ₂	20	30.87	28.93	1.94
SnMe ₃ Ph	25	28.90	25.61	3.29

* Els compostos ombrejats tenen residuals més grans que tres vegades el *rmse*

Model no lineal: Support Vector Machines

El model trobat amb el mètode de selecció heurístic-SVM, a partir dels 152 descriptors inicials triats pel mètode lineal, conté els set descriptors que es troben en la Taula 3.25.

Taula 3.25. Descriptors continguts en el model SVM.

Descriptor	Classe
Càrrega relativa positiva, RPCG	CPSA
Càrrega negativa superficial total, PNSA-2	CPSA
Diferència entre superfícies carregades, DPSA-3	CPSA
Energia LUMO	Quàntic
Diferència d'energia HOMO-LUMO	Quàntic
València mitjana d'un àtom de carboni	Quàntic
Energia de ressonància màxima d'un enllaç C-H	Quàntic

En aquest cas, els tipus de descriptors coincideixen amb els trobats amb BPNN. El model SVM conté tres descriptors de tipus CPSA: DPSA-3, PNSA-2 i RPCG; i quatre descriptors quàntics. Els descriptors RPCG, amb càrregues parcials calculades pel mètode de Zefirov,⁴¹ i diferència d'energies entre els orbitals moleculars HOMO i LUMO ja han aparegut en els models anteriors. La resta de descriptors són, també, molt semblants als que ja han aparegut en la resta de models. El descriptor DPSA-3 és la diferència entre les àrees carregades positiva i negativament. Aquestes àrees carregades es calculen:

$$PPSA-3 = \sum_P q_P \cdot S_P \quad PNSA-3 = \sum_N q_N \cdot S_N$$

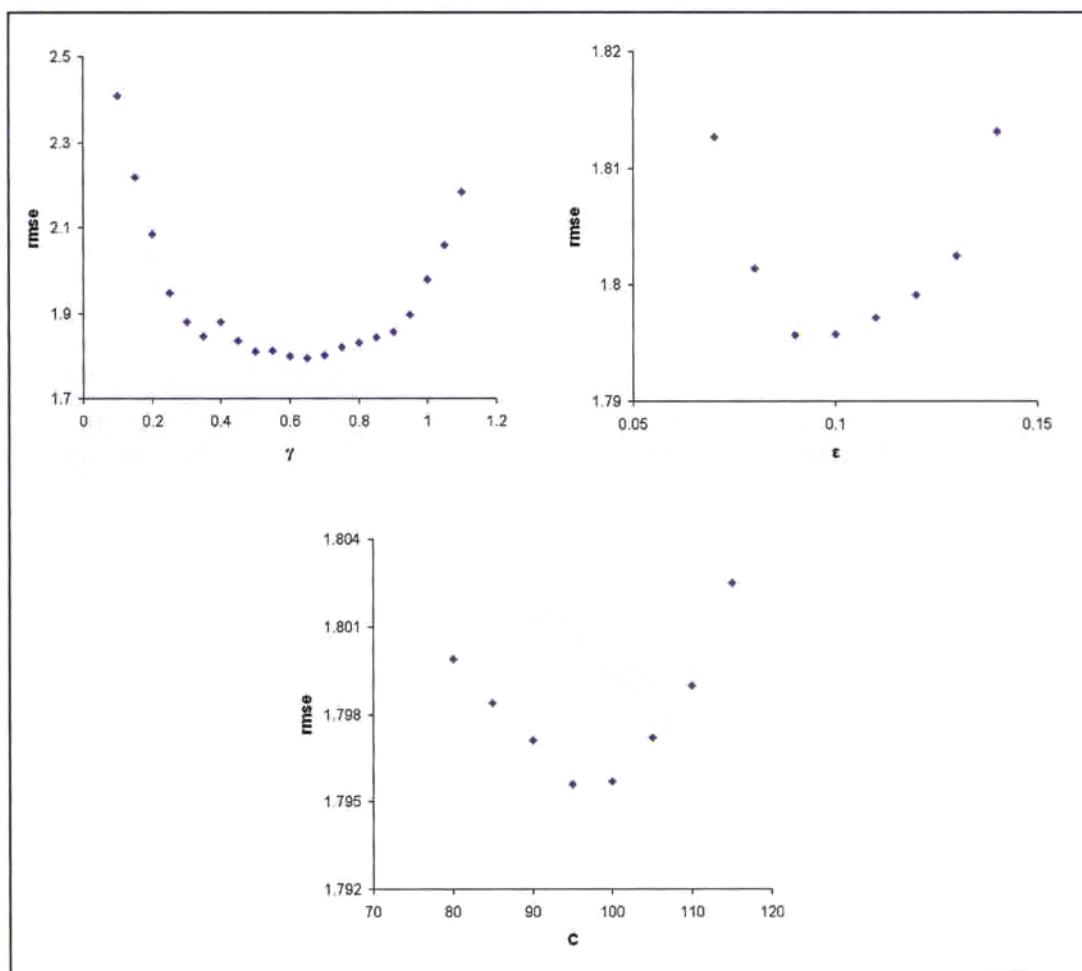
on q són les càrregues atòmiques parcials, S és la superfície total accessible al solvent carregada, P i N indiquen càrregues parcials positives i negatives, respectivament. El descriptor PNSA-2 s'obté de la següent manera:

$$PNSA-2 = \sum_N q_N \sum_N S_N$$

on q_N són les càrregues atòmiques parcials negatives, S_N són les superfícies atòmiques accessibles al solvent carregades negativament.

L'energia de l'orbital LUMO se sol associar a les característiques àcid/base d'un compost ja que és a través d'aquest orbital que es produeixen els enllaços donador/acceptor de Lewis. Els altres dos descriptors de tipus energètic, són semblants als ja vistos i es poden relacionar amb la capacitat de la molècula d'interaccionar a través de fenòmens electrostàtics. També en aquesta ocasió, els descriptors inclosos en el model permeten interpretar la propietat estudiada en funció de les interaccions intermoleculars que es donen en el compost en estat líquid. Així, els descriptors de tipus CPSA i els relacionats amb les energies dels orbitals molecular HOMO i LUMO contenen informació sobre les interaccions intermoleculars que s'estableixen per fenòmens electrostàtics i polars.

Per obtenir el model definitiu que permet el càlcul de la tensió superficial cal optimitzar els paràmetres característics γ , ϵ i C amb el mètode de minimització de l'error de validació creuada, Figura 3.11. L'optimització del paràmetre γ (amb $\epsilon=0.1$ i $C=100$) troba que el valor òptim per aquest paràmetre és 0.65. Un cop optimitzat aquest paràmetre, es passa a optimitzar ϵ (amb $\gamma=0.65$ i $C=100$), l'optimització d'aquest paràmetre fixa el seu valor òptim en 0.09. El paràmetre C s'optimitza amb $\gamma=0.65$ i $\epsilon=0.09$; el millor valor per aquest paràmetre és 95.


 Figura 3.11. Optimització dels paràmetres γ , ϵ i C .

El model construït consta de 98 vectors suport, és a dir, s'utilitza aquest nombre de compostos d'un total de 105 del *tset* per establir-lo. Els resultats estadístics que s'obtenen són força bons i són un punt intermediari entre el model lineal i el BPNN, la Taula 3.26 conté els valors del coeficient de determinació R^2 , el test- F i *rmse* per als dos subconjunts de dades analitzats: *tset* i *pset*.

Taula 3.26. Paràmetres estadístics obtinguts amb el model SVM.

Conjunt	n	R^2	F	<i>rmse</i>
<i>tset</i>	105	0.918	1137.2	1.22
<i>pset</i>	26	0.931	324.9	1.25

Els valors de tensió superficial calculats amb aquest model es troben en la Taula 3.19 i la seva representació gràfica enfront dels valors experimentals es pot veure en la Figura 3.12.

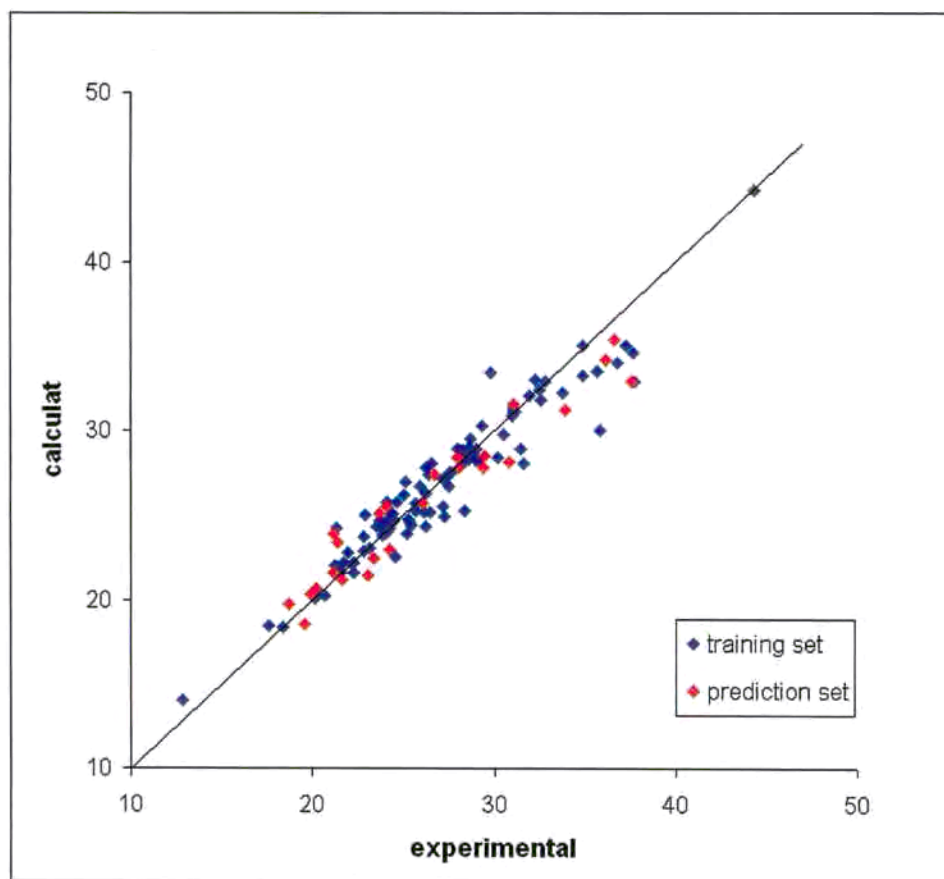


Figura 3.12. Representació gràfica de σ calculada vs. experimental amb el model SVM.

Aquesta representació gràfica permet veure amb claredat que alguns compostos no es troben ben descrits pel model, la majoria d'aquests compostos presenten valors de tensió superficial en l'interval 30-40 mN/m. La Taula 3.27 mostra els compostos amb residuals dues vegades més grans que el *rmse* del model (1.23). Com es pot comprovar, la majoria de compostos que hi apareixen es troben en l'esmentat interval i n'hi ha tres que es poden considerar *outliers* ja que el seu residual és tres vegades superior al *rmse*: $\text{PHO}(\text{OMe})_2$, $\text{PO}(\text{OMe})_3$ i $\text{SnBu}_3(\text{NCO})$.

Taula 3.27. Prediccions amb valors més grans que dues vegades el *rmse**

Compost	T (°C)	Exp.	BPNN	residual
Si(OEt) ₃ (3-OPh-Pr)	20	29.83	33.43	-3.60
SnMe ₃ (-CH=CH ₂)	25	21.30	24.23	-2.93
Ge(O ⁱ Pr) ₄	20	21.17	23.90	-2.73
As(OMe) ₃	20	31.44	28.88	2.55
P(OEtOMe) ₃	20	33.92	31.23	2.69
PHO(OEt) ₂	20	30.87	28.14	2.73
PO(OMe) ₂ Et	20	36.77	33.99	2.78
PHO(OEtOMe) ₂	20	37.68	34.60	3.08
Ge(Hex) ₄	22	28.39	25.21	3.18
SnBu ₃ Cl	20	31.60	28.08	3.52
PHO(OMe) ₂	20	37.60	32.87	4.73
PO(OMe) ₃	18	37.76	32.85	4.91
SnBu ₃ (NCO)	20	35.80	30.01	5.79

* Els compostos ombrejats tenen residuals més grans que tres vegades el *rmse*

En aquesta taula hi ha alguns compostos que també es troben en la Taula 3.24, corresponent al model BPNN: Si(OEt)₃(3-OPh-Pr) i PHO(OEt)₂, en tots dos casos els residuals són del mateix signe.

Comparació de resultats

Com ja s'ha comentat anteriorment, aquest és el primer estudi que es fa mitjançant la metodologia QSPR per predir la tensió superficial d'un conjunt de compostos organometàl·lics. Els treballs semblants trobats en la bibliografia només calculen aquesta propietat per conjunts de compostos orgànics. Degut a la major dificultat per determinar experimentalment la propietat estudiada per a compostos organometàl·lics, seria d'esperar que els models aquí desenvolupats donessin resultats pitjors que els que s'obtenen treballant amb compostos purament orgànics, on la tensió superficial és més fàcilment determinable. En la Taula 3.28 es presenten els resultats obtinguts en diferents estudis on s'ha emprat la metodologia QSPR per estimar la tensió superficial de diferents conjunts de compostos orgànics.

Taula 3.28. Estadística dels diferents estudis pel càlcul de la tensió superficial.

Referència	Compostos	Model ^a	Descriptors	R ²	rmse
60	31	MLR	6	0.908	2.32
61	146	MLR	10	0.983	0.40
20	197	MLR	8	0.848	3.03
	197	BPNN	8	0.959	1.63
62	19	MLR	3	0.976	0.29
	19	MLR	4	0.988	0.25
	19	MLR	5	0.992	0.24
63	68	MLR	10	0.976	0.30
64	68	PCR	10	0.794	0.89
65	189	MLR	9	0.744	3.27
	189	SVM	9	0.907	2.16

^a: MLR= regressió multilíneal, PCR= regressió per components principals, BPNN= *Back Propagation Neural Network*, SVM= *Support Vector Machine*

Com es pot comprovar, els resultats trobats en la bibliografia presenten una variabilitat molt elevada, degut probablement, a les grans diferències en la metodologia emprada en cada cas. Els treballs realitzats pel grup de Jurs^{20,60,61} utilitzen conjunts de compostos orgànics molt variats i obtenen resultats del mateix ordre que els aquí obtinguts, tant en els models lineals com en els no lineals. Els descriptors seleccionats en aquestes publicacions són del mateix tipus que els que s'han triat en aquest treball, per exemple, en la referència 20 alguns dels descriptors continguts pel model són: HDSA, PPSA-1, FNSA-3, RNCS i HASA/TMSA.

Altres estudis només utilitzen famílies de compostos molt homogènies, normalment alcans, per deduir models amb descriptors nous com poden ser índex de complexitat,⁶² vectors de distància d'electronegativitat⁶³ o els nous índexs topològics *Sh*.⁶⁴ En general, aquests estudis donen bons resultats però el fet de treballar amb un nombre tan restringit de compostos i, alhora, amb un nombre de descriptors relativament elevat, fan que siguin poc significatius.

El treball més recent⁶⁵ trobat, utilitza la regressió multilíneal i les *Support Vector Machines* per establir models lineals i no lineals per a un conjunt de 189 compostos orgànics. Alguns dels descriptors seleccionats en aquest treball, amb el mètode heurístic de CODESSA, són molt semblants o iguals als que s'han trobat aquí: DPSA-3, FPSA-3, energia de l'orbital LUMO+1 i HDCA-2/TMSA. Tot i treballar amb nou descriptors, els resultats que s'obtenen són una mica pitjors als aquí descrits per als 131 compostos organometàl·lics. En definitiva, pot dir-se que els resultats dels models lineals i no lineals desenvolupats en aquest treball proporcionen resultats molt bons, comparables

fins i tot amb els que s'obtenen en la resta d'estudis on s'han emprat només conjunts de compostos orgànics.

Els models lineal i no lineals desenvolupats en aquest treball contenen descriptors molt semblants, o iguals entre sí. La majoria d'aquests descriptors són de tipus CPSA: FNSA-3, FPSA-2, DPSA-3, RPCG, RPCS, entre d'altres. Aquests descriptors permeten descriure les possibles interaccions intermoleculares que es donen en els líquids i que, a la seva vegada, es poden relacionar amb la propietat estudiada. Els resultats obtinguts amb els tres models són bons, tot i així, els models no lineals superen clarament el model lineal quant a bondat estadística: el model SVM dona errors gairebé un 50% més baixos que el model lineal. Per la seva part, el model BPNN presenta uns resultats encara millors amb uns errors molt més baixos que els trobats amb els mètodes SVM i lineal.

Taula 3.19. Valors experimentals i calculats de σ (mN/m).

#	Compost	T (°C)	Exp.	MLR	BPNN	SVM
1	As(OBu) ₃ ^{a,b}	20	26.53	25.57	26.42	28.03
2	As(O'Pr) ₃ ^{a,c}	27	26.20	25.04	26.95	26.29
3	As(OMe) ₃ ^{a,b}	20	31.44	29.32	31.05	28.88
4	As(OPent) ₃ ^{a,b}	22	26.27	26.47	26.29	27.81
5	As(OPr) ₃ ^d	27	29.41	25.12	27.62	27.80
6	GeHex ₄ ^{a,b}	22	28.39	26.38	26.70	25.21
7	Ge(O(2,2-Me ₂ -Pr)) ₄ ^{a,b}	20	23.86	22.98	24.22	24.58
8	Ge(OBu) ₄ ^{a,b}	20	24.45	24.73	24.81	25.06
9	Ge(OEt) ₄ ^d	20	23.67	23.86	23.88	25.03
10	Ge(O'Bu) ₄ ^{a,b}	20	23.68	23.62	23.54	24.62
11	Ge(O'Pr) ₄ ^d	20	21.17	24.61	22.63	23.90
12	Ge(OMe) ₄ ^{a,b}	20	22.89	28.65	23.42	25.01
13	Ge(OPent) ₄ ^{a,b}	20	24.65	25.41	25.71	25.76
14	Ge(OPr) ₄ ^{a,b}	20	24.06	24.09	24.29	24.68
15	Ge(O ^f Bu) ₄ ^{a,b}	20	23.93	24.40	23.29	24.02
16	Ge(O'Bu) ₄ ^{a,b}	20	23.14	25.02	23.25	23.05
17	GePent ₄ ^{a,b}	22	27.25	25.94	26.34	24.90
18	GeEt ₄ ^{a,c}	20	24.06	22.89	23.40	24.29
19	GeMe ₄ ^d	20	24.25	20.81	25.90	22.92
20	GePr ₄ ^{a,b}	24	24.08	25.03	25.98	23.99
21	P(OBu) ₃ ^d	20	26.07	25.12	25.10	25.62
22	P(OEt) ₃ ^{a,b}	20	24.06	25.49	23.61	25.72
23	P(OEtOBu) ₃ ^{a,b}	20	29.33	29.81	29.86	30.28
24	P(OEtOEt) ₃ ^{a,b}	20	30.54	30.30	30.13	29.78
25	P(OEtOMe) ₃ ^d	20	33.92	32.70	32.45	31.23
26	P(OEtOPr) ₃ ^{a,b}	20	29.04	29.87	29.09	28.21
27	P(OHept) ₃ ^{a,b}	20	28.19	26.14	27.96	28.00
28	P(OHex) ₃ ^{a,c}	20	27.54	26.09	27.18	26.64
29	P(OMe) ₃ ^{a,b}	20	27.18	29.77	26.93	27.09
30	P(OOct) ₃ ^{a,b}	20	28.18	26.44	28.31	28.27
31	P(OPr) ₃ ^{a,b}	20	25.04	24.73	24.77	26.18
32	PHO(OBu) ₂ ^{a,b}	20	28.26	29.89	28.60	28.46
33	PHO(OEt) ₂ ^d	20	30.87	30.46	28.93	28.14
34	PHO(OEtOBu) ₂ ^{a,b}	20	31.10	31.81	31.46	31.36
35	PHO(OEtOEt) ₂ ^{a,b}	20	33.78	33.47	32.38	32.19
36	PHO(OEtOHept) ₂ ^{a,b}	20	30.96	30.92	30.44	30.89
37	PHO(OEtOHex) ₂ ^{a,b}	20	31.17	31.11	30.55	31.08
38	PHO(OEtOMe) ₂ ^{a,b}	20	37.68	36.03	36.88	34.60
39	PHO(OEtOPent) ₂ ^{a,c}	20	31.04	31.48	31.25	30.98
40	PHO(OEtOPr) ₂ ^d	20	31.07	32.47	31.41	31.50
41	PHO(OHept) ₂ ^{a,b}	20	28.67	28.52	28.44	28.76
42	PHO(OHex) ₂ ^{a,b}	20	28.70	29.67	29.63	28.79
43	PHO(O'Pr) ₂ ^{a,b}	20	26.40	27.04	26.17	27.37
44	PHO(OMe) ₂ ^d	20	37.60	36.30	37.36	32.87
45	PHO(OOct) ₂ ^{a,b}	20	29.00	27.92	29.18	28.91
46	PHO(OPent) ₂ ^{a,b}	20	28.46	29.86	28.42	28.37
47	PHO(OPr) ₂ ^{a,b}	20	29.23	30.55	28.81	28.21
48	PO(OBu) ₂ Et ^d	20	28.05	27.97	27.93	27.77
49	PO(OBu) ₃ ^{a,b}	20	27.55	28.00	27.66	27.41
50	PO(OEt) ₂ Et ^{a,b}	20	28.68	29.41	28.75	29.50

#	Compost	T (°C)	Exp.	MLR	BPNN	SVM
51	PO(OEt) ₃ ^{a,b}	18	30.22	29.55	30.44	28.38
52	PO(OEtOMe) ₃ ^d	20	36.61	35.04	38.04	35.34
53	PO(OEtOPr) ₃ ^{a,c}	20	32.24	31.45	33.62	32.94
54	PO(OHept) ₂ Et ^{a,b}	20	28.46	28.70	28.12	28.61
55	PO(OHept) ₃ ^{a,b}	20	28.11	28.63	28.32	28.82
56	PO(OHex) ₂ Et ^{a,b}	20	28.11	28.81	28.12	28.23
57	PO(OHex) ₃ ^{a,b}	20	28.11	28.63	28.18	28.48
58	PO(O'Bu) ₃ ^d	15	26.74	27.69	27.07	27.33
59	PO(O'Pr) ₃ ^{a,b}	15	25.68	28.01	26.00	25.24
60	PO(OMe)(OPh) ₂ ^{a,b}	25	44.32	40.29	44.32	44.23
61	PO(OMe) ₂ Et ^{a,b}	20	36.77	34.02	36.75	33.99
62	PO(OMe) ₃ ^{a,b}	18	37.76	37.24	38.88	32.85
63	PO(OOct) ₂ Et ^{a,b}	20	28.87	28.71	28.52	28.64
64	PO(OPent) ₂ Et ^d	20	27.97	28.79	28.77	28.39
65	PO(OPent) ₃ ^{a,b}	20	27.60	28.62	28.10	27.51
66	PO(OPr) ₂ Bu ^{a,c}	20	27.98	28.27	28.42	28.41
67	PO(OPr) ₂ Et ^{a,b}	20	28.36	28.78	28.12	28.40
68	PO(OPr) ₂ Hept ^{a,b}	20	28.58	28.50	28.36	28.34
69	PO(OPr) ₂ Hex ^d	20	28.07	28.60	28.41	28.26
70	PO(OPr) ₂ Me ^{a,b}	20	28.64	29.00	29.05	28.94
71	PO(OPr) ₂ Oct ^{a,b}	20	28.83	28.66	28.32	28.74
72	PO(OPr) ₂ Pent ^{a,b}	20	28.25	28.15	28.41	28.34
73	PO(OPr) ₂ Pr ^{a,b}	20	27.98	28.16	28.22	28.92
74	PO(OPr) ₃ ^d	17	29.46	28.67	30.32	28.51
75	POCl(OEt)(NMe ₂) ^{a,b}	20	34.90	33.07	35.25	33.20
76	POCl ₂ (OEt) ^{a,b}	20	32.80	32.60	32.53	32.89
77	POCl ₂ (OMe) ^{a,b}	20	34.90	32.32	34.88	34.99
78	POCl ₂ (NEt ₂) ^{a,b}	20	35.70	35.38	35.63	33.48
79	POCl ₂ (NMe ₂) ^d	20	36.10	34.02	36.25	34.14
80	POF(OEt) ₂ ^{a,c}	20	25.90	29.25	27.03	26.65
81	POF ₂ (NMe ₂) ^{a,b}	20	25.10	24.03	25.07	26.91
82	PPh(OEt) ₂ ^{a,b}	20	31.95	29.53	31.54	32.04
83	Si(O(1-Me-Bu)) ₄ ^d	30	23.36	23.65	23.02	22.43
84	Si(O(2-Et-Bu)) ₄ ^{a,c}	30	25.17	23.85	24.68	23.85
85	Si(O(2-Et-Hex)) ₄ ^{a,b}	30	25.38	24.59	26.15	24.59
86	Si(OBu) ₄ ^{a,b}	30	22.85	23.54	22.60	22.81
87	Si(OEt) ₃ (3-OPh-Pr) ^{a,b}	20	29.83	27.73	32.00	33.43
88	Si(OEt) ₄ ^d	20	21.67	21.42	19.79	21.18
89	Si(OHept) ₄ ^{a,b}	30	26.19	25.64	25.51	25.15
90	Si(OHex) ₄ ^{a,b}	30	25.44	25.11	24.75	24.40
91	Si(O'Bu) ₄ ^{a,b}	30	21.96	22.79	23.04	22.80
92	Si(O'Pent) ₄ ^{a,b}	30	22.80	24.09	24.05	23.70
93	Si(O'Pr) ₄ ^d	30	21.18	22.67	20.75	21.54
94	Si(OOct) ₄ ^{a,b}	30	27.15	26.00	25.81	25.52
95	Si(OPent) ₄ ^{a,b}	30	23.89	24.49	23.95	23.78
96	Si(O'fBu) ₄ ^{a,b}	30	22.29	22.87	22.42	22.20
97	SiCl ₂ Ph ₂ ^{a,b}	30	37.24	39.85	36.19	35.06
98	SiEt ₄ ^d	20	23.04	19.02	21.63	21.37
99	SiH(OEt) ₃ ^{a,b}	20	21.25	24.39	20.06	22.01
100	SiHEt ₂ Cl ^{a,b}	20	22.30	20.51	21.39	21.54
101	SiHEt ₃ ^{a,c}	20	20.71	17.70	19.43	20.20

#	Compost	T (°C)	Exp.	MLR	BPNN	SVM
102	SiH ₂ EtCl ₂ ^{a,b}	20	21.70	22.88	21.88	22.16
103	SiHMeCl ₂ ^d	20	18.70	18.38	20.10	19.71
104	SiMe ₂ Br ₂ ^{a,b}	25	24.57	24.38	22.71	22.53
105	SiMe ₂ Cl ₂ ^{a,b}	20	20.16	21.13	21.73	20.07
106	SiMe ₂ Ph ₂ ^{a,b}	25	32.47	29.53	32.22	32.38
107	SiMe ₂ PhBr ^{a,b}	25	28.30	32.61	28.20	28.85
108	SiMe ₃ Br ^d	25	19.87	21.67	20.64	20.29
109	SiMe ₃ Cl ^{a,b}	20	17.58	19.16	18.68	18.39
110	SiMe ₃ (OH) ^{a,b}	20	18.40	15.47	18.77	18.31
111	SiMe ₃ Ph ^{a,b}	25	24.29	25.54	23.89	24.20
112	SiMe ₄ ^{a,c}	20	12.85	13.81	13.19	14.00
113	SiMeCl ₃ ^d	20	20.24	16.03	21.43	20.61
114	SiMeEt ₂ Br ^{a,b}	25	23.55	22.89	25.34	24.29
115	SiPhCl ₃ ^{a,b}	20	31.09	34.86	31.21	31.18
116	SnPent ₄ ^d	25	24.10	26.69	25.64	25.53
117	SnBu ₃ (NCO)	20	35.80	30.45	35.68	30.01
118	SnBu ₃ Cl ^{a,b}	20	31.60	28.42	31.35	28.08
119	SnBu ₄ ^{a,c}	20	26.50	25.99	26.18	25.13
120	SnBuCl ₃ ^{a,b}	20	32.60	30.43	33.19	31.80
121	SnEt ₄ ^{a,b}	20	26.20	24.05	24.42	24.26
122	Sn ⁱ Bu ₄ ^{a,b}	25	24.50	26.21	26.34	24.72
123	Sn ⁱ Pr ₄ ^{a,b}	25	25.70	25.32	26.10	25.64
124	SnMe ₃ (CH=CH ₂) ^{a,b}	25	21.30	25.85	21.83	24.23
125	SnMe ₃ cHex ^{a,b}	25	25.30	25.49	24.66	24.70
126	SnMe ₃ Et ^{a,b}	25	21.60	22.52	21.93	21.55
127	SnMe ₃ ⁱ Pr ^d	25	21.40	23.00	22.53	23.35
128	SnMe ₃ Ph ^{a,b}	25	28.90	30.27	25.61	28.81
129	SnMe ₃ Pr ^{a,b}	25	22.20	22.93	22.47	22.11
130	SnMe ₄ ^d	20	19.56	22.13	18.33	18.53
131	SnPr ₄ ^{a,b}	25	24.50	25.32	25.68	24.63

^a Training set de MLR i SVM. ^b Training set de BPNN.

^c Cross-validation set de BPNN. ^d Prediction set.

3.2.2. Tensió superficial a temperatura variable

Com ja s'ha explicat, la tensió superficial dels compostos químics depèn de manera molt marcada de la temperatura. Aquesta dependència, tal i com es deriva de la relació d'Eötvös, és lineal i d'efecte negatiu, és a dir, la tensió superficial disminueix quan la temperatura augmenta. Aquest comportament és lògic ja que el creixement de la temperatura fa que augmentin les vibracions tèrmiques, que fan alhora que l'establiment d'interaccions intermoleculares en el líquid sigui més difícil.

Conjunt de dades

La base de dades construïda consta de 468 valors experimentals de tensió superficial de 131 compostos organometàl·lics a diferents temperatures, tots aquests valors experimentals han estat extrets de la compilació Landolt-Börnstein.⁵⁹ En aquest cas es tenen 22 valors de compostos d'arsènic, 90 de germani, 229 de fòsfor, 83 de silici i 44 d'estany. El rang que cobreixen aquests valors va des de 12.85 mN/m del SiMe_4 a 20°C fins a 44.32 mN/m del $\text{PO}(\text{OMe})(\text{OPh})_2$ a 25°C, amb un valor mitjà de 25.58 mN/m. El rang de temperatures d'aquests valors va des de 0 a 241°C, amb un valor mitjà de 46.8°C. Tots els valors experimentals i calculats es troben en la Taula 3.29 així com la temperatura exacta de la determinació experimental per a cada compost.

Quan es divideix aquest conjunt de dades en *tset* i *pset* s'obtenen dos subconjunts de 374 i 94 valors, respectivament. Aquests dos nous conjunts de dades s'empraran en la construcció dels models lineals amb regressió multilíneal i no lineals amb SVM. El *cvset*, necessari per establir els models no lineals amb BPNN, ha de contenir un 10% dels compostos del *tset* inicial. En aquest cas, pels models construïts amb BPNN, el *tset* i el *cvset* contenen 335 i 39 valors de σ , respectivament.

Model lineal

Els models lineals s'estableixen mitjançant els procediments de regressió multilíneal implementats en el mètode heurístic de CODESSA. Als 885 descriptors calculats inicialment cal afegir-hi la temperatura en Kelvin, corresponent la determinació de cada valor experimental de tensió superficial. El procediment de reducció de descriptors realitzat per heurístic deixa el nombre inicial d'aquests en 184.

El model lineal obtingut amb el mètode heurístic pels 374 compostos del *tset* es troba en la Taula 3.30.

Taula 3.30. Model obtingut per regressió lineal.

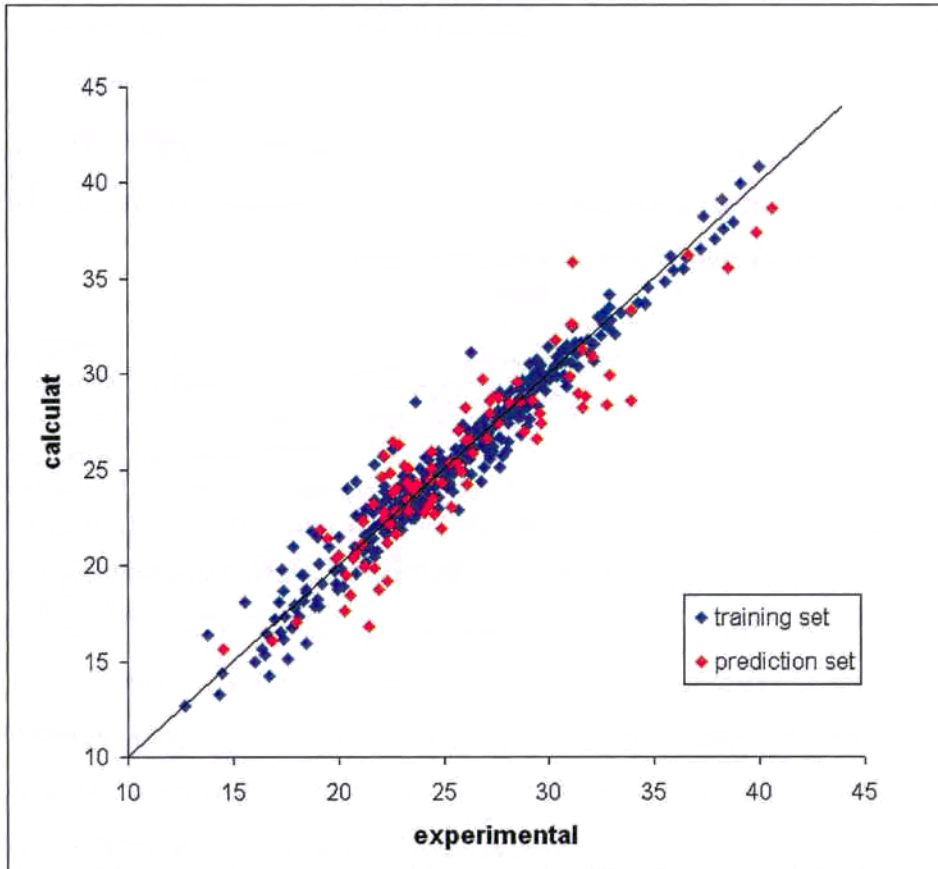
Descriptor	coef.	sd	β	test-t	VIF
Terme independent	-76.18	9.93	-	-7.67	-
Índex de Kier & Hall d'ordre 2	0.379	0.035	0.38	10.84	2.79
Àrea superficial parcial negativa, PNSA-3	-0.119	0.006	-0.56	-21.49	1.59
Paràmetre de polaritat/(distància) ²	-23.15	2.66	-0.21	-8.71	1.34
Interacció coulòmbica màxima d'un enllaç C-H	23.00	3.00	0.20	7.67	1.51
Energia molecular de ressonància/nombre d'àtoms	-3.11	0.13	-0.70	-24.72	1.85
València de l'àtom metàl·lic	-4.06	0.35	-0.31	-11.54	1.65
Temperatura	-0.086	0.004	-0.53	-22.61	1.26

La correlació entre els set descriptors es troba en el rang 0.02-0.56, amb un valor mitjà de 0.24. El model presenta un valor VIF mitjà de 1.71, indicatiu d'una col·linealitat mínima entre els descriptors del model. Els descriptors més importants dintre la correlació, segons els valors de β , són l'energia molecular de ressonància entre el nombre d'àtoms, PNSA-3 i la temperatura.

Els resultats que s'obtenen amb aquest model per als conjunts de dades estudiats es mostren en la Taula 3.31. En aquest cas, el valor de R^2_{cv} és de 0.834, també molt pròxim al valor per al *tset*. Els valors de tensió superficial calculats amb aquest model es troben en la Taula 3.29 i la seva representació gràfica enfront dels valors experimentals es pot veure en la Figura 3.13.

Taula 3.31. Paràmetres estadístics obtinguts amb el model lineal.

Conjunt	n	R ²	F	rmse
<i>tset</i>	374	0.843	1999.5	1.84
<i>pset</i>	94	0.850	521.2	1.84

Figura 3.13. Representació gràfica de σ calculada vs. experimental amb el model lineal.

Com es pot comprovar, els descriptors continguts en aquest model s'assemblen molt als que ja han sortit abans en l'estimació de la tensió superficial a temperatura ambient. Els descriptors PNSA-3, interacció coulòmbica màxima d'un enllaç C-H i l'energia molecular de ressonància entre el nombre d'àtoms ja han aparegut amb anterioritat en aquest capítol i, per tant, no se'ls tornarà a explicar.

El descriptor topològic Índex de Kier & Hall d'ordre 2 conté informació sobre els tipus d'àtoms i enllaços, la seva connectivitat i hibridació en la molècula.⁴³ Aquest descriptor es calcula de la següent manera:

$${}^2\chi = \sum_i^n \sum_{j>1} \sum_{k>j} \left(\frac{1}{\delta_i \delta_j \delta_k} \right)^{\frac{1}{2}}$$

on n és el nombre total d'àtoms de la molècula (sense comptar hidrògens) i δ_i , δ_j i δ_k són la connectivitat atòmica d'ordre zero de cada àtom. El descriptor, de tipus electrostàtic, paràmetre de polaritat entre distància al quadrat s'obté:⁶⁶

$$\frac{PolPar}{d^2} = \frac{q_{max} - q_{min}}{r_{max-min}^2}$$

on q_{max} i q_{min} són les càrregues atòmiques parcials màxima i mínima, respectivament; i $r_{max-min}^2$ és la distància que separa els dos àtoms que posseeixen aquestes càrregues. La valència de l'àtom metàl·lic és un descriptor anàleg a un que ja ha sortit anteriorment: l'ordre d'enllaç d'aquest àtom. Aquest descriptor mesura la força i el tipus d'enllaç que s'estableix entre el metall i els lligands que l'envolten, sent més gran el seu valor més polars són els enllaços M-L que es formen. La informació continguda en els descriptors del model relaciona la propietat estudiada amb les interaccions intermoleculares que es donen en el líquid i la temperatura. Així, les interaccions que es donen per fenòmens de polaritat i electrostàtica venen explicades pels descriptors electrostàtics. Els descriptor topològic índex de Kier & Hall conté informació sobre les interaccions intermoleculares degut a la ramificació i hibridació dels àtoms en els substituents orgànics. Per la seva part, els descriptors energètics donen informació sobre les interaccions intramoleculares que s'estableixen en els compostos estudiats, que possiblement afectin el valor de tensió superficial. La temperatura, que és un dels descriptors més importants del model, presenta signe negatiu, la qual cosa s'adiu perfectament amb el comportament experimental: com més gran sigui la temperatura, més petita es fa la tensió superficial ja que l'augment de temperatura dificulta l'establiment de les interaccions intermoleculares.

Model no lineal: Xarxes neuronals BPNN

Per tal d'establir els models no lineals amb BPNN és necessària una altra reducció en el nombre de descriptors. Així, les rutines de selecció objectiva incorporades per ADAPT redueixen aquest nombre des de 184 fins a 100. L'arquitectura seleccionada en aquesta ocasió és 7-5-1, d'aquesta manera el nombre de paràmetres ajustables és 46; la relació que té amb el nombre de compostos del *tset*, 335, és de 7.28, bastant per sobre del 2.0 recomanat.

Els descriptors continguts en el model desenvolupat amb les xarxes neuronals BPNN d'arquitectura 7-5-1 es troben en la Taula 3.32.

Taula 3.32. Descriptors continguts en el model BPNN.

Descriptor	Classe
Nombre relatiu d'àtoms d'oxigen	Constitucional
Contingut mitjà d'informació d'ordre 0	Topològic
Càrrega relativa positiva, RPCG	CPSA
Diferència entre superfícies carregades, DPSA-3	CPSA
Energia LUMO+1	Quàntic
Ordre d'enllaç màxim d'un àtom de carboni	Quàntic
Temperatura	-

Tal i com es pot veure, el model conté un descriptor constitucional, un de topològic, dos electrostàtics, ambdós de tipus CPSA; dos quàntics i la temperatura. Alguns d'aquests descriptors ja han aparegut en models anteriors i, per tant, no és sorprenent que tornin a ser inclosos en aquest model. Així, els descriptors RPCG, on les càrregues parcials han estat calculades amb el mètode de Zefirov,⁴¹ i DPSA-3 ja han estat seleccionats en models anteriors on ja han estat explicats. Altres descriptors d'aquest model són molt semblant als que ja s'han anat veient al llarg d'aquest capítol; l'energia de l'orbital molecular LUMO+1 sol estar associada a les característiques àcid/base de Lewis de cada compost, tal i com ho està l'energia de l'orbital molecular LUMO. El valor d'aquest descriptor es pot relacionar amb la segona afinitat electrònica de la molècula i, per tant, amb la capacitat acceptora que presenta el compost. El descriptor: contingut mitjà d'informació d'ordre 0, que es deriva de la teoria d'informació de Shannon,^{44,45} conté informació sobre la diversitat atòmica dintre de la molècula i es pot relacionar amb la seva forma i simetria. El descriptor ordre d'enllaç màxim d'un àtom de carboni, derivat de la teoria de Mulliken,^{47,48} presenta uns valors

que depenen fortament de la naturalesa dels lligands orgànics; els valors més alts d'aquest descriptor els tenen compostos amb lligands fenílics mentre que els més baixos corresponen a compostos amb substituents alquílics o alcoxílics. Així doncs, es pot dir que aquest descriptor recull informació sobre les possibles interaccions intermoleculares que es poden donar per efecte de *pi-stacking* dels lligands aromàtics. Els descriptors inclosos en el model contenen informació útil per relacionar la propietat estudiada amb les interaccions intermoleculares de les quals en depèn. Així, els descriptors electrostàtics DPSA-3 i RPCG recullen la informació relacionada amb les interaccions que s'estableixen per efectes de polaritat i de càrrega, el descriptor contingut mitjà d'informació d'ordre 0 permet explicar les interaccions que es donen degut a la ramificació dels àtoms en els substituents orgànics, l'ordre d'enllaç màxim d'un àtom de carboni es relaciona directament amb el nombre d'anells aromàtics presents i mesura les interaccions que es poden formar per *pi-stacking* entre aquest tipus de substituents. Per últim, els descriptors que queden: nombre relatiu d'àtoms d'oxigen i energia de l'orbital molecular LUMO+1, es poden relacionar amb les possibles interaccions acceptor/donador que s'estableixen entre les molècules. L'efecte de la temperatura es troba recollit en el model amb la incorporació d'aquest paràmetre com a descriptor.

Els resultats que s'obtenen amb aquest model per als subconjunts de compostos *tset*, *cvset* i *pset*, són molt bons i són molt semblants entre sí, demostrant la robustesa del model construït. Aquests resultats milloren significativament els que s'obtenen amb el mètode lineal. Els paràmetres estadístics per a aquests conjunts de compostos es troben en la Taula 3.33, els valors de tensió superficial calculats amb aquest model es poden veure en la Taula 3.29 i la seva representació gràfica enfront dels valors experimentals es mostra en la Figura 3.14.

Taula 3.33. Paràmetres estadístics obtinguts amb el model BPNN.

Conjunt	n	R ²	F	rmse
<i>tset</i>	335	0.961	8145.1	0.96
<i>cvset</i>	39	0.972	1303.6	0.95
<i>pset</i>	94	0.961	2257.7	0.97

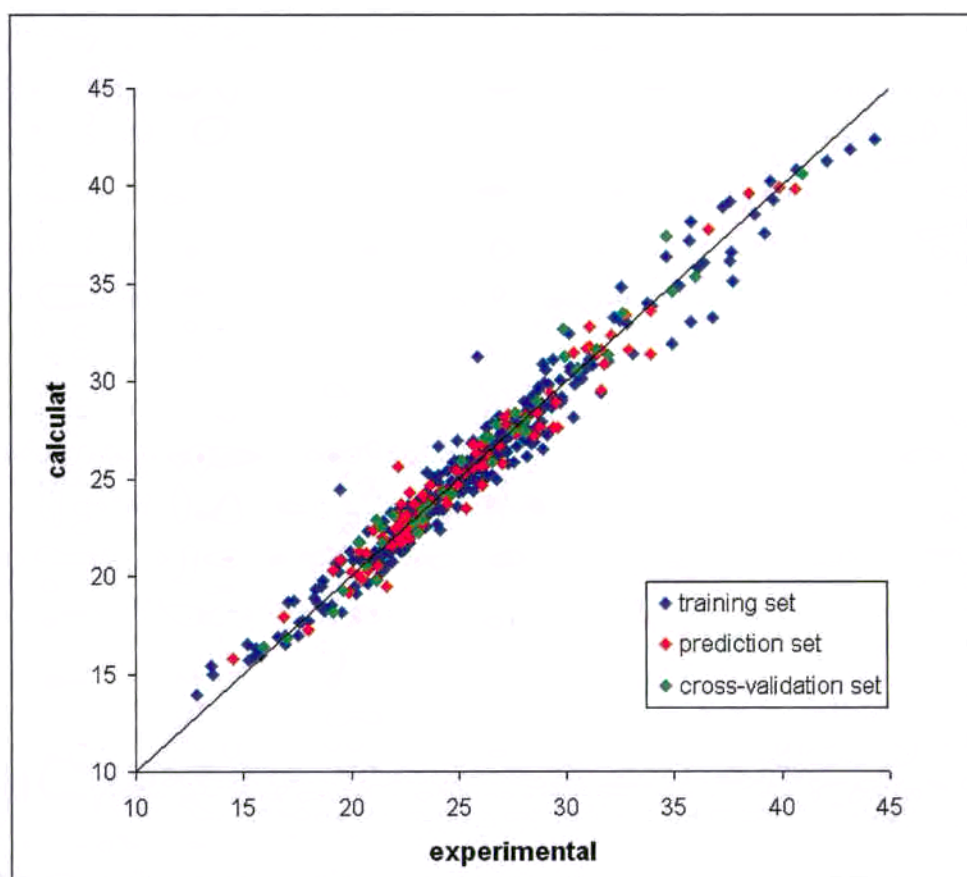


Figura 3.14. Representació gràfica de σ calculada vs. experimental amb el model BPNN.

El valor de *rmse* obtingut per al conjunt de tots els compostos és de 0.97. Dels 468 compostos estudiats, només hi ha 23 que presenten residuals més grans que dues vegades aquest valor, Taula 3.34. Dintre d'aquest conjunt n'hi ha quatre que tenen residuals tres cops més gran que el valor de *rmse* del model i, per tant, es poden considerar *outliers*. Alguns d'aquests compostos corresponen a valors d'una mateixa substància, per exemple, $\text{PO}(\text{OEtOMe})_3$, SnBu_3Cl o $\text{Si}(\text{OEt})_3(3\text{-OPh-Pr})$, a diferents temperatures. Tot i així, no s'aprecia una tendència clara que relacioni aquests errors amb l'estructura dels compostos o amb la temperatura a què s'ha fet la mesura. D'entre tots els compostos presents en aquesta taula, l'únic que es repeteix respecte els trobats amb el model BPNN a 20°C és el SnMe_3Ph a 25°C, en tots dos casos els residuals són positius, encara que amb aquest nou model aquest compost no es podria considerar *outlier*. Hi ha alguns compostos que ja sortien amb errors grans en l'altre model BPNN que aquí apareixen a altres temperatures: $\text{Si}(\text{OEt})_3(3\text{-OPh-Pr})$ i $\text{As}(\text{OPr})_3$.

Taula 3.34. Prediccions amb valors més grans que dues vegades el *rmse**.

Compost	T (°C)	exp.	BPNN	residual
POF(OEt) ₂	20	25.9	31.25	-5.35
SnBu ₃ (NCO)	80	19.5	24.46	-4.96
SnBu ₃ (NCO)	70	22.2	25.66	-3.46
Si(OEt) ₃ (3-OPh-Pr)	20	29.83	32.73	-2.90
PPh ₃	107	34.65	37.47	-2.82
SnPent ₄	25	24.1	26.65	-2.55
PPh ₃	96	35.76	38.16	-2.40
PO(OEtOMe) ₃	85	30.14	32.49	-2.35
PO(OEtOMe) ₃	60	32.53	34.82	-2.29
SnBu ₃ (NCO)	60	24.92	26.97	-2.05
Si(OEt) ₃ (3-OPh-Pr)	30	28.94	30.91	-1.97
As(OPr) ₃	40	28.14	26.17	1.97
SnBu ₃ Cl	20	31.6	29.61	1.99
GeHex ₄	12	29.62	27.62	2.00
GeHex ₄	5	30.34	28.19	2.15
PO(OMe) ₂ Et	60	31.58	29.42	2.16
SnMe ₄	0	21.65	19.47	2.18
SnMe ₃ Ph	25	28.9	26.54	2.36
PO(OMe) ₂ Et	40	33.89	31.4	2.49
PO(OMe) ₃	18	37.76	35.14	2.62
SnBu ₃ (NCO)	20	35.8	33.06	2.74
POCl(OEt)(NMe ₂)	20	34.9	31.97	2.93
PO(OMe) ₂ Et	20	36.77	33.27	3.50

* Els compostos ombrejats tenen residuals més grans que tres vegades el *rmse*

L'anàlisi dels paràmetres estadístics que s'obtenen amb aquest model per als subconjunts de cada metall és una altra manera de comprovar la robustesa del model, Taula 3.35. Els resultats que s'obtenen amb aquest procediment són semblants entre sí i als obtinguts amb el model global, demostrant així la seva bondat. Tot i així, és fàcil comprovar que hi ha alguns subconjunts pels quals s'obtenen uns errors apreciablement més elevats que per a la resta, per exemple, els compostos d'arsènic i, en especial, els d'estany.

Taula 3.35. Estadística per als subconjunts de compostos.

Conjunt	n	R ²	F	rmse
As	22	0.919	227.7	1.13
Ge	90	0.915	953.1	0.62
P	229	0.960	5463.6	0.95
Si	83	0.967	2338.5	0.85
Sn	44	0.916	459.8	1.36

En la Taula 3.34 només hi ha un compost d'arsènic que presenti un residual dues vegades més gran que el *rmse* del model: As(OPr)₃, la qual cosa fa pensar que l'error mitjà d'aquesta família no és degut a causes estructurals, sinó més aviat sembla conseqüència del baix nombre de compostos d'aquest element. Dels vuit compostos d'estany que apareixen en la Taula 3.34, quatre corresponen a mesures del SnBu₃(NCO) a diferents temperatures. Això indica que, possiblement, els valors experimentals per a aquest compost no són prou acurats.

Model no lineal: Support Vector Machines

El model desenvolupat amb SVM utilitza el mètode de selecció de descriptors heurístic-SVM, a partir dels 185 descriptors seleccionats en el mètode lineal. Els descriptors escollits amb aquest mètode es troben en la Taula 3.36.

Taula 3.36. Descriptors continguts en el model SVM.

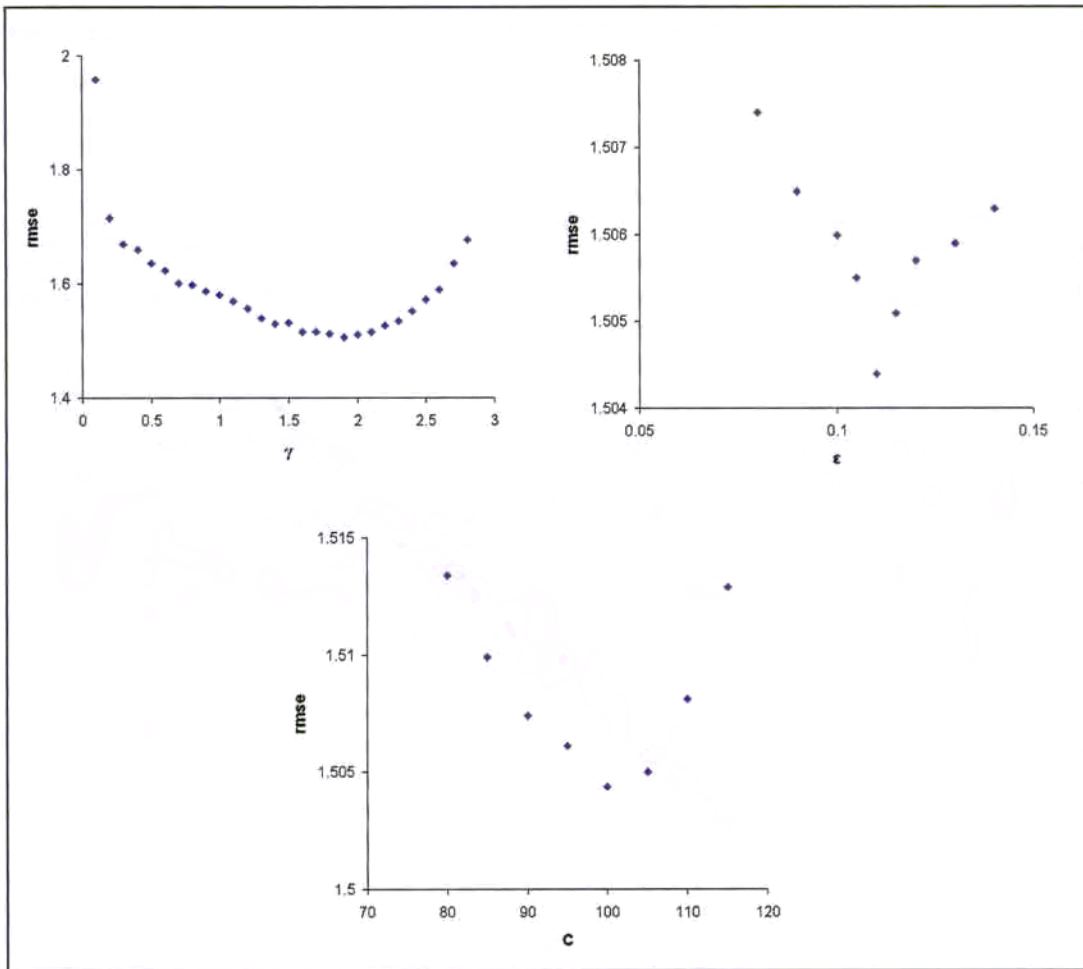
Descriptor	Classe
Índex de Kier & Hall d'ordre 3	Topològic
Càrrega relativa negativa, RPCS	CPSA
Diferència entre superfícies carregades, DPSA-3	CPSA
Ordre d'enllaç σ-σ màxim	Quàntic
Energia de ressonància màxima d'un enllaç C-H	Quàntic
Energia d'atracció electró-nucli de l'àtom metàl·lic	Quàntic
Temperatura	-

El model obtingut conté diversos descriptors que ja han aparegut en models anteriors. Així, en aquest cas es repeteixen la temperatura, RPCS i DPSA-3. Altres descriptors, l'índex de Kier & Hall d'ordre 3, l'energia de ressonància màxima d'un enllaç C-H i l'energia d'atracció electró-nucli de l'àtom metàl·lic, són molt semblants a d'altres que ja han aparegut en la resta de models. El descriptor índex de Kier & Hall d'ordre 3 es calcula de manera anàloga a la resta de descriptors d'aquest tipus:⁴³

$${}^3\chi = \sum_i^n \sum_{j>i} \sum_{k>j} \sum_{l>k} \left(\frac{1}{\delta_i \delta_j \delta_k \delta_l} \right)^{\frac{1}{2}}$$

on n és el nombre total d'àtoms de la molècula (sense comptar hidrògens) i δ_i , δ_j , δ_k i δ_l són la connectivitat atòmica d'ordre zero de cada àtom. L'energia d'atracció electrònica de l'àtom metàl·lic, que es calcula amb el mètode proposat per Clementi,⁶⁷ permet saber quin tipus d'enllaç es forma entre aquest àtom i els lligands que l'envolten. Així, els compostos amb lligands alquílics presenten valors més elevats per aquest descriptor, donant a entendre que en ells la separació de càrregues és menor i, per tant, l'enllaç format entre els dos àtoms té un major grau de covalència. En l'últim descriptor cal no confondre el tipus d'enllaç del qual es parla: sigma (σ), amb la propietat estudiada: σ . L'ordre d'enllaç σ - σ màxim es correspon al valor màxim d'ordre d'enllaç sigma entre dos àtoms de la molècula. En els compostos estudiats aquest enllaç se sol trobar sobre l'enllaç C-H més llunyà al metall del substituent orgànic més llarg del compost. Potser aquest descriptor es podria relacionar amb la densitat electrònica en els lligands orgànics que els permet interaccionar de manera polar però, realment es fa molt difícil relacionar aquest descriptor amb la propietat estudiada. Tot i així, la resta de descriptors inclosos en el model permeten una interpretació fisicoquímica de la tensió superficial en base a la informació que contenen sobre les interaccions intermoleculares que es produeixen en el líquid. Les interaccions de tipus electrostàtic venen descrites pels descriptors d'aquest mateix tipus: RPCS i DPSA-3. Les interaccions que es produeixen per la ramificació en les cadenes es troben recollides en el descriptor índex de Kier & Hall d'ordre 3.

Com tots els models desenvolupats per SVM, cal optimitzar els tres paràmetres característics per tal d'ajustar el model al càlcul definitiu de la tensió superficial, Figura 3.15. El valor òptim trobat pel paràmetre γ (amb $\epsilon=0.1$ i $C=100$) és 1. L'optimització de ϵ (amb $\gamma=1.9$ i $C=100$) condueix a un valor de 0.11 per a aquest paràmetre. El paràmetre C s'optimitza amb $\gamma=1.9$ i $\epsilon=0.11$ i el valor òptim trobat és 100.


 Figura 3.15. Optimització dels paràmetres γ , ϵ i C .

El model construït amb aquests procediments d'optimització utilitza 318 compostos del *tset*, d'entre els 374 disponibles, per calcular la tensió superficial. Els resultats estadístics obtinguts amb aquest model són bons i com en el cas anterior estan a mig camí entre el model lineal i el BPNN, la Taula 3.37 conté els valors del coeficient de determinació R^2 , el test- F i *rmse* per els *tset* i *pset*.

Taula 3.37. Paràmetres estadístics obtinguts amb el model SVM.

Conjunt	n	R^2	F	rmse
<i>tset</i>	374	0.947	6650.3	1.11
<i>pset</i>	94	0.954	1928.8	1.04

Els valors de tensió superficial calculats amb aquest model es troben en la Taula 3.29 i la seva representació gràfica enfront dels valors experimentals es pot veure en la Figura 3.16.

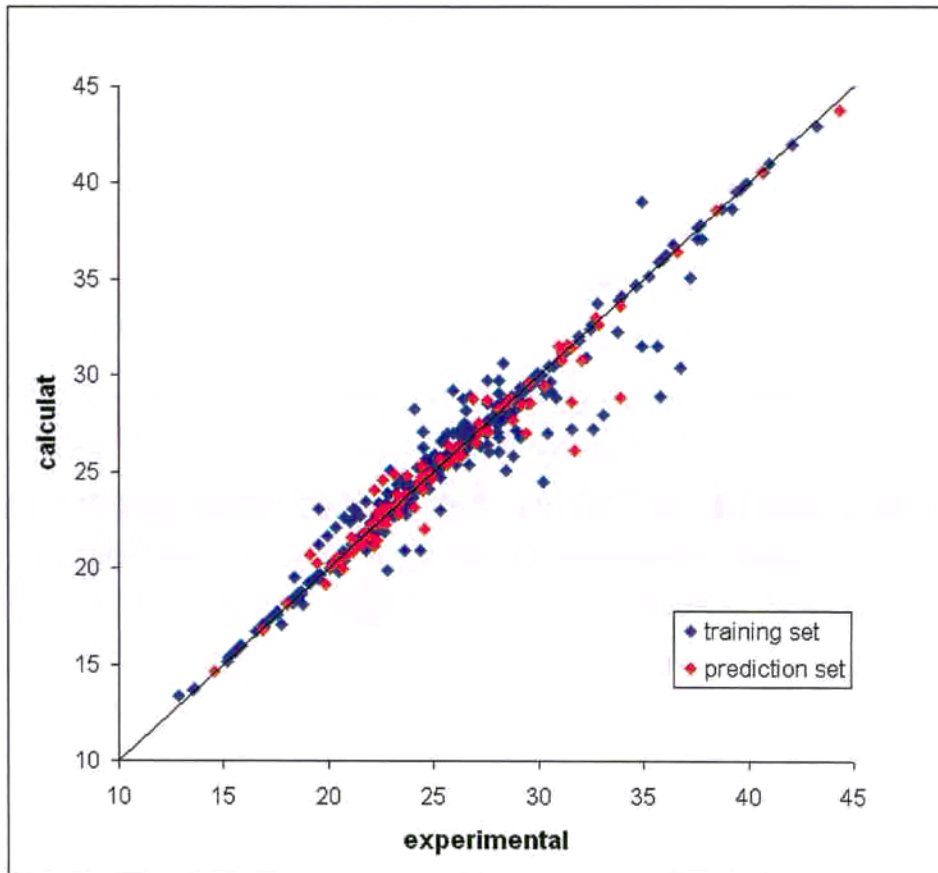


Figura 3.16. Representació gràfica de σ calculada vs. experimental amb el model SVM.

Resulta fàcilment observable l'alineació que presenten els valors d'alguns compostos a diferents temperatures per sobre o per sota de la diagonal. Així, per sobre de la diagonal, amb prediccions per sobre dels valor experimentals (20-25 mN/m) apareixen tots els valors de $\text{Ge}(\text{O}^i\text{Pr})_4$ i $\text{SnEt}(\text{NCO})_4$ a les diferents temperatures; mentre que per sota de la diagonal, amb valors estimats inferiors als experimentals, apareix el SiMe_2Br_2 (23-28 mN/m) a totes les temperatures en què s'ha determinat. Aquest comportament es deu, probablement, a diferències sistemàtiques grans en els valors determinats experimentalment respecte a la resta d'observacions.

Els compostos que presenten valors de residual més gran que dues vegades el *rmse* del model es troben en la Taula 3.38. En aquest cas, la majoria de residuals són positius i hi ha alguns que corresponen a determinacions d'un únic compost a diferents temperatures: SnBu₃(NCO), As(OPr)₃, SiMe₂Br₂, SnBu₃Cl i PO(OMe)₂Et. En aquest cas hi ha un bon grapat de compostos que presenten residuals superior a tres vegades el *rmse* del model, 13 en total, alguns d'ells ja apareixien en el model BPNN: SnBu₃(NCO) a 20 i 80°C, PO(OMe)₂Et a 20 i 60°C i SnBu₃Cl a 20°C, entre d'altres.

 Taula 3.38. Prediccions amb valors més grans que dues vegades el *rmse*^{*}.

Compost	T (°C)	Exp.	BPNN	residual
Sn(Pent) ₄	25	24.10	28.22	-4.12
POCl ₂ (OMe)	20	34.90	39.01	-4.11
SnBu ₃ (NCO)	80	19.50	23.02	-3.52
POF(OEt) ₂	20	25.90	29.19	-3.29
Sn ^t Bu ₄	25	24.50	27.02	-2.52
PHO(O ^t Pr) ₂	20	26.40	28.78	-2.38
SiMe ₂ PhBr	25	28.30	30.62	-2.32
SnMe ₃ (cC ₆ H ₁₁)	25	25.30	22.99	2.31
As(OPr) ₃	30	29.11	26.75	2.36
As(OPr) ₃	27	29.41	26.99	2.42
SiMe ₂ Br ₂	25	24.57	21.99	2.58
SiMe ₂ Br ₂	35	23.63	20.93	2.70
SiMe ₂ Br ₂	45	22.80	19.85	2.95
SnBu ₃ Cl	20	31.60	28.60	3.00
SnBu ₃ Cl	50	28.80	25.77	3.03
POCl(OEt)N(Me) ₂	20	34.90	31.54	3.36
SnBu ₃ (NCO)	40	30.40	27.01	3.39
PO(OMe) ₂ Et	85	28.48	25.07	3.41
SnBu ₃ Cl	100	24.40	20.91	3.49
POCl ₂ N(Et) ₂	20	35.70	31.49	4.21
PO(OMe) ₂ Et	60	31.58	27.22	4.36
PO(OMe) ₂ Et	40	33.89	28.87	5.02
SnBu ₃ (NCO)	30	33.10	27.98	5.12
SnBuCl ₃	20	32.60	27.22	5.38
SnBu ₂ Cl ₂	45	31.74	26.08	5.66
SnBuCl ₃	50	30.20	24.46	5.74
PO(OMe) ₂ Et	20	36.77	30.40	6.37
SnBu ₃ (NCO)	20	35.80	28.91	6.89

^{*} Els compostos ombrejats tenen residuals més grans que tres vegades el *rmse*

Si s'analitzen els resultats que s'obtenen per als diferents subconjunts de compostos classificats pel metall que contenen, Taula 3.39, es poden apreciar grans similituds amb els resultats que s'obtenen amb el model BPNN.

Taula 3.39. Estadística per als conjunts de compostos.

Conjunt	n	R ²	F	rmse
As	22	0.950	380.7	0.90
Ge	90	0.944	1495.8	0.55
P	229	0.957	5062.4	0.97
Si	83	0.971	2673.3	0.77
Sn	44	0.768	138.7	1.87

De nou, els pitjors resultats s'obtenen amb els compostos d'estany, de fet, SnBu₃(NCO) i SnBu₃Cl apareixen quatre vegades en la Taula 3.38, la qual cosa fa pensar que realment la determinació experimental de la tensió superficial d'aquests compostos és errònia. De la resta de compostos d'estany que apareixen en la Taula 3.38, només el SnMe₃(c-C₆H₁₁) no presenta grups butílics. Sembla ser que alguns dels compostos d'estany que presenten aquest tipus de substituent no es troben ben descrits pel model proposat.

Comparació de resultats

Els tres models obtinguts, on la temperatura es troba inclosa com a descriptor, presenten descriptors de tipus CPSA, adequats per descriure interaccions intermoleculares. Alguns d'aquests descriptors són molt semblants i, fins i tot, es repeteixen: RPCG, RPCS, DPSA-3 i PNSA-3. Aquests descriptors són, també, semblants als que s'obtenen en els models desenvolupats per calcular aquesta mateixa propietat a 20°C.

Com era d'esperar, els models desenvolupats amb mètodes no lineals permeten obtenir resultats més acurats que el model establert amb regressió lineal. Així, el model lineal presenta un coeficient de correlació un 10% més baix i uns errors gairebé el doble de grans que els que s'obtenen amb els models construïts amb xarxes neuronals BPNN i *Support Vector Machines*. Dintre dels models no lineals, sembla que el model BPNN és capaç d'ajustar millor el càlcul, fent que els resultats que s'obtenen per tots els subconjunts de compostos siguin semblants; per contra, el model SVM sembla fallar pels compostos que contenen estany.

Taula 3.29: Valors experimentals i calculats de σ (mN/m) a diferents temperatures.

#	Compost	T (°C)	Exp.	MLR	BPNN	SVM
1	As(OBu) ₃ ^{a,b}	20	26.53	27.90	27.87	27.03
2	As(OBu) ₃ ^d	30	25.71	27.04	26.79	26.28
3	As(OBu) ₃ ^{a,b}	40	24.88	26.19	25.73	25.53
4	As(O'Pr) ₃ ^{a,b}	27	26.20	25.98	25.12	26.09
5	As(O'Pr) ₃ ^{a,b}	30	25.91	25.71	24.74	25.84
6	As(O'Pr) ₃ ^{a,b}	40	24.96	24.86	23.55	25.03
7	As(OMe) ₃ ^{a,c}	20	31.44	29.66	31.69	31.33
8	As(OMe) ₃ ^{a,b}	30	30.50	28.81	30.33	30.48
9	As(OMe) ₃ ^d	40	29.54	27.95	28.94	29.61
10	As(OPent) ₃ ^{a,b}	22	26.27	27.85	27.68	26.38
11	As(OPent) ₃ ^{a,b}	30	25.59	27.15	26.85	25.78
12	As(OPent) ₃ ^{a,b}	40	24.75	26.30	25.85	25.03
13	As(OPr) ₃ ^d	27	29.41	26.61	27.68	26.99
14	As(OPr) ₃ ^{a,b}	30	29.11	26.34	27.31	26.75
15	As(OPr) ₃ ^{a,b}	40	28.14	25.49	26.17	25.99
16	AsPh ₃ ^{a,b}	80	39.20	35.53	37.57	38.61
17	AsPh ₃ ^{a,b}	101	36.40	33.74	36.12	36.76
18	AsPh ₃ ^{a,b}	131	33.80	31.11	33.70	33.91
19	AsPh ₃ ^d	156	31.40	28.98	31.45	31.55
20	AsPh ₃ ^{a,c}	196	28.10	25.60	27.44	27.93
21	AsPh ₃ ^{a,b}	221	25.90	23.48	24.82	25.79
22	AsPh ₃ ^{a,b}	242	24.00	21.69	22.71	24.11
23	GeHex ₄ ^{a,b}	5	30.34	28.02	28.19	29.09
24	GeHex ₄ ^d	12	29.62	27.45	27.62	28.52
25	GeHex ₄ ^{a,b}	22	28.39	26.63	26.82	27.70
26	GeHex ₄ ^{a,b}	45	25.61	24.64	25.02	25.61
27	GeHex ₄ ^{a,b}	52	24.67	24.07	24.55	25.01
28	GeHex ₄ ^{a,b}	59	23.90	23.44	24.04	24.33
29	Ge(O(2,2-Me ₂ -Pr)) ₄ ^{a,b}	20	23.86	26.12	23.71	23.77
30	Ge(O(2,2-Me ₂ -Pr)) ₄ ^{a,b}	25	23.39	25.69	23.21	23.29
31	Ge(O(2,2-Me ₂ -Pr)) ₄ ^{a,c}	30	22.91	25.27	22.72	22.80
32	Ge(O(2,2-Me ₂ -Pr)) ₄ ^d	35	22.44	24.84	22.25	22.31
33	Ge(O(2,2-Me ₂ -Pr)) ₄ ^{a,b}	40	21.96	24.41	21.79	21.80
34	Ge(O(2,2-Me ₂ -Pr)) ₄ ^{a,b}	45	21.48	23.98	21.35	21.30
35	Ge(OBu) ₄ ^{a,b}	20	24.45	24.53	24.48	24.57
36	Ge(OBu) ₄ ^{a,b}	25	24.00	24.10	23.96	24.11
37	Ge(OBu) ₄ ^{a,b}	30	23.54	23.67	23.44	23.65
38	Ge(OBu) ₄ ^{a,b}	35	23.09	23.25	22.94	23.18
39	Ge(OBu) ₄ ^{a,b}	40	22.64	22.82	22.46	22.71
40	Ge(OBu) ₄ ^d	45	22.19	22.39	21.98	22.24
41	Ge(OEt) ₄ ^d	20	23.67	23.85	24.66	23.80
42	Ge(OEt) ₄ ^{a,b}	25	23.23	23.42	24.05	23.32
43	Ge(OEt) ₄ ^{a,b}	30	22.79	22.99	23.44	22.84
44	Ge(OEt) ₄ ^{a,b}	35	22.35	22.57	22.85	22.35
45	Ge(OEt) ₄ ^{a,b}	40	21.91	22.14	22.28	21.86
46	Ge(OEt) ₄ ^{a,c}	45	21.47	21.71	21.72	21.36
47	Ge(O'Bu) ₄ ^{a,b}	20	23.68	24.72	23.55	23.70
48	Ge(O'Bu) ₄ ^d	25	23.25	24.29	23.04	23.23
49	Ge(O'Bu) ₄ ^{a,b}	30	22.81	23.86	22.53	22.75
50	Ge(O'Bu) ₄ ^{a,b}	35	22.38	23.43	22.04	22.27

#	Compost	T (°C)	Exp.	MLR	BPNN	SVM
51	Ge(O'Bu) ₄ ^{a,b}	40	21.95	23.01	21.57	21.78
52	Ge(O'Bu) ₄ ^{a,b}	45	21.52	22.58	21.11	21.29
53	Ge(O'Pr) ₄ ^{a,c}	20	21.17	23.97	22.88	23.03
54	Ge(O'Pr) ₄ ^{a,b}	25	20.76	23.54	22.34	22.58
55	Ge(O'Pr) ₄ ^{a,b}	30	20.35	23.12	21.82	22.12
56	Ge(O'Pr) ₄ ^{a,b}	35	19.95	22.69	21.31	21.65
57	Ge(O'Pr) ₄ ^{a,b}	40	19.54	22.26	20.81	21.19
58	Ge(O'Pr) ₄ ^d	45	19.13	21.83	20.33	20.71
59	Ge(OMe) ₄ ^d	20	22.89	26.35	23.75	23.13
60	Ge(OMe) ₄ ^{a,b}	25	22.49	25.92	23.06	22.66
61	Ge(OMe) ₄ ^{a,b}	30	22.08	25.50	22.38	22.19
62	Ge(OMe) ₄ ^{a,b}	35	21.68	25.07	21.72	21.71
63	Ge(OMe) ₄ ^{a,b}	40	21.27	24.64	21.08	21.23
64	Ge(OMe) ₄ ^{a,b}	45	20.86	24.21	20.47	20.75
65	Ge(OPent) ₄ ^{a,b}	20	24.65	25.11	24.87	25.64
66	Ge(OPent) ₄ ^{a,b}	25	24.20	24.69	24.37	25.19
67	Ge(OPent) ₄ ^d	30	23.76	24.26	23.89	24.74
68	Ge(OPent) ₄ ^{a,c}	35	23.31	23.83	23.41	24.28
69	Ge(OPent) ₄ ^{a,b}	40	22.87	23.40	22.95	23.82
70	Ge(OPent) ₄ ^{a,b}	45	22.42	22.98	22.50	23.35
71	Ge(OPr) ₄ ^{a,b}	20	24.06	24.14	24.32	23.62
72	Ge(OPr) ₄ ^{a,b}	25	23.61	23.71	23.75	23.16
73	Ge(OPr) ₄ ^{a,b}	30	23.16	23.28	23.20	22.70
74	Ge(OPr) ₄ ^d	35	22.72	22.86	22.66	22.23
75	Ge(OPr) ₄ ^{a,b}	40	22.27	22.43	22.13	21.76
76	Ge(OPr) ₄ ^{a,b}	45	21.82	22.00	21.62	21.29
77	Ge(O ^s Bu) ₄ ^{a,b}	20	23.93	24.42	24.19	24.04
78	Ge(O ^s Bu) ₄ ^d	25	23.47	23.99	23.65	23.59
79	Ge(O ^s Bu) ₄ ^{a,b}	30	23.00	23.57	23.14	23.12
80	Ge(O ^s Bu) ₄ ^{a,b}	35	22.53	23.14	22.63	22.66
81	Ge(O ^s Bu) ₄ ^{a,b}	40	22.07	22.71	22.14	22.18
82	Ge(O ^s Bu) ₄ ^{a,b}	45	21.57	22.28	21.66	21.71
83	Ge(O'Bu) ₄ ^{a,c}	20	23.14	24.45	22.26	23.30
84	Ge(O'Bu) ₄ ^{a,b}	25	22.74	24.02	21.75	22.82
85	Ge(O'Bu) ₄ ^{a,b}	30	22.34	23.59	21.25	22.33
86	Ge(O'Bu) ₄ ^{a,b}	35	21.94	23.16	20.77	21.83
87	Ge(O'Bu) ₄ ^{a,b}	40	21.54	22.74	20.31	21.33
88	Ge(O'Bu) ₄ ^d	45	21.14	22.31	19.86	20.82
89	GePent ₄ ^d	7	28.79	26.96	27.65	27.67
90	GePent ₄ ^{a,b}	14	28.02	26.34	27.02	27.05
91	GePent ₄ ^{a,b}	22	27.25	25.67	26.36	26.35
92	GePent ₄ ^{a,b}	34	26.07	24.71	25.46	25.33
93	GePent ₄ ^{a,b}	40	25.30	24.13	24.94	24.71
94	GePent ₄ ^d	48	24.49	23.51	24.40	24.03
95	GePent ₄ ^{a,b}	56	23.40	22.77	23.78	23.22
96	GeEt ₄ ^{a,b}	20	24.06	22.92	25.23	24.17
97	GeEt ₄ ^{a,b}	40	22.21	21.21	23.20	22.32
98	GeEt ₄ ^d	60	20.38	19.50	21.29	20.42
99	GeEt ₄ ^{a,b}	80	18.60	17.79	19.52	18.55
100	GeEt ₄ ^d	100	16.86	16.08	17.94	16.75
101	GeEt ₄ ^{a,b}	120	15.16	14.37	16.56	15.09

#	Compost	T (°C)	Exp.	MLR	BPNN	SVM
102	GeEt ₄ ^{a,b}	140	13.51	12.67	15.41	13.62
103	GeMe ₄ ^d	15	24.92	21.94	24.71	24.58
104	GeMe ₄ ^{a,b}	20	24.25	21.51	24.22	24.14
105	GeMe ₄ ^{a,c}	26	23.58	21.00	23.64	23.60
106	GePr ₄ ^d	8	25.85	24.93	26.32	25.59
107	GePr ₄ ^{a,b}	16	24.81	24.21	25.54	24.85
108	GePr ₄ ^{a,b}	24	24.08	23.55	24.83	24.15
109	GePr ₄ ^{a,b}	32	23.29	22.85	24.10	23.40
110	GePr ₄ ^{a,b}	37	22.63	22.44	23.68	22.96
111	GePr ₄ ^{a,c}	44	22.00	21.87	23.12	22.34
112	GePr ₄ ^d	52	21.07	21.15	22.42	21.55
113	P(OBu) ₃ ^d	20	26.07	26.51	26.79	25.89
114	P(OBu) ₃ ^{a,b}	40	23.84	24.80	25.07	24.15
115	P(OBu) ₃ ^{a,b}	60	22.28	23.09	23.48	22.39
116	P(OBu) ₃ ^{a,c}	85	20.37	20.95	21.73	20.26
117	P(OEt) ₃ ^{a,b}	20	24.06	26.30	24.34	24.08
118	P(OEt) ₃ ^d	40	22.03	24.59	22.47	22.25
119	P(OEt) ₃ ^{a,b}	60	20.51	22.88	20.75	20.40
120	P(OEt) ₃ ^{a,b}	85	18.30	20.74	18.89	18.19
121	P(OEtOBu) ₃ ^{a,b}	20	29.33	31.67	28.67	29.22
122	P(OEtOBu) ₃ ^{a,b}	40	27.70	29.96	26.68	27.72
123	P(OEtOBu) ₃ ^d	60	26.07	28.25	24.72	26.17
124	P(OEtOBu) ₃ ^{a,b}	85	24.12	26.12	22.41	24.23
125	P(OEtOEt) ₃ ^{a,b}	40	28.83	29.80	28.51	28.14
126	P(OEtOEt) ₃ ^{a,b}	60	26.71	28.09	26.31	26.60
127	P(OEtOEt) ₃ ^d	85	24.45	25.96	23.70	24.67
128	P(OEtOEt) ₃ ^{a,b}	20	30.54	31.51	30.70	29.62
129	P(OEtOMe) ₃ ^{a,c}	40	31.91	31.65	31.35	31.80
130	P(OEtOMe) ₃ ^{a,b}	60	29.75	29.94	28.89	29.86
131	P(OEtOMe) ₃ ^{a,b}	85	27.28	27.81	25.79	27.37
132	P(OEtOMe) ₃ ^d	20	33.92	33.36	33.64	33.61
133	P(OEtOPr) ₃ ^d	40	27.30	28.76	28.34	27.19
134	P(OEtOPr) ₃ ^{a,b}	60	25.67	27.05	26.09	25.56
135	P(OEtOPr) ₃ ^{a,b}	85	23.85	24.91	23.44	23.52
136	P(OEtOPr) ₃ ^{a,b}	20	29.04	30.47	30.61	28.76
137	P(OHept) ₃ ^{a,b}	40	26.62	26.01	26.92	26.51
138	P(OHept) ₃ ^d	60	24.86	24.30	25.43	24.90
139	P(OHept) ₃ ^{a,b}	85	23.04	22.17	23.75	22.92
140	P(OHept) ₃ ^{a,b}	20	28.19	27.72	28.50	28.08
141	P(OHex) ₃ ^{a,b}	40	25.94	25.08	25.82	26.02
142	P(OHex) ₃ ^{a,c}	60	24.28	23.37	24.34	24.36
143	P(OHex) ₃ ^d	85	22.32	21.23	22.71	22.32
144	P(OHex) ₃ ^{a,b}	20	27.54	26.78	27.42	27.65
145	P(OMe) ₃ ^{a,b}	40	24.88	26.94	25.64	24.99
146	P(OMe) ₃ ^d	20	27.18	28.65	28.19	26.99
147	P(OOct) ₃ ^{a,b}	20	28.18	28.27	28.96	28.44
148	P(OOct) ₃ ^{a,b}	40	26.98	26.56	27.38	26.93
149	P(OOct) ₃ ^{a,b}	60	25.35	24.85	25.88	25.38
150	P(OOct) ₃ ^{a,b}	85	23.36	22.71	24.19	23.47
151	P(OPr) ₃ ^{a,b}	20	25.04	26.73	25.94	25.11
152	P(OPr) ₃ ^d	40	23.32	25.02	24.15	23.36

#	Compost	T (°C)	Exp.	MLR	BPNN	SVM
153	P(OPr) ₃ ^{a,b}	60	21.52	23.31	22.49	21.57
154	P(OPr) ₃ ^{a,b}	85	19.30	21.17	20.65	19.42
155	PHO(OBu) ₂ ^{a,b}	20	28.26	28.76	28.96	28.60
156	PHO(OBu) ₂ ^{a,c}	40	26.20	27.05	27.11	27.16
157	PHO(OBu) ₂ ^d	60	25.31	25.34	25.30	25.64
158	PHO(OBu) ₂ ^{a,b}	85	23.59	23.20	23.18	23.70
159	PHO(OEt) ₂ ^{a,b}	20	30.87	28.30	30.60	28.82
160	PHO(OEt) ₂ ^{a,b}	40	28.96	26.59	28.72	27.13
161	PHO(OEt) ₂ ^{a,b}	60	26.66	24.88	26.78	25.35
162	PHO(OEt) ₂ ^d	85	24.11	22.75	24.38	23.10
163	PHO(OEtOBu) ₂ ^d	20	31.10	32.63	31.83	30.80
164	PHO(OEtOBu) ₂ ^{a,b}	40	29.69	30.92	30.07	29.53
165	PHO(OEtOBu) ₂ ^{a,b}	60	28.26	29.21	28.21	28.16
166	PHO(OEtOBu) ₂ ^{a,b}	85	27.56	27.08	25.86	26.37
167	PHO(OEtOEt) ₂ ^{a,b}	20	33.78	32.60	33.99	32.24
168	PHO(OEtOEt) ₂ ^d	40	32.09	30.89	32.40	30.78
169	PHO(OEtOEt) ₂ ^{a,c}	60	30.56	29.18	30.63	29.19
170	PHO(OEtOEt) ₂ ^{a,b}	85	26.92	27.05	28.27	27.11
171	PHO(OEtOHept) ₂ ^{a,b}	20	30.96	33.18	30.81	30.74
172	PHO(OEtOHept) ₂ ^{a,b}	40	29.40	31.47	29.02	29.49
173	PHO(OEtOHept) ₂ ^{a,b}	60	28.26	29.76	27.21	28.15
174	PHO(OEtOHept) ₂ ^{a,b}	85	26.22	27.62	24.99	26.43
175	PHO(OEtOHex) ₂ ^{a,b}	20	31.17	33.01	30.86	31.06
176	PHO(OEtOHex) ₂ ^{a,b}	40	29.82	31.30	29.06	29.81
177	PHO(OEtOHex) ₂ ^d	60	28.48	29.59	27.22	28.47
178	PHO(OEtOHex) ₂ ^{a,b}	85	26.73	27.45	24.98	26.73
179	PHO(OEtOMe) ₂ ^{a,b}	20	37.68	35.43	36.62	37.79
180	PHO(OEtOMe) ₂ ^{a,c}	40	35.96	33.72	35.35	36.04
181	PHO(OEtOMe) ₂ ^{a,b}	60	33.97	32.01	33.87	34.08
182	PHO(OEtOMe) ₂ ^d	85	30.96	29.87	31.76	31.48
183	PHO(OEtOPent) ₂ ^{a,b}	20	31.04	32.91	31.16	30.74
184	PHO(OEtOPent) ₂ ^{a,b}	40	29.35	31.20	29.36	29.46
185	PHO(OEtOPent) ₂ ^{a,b}	60	28.48	29.49	27.51	28.09
186	PHO(OEtOPent) ₂ ^{a,b}	85	26.57	27.35	25.22	26.31
187	PHO(OEtOPr) ₂ ^d	20	31.07	32.63	32.86	30.85
188	PHO(OEtOPr) ₂ ^{a,b}	40	29.38	30.92	31.16	29.53
189	PHO(OEtOPr) ₂ ^{a,b}	60	28.51	29.21	29.33	28.11
190	PHO(OEtOPr) ₂ ^{a,b}	85	26.59	27.08	26.96	26.25
191	PHO(OHept) ₂ ^{a,c}	20	28.67	31.08	29.00	28.82
192	PHO(OHept) ₂ ^{a,b}	40	27.38	29.37	27.18	27.42
193	PHO(OHept) ₂ ^{a,b}	60	25.73	27.66	25.42	25.97
194	PHO(OHept) ₂ ^{a,b}	85	24.04	25.53	23.39	24.17
195	PHO(OHex) ₂ ^{a,b}	20	28.70	29.65	29.55	28.89
196	PHO(OHex) ₂ ^d	40	27.19	27.94	27.78	27.49
197	PHO(OHex) ₂ ^{a,b}	60	25.74	26.23	26.06	26.01
198	PHO(OHex) ₂ ^{a,b}	85	23.70	24.10	24.06	24.14
199	PHO(O'Pr) ₂ ^{a,b}	20	26.40	26.65	27.13	28.78
200	PHO(OMe) ₂ ^{a,b}	20	37.60	31.81	36.20	37.09
201	PHO(OMe) ₂ ^{a,b}	40	35.24	30.10	34.90	35.13
202	PHO(OMe) ₂ ^d	60	32.76	28.39	33.40	32.97
203	PHO(OMe) ₂ ^{a,c}	85	29.92	26.25	31.26	30.12

#	Compost	T (°C)	Exp.	MLR	BPNN	SVM
204	PHO(OOct) ₂ ^{a,b}	20	29.00	31.61	29.95	28.89
205	PHO(OOct) ₂ ^{a,b}	40	27.57	29.90	28.18	27.52
206	PHO(OOct) ₂ ^{a,b}	60	26.11	28.19	26.44	26.11
207	PHO(OOct) ₂ ^{a,b}	85	24.44	26.06	24.42	24.36
208	PHO(OPent) ₂ ^{a,b}	20	28.46	28.97	29.01	28.65
209	PHO(OPent) ₂ ^{a,b}	40	26.98	27.27	27.21	27.22
210	PHO(OPent) ₂ ^{a,b}	60	25.31	25.56	25.45	25.71
211	PHO(OPent) ₂ ^d	85	23.26	23.42	23.42	23.80
212	PHO(OPr) ₂ ^d	20	29.23	28.61	29.40	28.47
213	PHO(OPr) ₂ ^{a,c}	40	27.48	26.91	27.51	27.01
214	PHO(OPr) ₂ ^{a,b}	60	25.56	25.20	25.62	25.45
215	PHO(OPr) ₂ ^{a,b}	85	23.36	23.06	23.36	23.47
216	PO(OBu) ₂ Et ^{a,b}	20	28.05	27.61	27.97	27.44
217	PO(OBu) ₂ Et ^d	40	26.37	25.91	26.25	25.80
218	PO(OBu) ₂ Et ^{a,b}	60	24.81	24.20	24.62	24.07
219	PO(OBu) ₂ Et ^{a,b}	85	22.67	22.06	22.78	21.89
220	PO(OBu) ₃ ^{a,b}	20	27.55	28.47	27.79	28.55
221	PO(OBu) ₃ ^{a,b}	40	25.96	26.76	26.02	26.96
222	PO(OBu) ₃ ^d	60	24.44	25.05	24.35	25.28
223	PO(OBu) ₃ ^{a,b}	85	23.25	22.91	22.46	23.15
224	PO(OEt) ₂ Et ^{a,b}	20	28.68	27.81	29.72	28.79
225	PO(OEt) ₂ Et ^{a,c}	40	26.73	26.10	27.81	27.04
226	PO(OEt) ₂ Et ^{a,b}	60	24.78	24.39	25.93	25.19
227	PO(OEt) ₂ Et ^d	85	22.34	22.26	23.71	22.82
228	PO(OEt) ₃ ^{a,b}	18	30.22	29.19	30.57	29.40
229	PO(OEt) ₃ ^{a,b}	62	25.99	25.48	26.41	25.46
230	PO(OEt) ₃ ^{a,b}	87	23.53	23.28	24.00	23.02
231	PO(OEt) ₃ ^d	121	20.72	20.40	21.18	19.97
232	PO(OEtOMe) ₃ ^d	20	36.61	36.25	37.81	36.37
233	PO(OEtOMe) ₃ ^{a,b}	40	34.66	34.54	36.43	34.60
234	PO(OEtOMe) ₃ ^{a,b}	60	32.53	32.83	34.82	32.64
235	PO(OEtOMe) ₃ ^{a,b}	85	30.14	30.70	32.49	30.03
236	PO(OEtOPr) ₃ ^{a,b}	20	32.24	33.50	33.27	30.89
237	PO(OEtOPr) ₃ ^d	40	30.30	31.79	31.49	29.44
238	PO(OEtOPr) ₃ ^{a,b}	60	28.69	30.08	29.59	27.88
239	PO(OEtOPr) ₃ ^{a,c}	85	26.40	27.94	27.15	25.88
240	PO(OHept) ₂ Et ^{a,b}	20	28.46	28.74	28.59	28.63
241	PO(OHept) ₂ Et ^{a,b}	40	27.13	27.03	27.04	27.02
242	PO(OHept) ₂ Et ^d	60	25.65	25.32	25.59	25.34
243	PO(OHept) ₂ Et ^{a,b}	85	23.93	23.19	23.97	23.23
244	PO(OHept) ₃ ^{a,b}	20	28.11	29.55	28.35	29.70
245	PO(OHept) ₃ ^{a,b}	40	26.53	27.84	26.74	28.19
246	PO(OHept) ₃ ^{a,b}	60	25.37	26.13	25.23	26.60
247	PO(OHept) ₃ ^{a,b}	85	23.57	23.99	23.55	24.61
248	PO(OHex) ₂ Et ^d	20	28.11	28.49	28.24	28.21
249	PO(OHex) ₂ Et ^{a,b}	40	26.74	26.78	26.65	26.59
250	PO(OHex) ₂ Et ^{a,b}	60	25.15	25.07	25.16	24.90
251	PO(OHex) ₂ Et ^{a,b}	85	23.71	22.93	23.51	22.77
252	PO(OHex) ₃ ^{a,c}	20	28.11	29.20	28.09	29.02
253	PO(OHex) ₃ ^{a,b}	40	26.50	27.49	26.45	27.48
254	PO(OHex) ₃ ^{a,b}	60	25.03	25.78	24.90	25.88

#	Compost	T (°C)	Exp.	MLR	BPNN	SVM
255	PO(OHex) ₃ ^{a,b}	85	23.24	23.65	23.17	23.86
256	PO(O'Bu) ₃ ^{a,b}	15	26.74	29.12	27.05	28.92
257	PO(O'Bu) ₃ ^d	61	23.13	25.19	23.16	24.87
258	PO(O'Bu) ₃ ^{a,b}	87	21.08	22.96	21.29	22.45
259	PO(O'Bu) ₃ ^{a,b}	121	18.35	20.12	19.36	19.52
260	PO(O'Pr) ₃ ^{a,b}	15	25.68	27.55	27.01	26.98
261	PO(O'Pr) ₃ ^{a,b}	62	21.51	23.58	22.81	22.69
262	PO(O'Pr) ₃ ^d	87	19.48	21.41	20.82	20.23
263	PO(O'Pr) ₃ ^{a,b}	120	17.06	18.57	18.65	17.17
264	PO(OMe)(OPh) ₂ ^{a,b}	25	44.32	40.84	42.38	43.72
265	PO(OMe)(OPh) ₂ ^{a,b}	35	43.21	39.99	41.87	42.90
266	PO(OMe)(OPh) ₂ ^{a,b}	45	42.10	39.13	41.29	41.99
267	PO(OMe)(OPh) ₂ ^{a,c}	55	40.99	38.28	40.66	41.02
268	PO(OMe)(OPh) ₂ ^d	65	39.88	37.42	39.96	39.99
269	PO(OMe) ₂ Et ^{a,b}	20	36.77	30.29	33.27	30.40
270	PO(OMe) ₂ Et ^d	40	33.89	28.58	31.40	28.87
271	PO(OMe) ₂ Et ^{a,b}	60	31.58	26.87	29.42	27.22
272	PO(OMe) ₂ Et ^{a,b}	85	28.48	24.73	26.88	25.07
273	PO(OMe) ₃ ^{a,b}	18	37.76	33.69	35.14	37.10
274	PO(OMe) ₃ ^d	61	32.88	29.98	31.69	32.65
275	PO(OMe) ₃ ^{a,b}	87	29.82	27.76	29.22	29.71
276	PO(OMe) ₃ ^{a,b}	121	26.06	24.87	25.80	25.95
277	PO(OOct) ₂ Et ^{a,b}	20	28.87	29.11	28.88	28.68
278	PO(OOct) ₂ Et ^d	40	27.63	27.40	27.35	27.08
279	PO(OOct) ₂ Et ^{a,c}	60	25.14	25.69	25.93	25.42
280	PO(OOct) ₂ Et ^{a,b}	85	23.45	23.55	24.33	23.34
281	PO(OPent) ₂ Et ^{a,b}	20	27.97	29.34	28.04	28.42
282	PO(OPent) ₂ Et ^{a,b}	40	26.25	27.63	26.40	26.84
283	PO(OPent) ₂ Et ^{a,b}	60	24.85	25.92	24.86	25.19
284	PO(OPent) ₂ Et ^d	85	22.60	23.78	23.13	23.08
285	PO(OPent) ₃ ^d	20	27.60	28.78	27.58	28.71
286	PO(OPent) ₃ ^{a,b}	40	26.19	27.07	25.90	27.13
287	PO(OPent) ₃ ^{a,b}	60	25.00	25.36	24.32	25.47
288	PO(OPent) ₃ ^{a,b}	85	23.46	23.23	22.56	23.38
289	PO(OPh) ₃ ^d	66	40.63	38.68	39.86	40.53
290	PO(OPh) ₃ ^{a,b}	74	39.64	37.95	39.29	39.66
291	PO(OPh) ₃ ^{a,b}	84	38.71	37.09	38.58	38.60
292	PO(OPr) ₂ Bu ^{a,c}	20	27.98	28.24	27.49	27.94
293	PO(OPr) ₂ Bu ^d	40	26.27	26.53	25.79	26.31
294	PO(OPr) ₂ Bu ^{a,b}	60	24.59	24.82	24.19	24.60
295	PO(OPr) ₂ Bu ^{a,b}	85	22.24	22.68	22.41	22.42
296	PO(OPr) ₂ Et ^{a,b}	20	28.36	27.75	28.45	27.82
297	PO(OPr) ₂ Et ^{a,b}	40	26.29	26.04	26.64	26.20
298	PO(OPr) ₂ Et ^{a,b}	60	24.66	24.33	24.90	24.49
299	PO(OPr) ₂ Et ^d	85	22.45	22.20	22.91	22.31
300	PO(OPr) ₂ Hept ^{a,b}	20	28.58	28.54	27.62	28.19
301	PO(OPr) ₂ Hept ^{a,b}	40	26.83	26.83	26.02	26.58
302	PO(OPr) ₂ Hept ^{a,b}	60	25.19	25.12	24.52	24.89
303	PO(OPr) ₂ Hept ^{a,b}	85	23.17	22.98	22.87	22.75
304	PO(OPr) ₂ Hex ^{a,b}	20	28.07	28.40	27.54	28.19
305	PO(OPr) ₂ Hex ^{a,c}	40	26.59	26.69	25.90	26.57

#	Compost	T (°C)	Exp.	MLR	BPNN	SVM
306	PO(OPr) ₂ Hex ^{a,b}	60	25.21	24.98	24.38	24.87
307	PO(OPr) ₂ Hex ^d	85	23.35	22.84	22.69	22.72
308	PO(OPr) ₂ Me ^d	20	28.64	28.57	28.39	28.75
309	PO(OPr) ₂ Me ^{a,b}	40	26.66	26.86	26.56	27.17
310	PO(OPr) ₂ Me ^{a,b}	60	24.89	25.15	24.79	25.50
311	PO(OPr) ₂ Me ^{a,b}	85	22.72	23.01	22.76	23.35
312	PO(OPr) ₂ Oct ^{a,b}	20	28.83	28.69	27.92	28.60
313	PO(OPr) ₂ Oct ^{a,b}	40	26.97	26.98	26.33	26.99
314	PO(OPr) ₂ Oct ^{a,b}	60	25.34	25.27	24.85	25.31
315	PO(OPr) ₂ Oct ^{a,b}	85	23.26	23.13	23.20	23.20
316	PO(OPr) ₂ Pent ^{a,b}	20	28.25	28.33	27.48	28.08
317	PO(OPr) ₂ Pent ^d	40	27.04	26.62	25.82	26.46
318	PO(OPr) ₂ Pent ^{a,c}	60	24.64	24.91	24.27	24.75
319	PO(OPr) ₂ Pent ^{a,b}	85	22.70	22.77	22.53	22.59
320	PO(OPr) ₂ Pr ^{a,b}	20	27.98	27.92	27.56	27.86
321	PO(OPr) ₂ Pr ^{a,b}	40	26.20	26.21	25.83	26.20
322	PO(OPr) ₂ Pr ^d	60	24.37	24.50	24.19	24.48
323	PO(OPr) ₂ Pr ^{a,b}	85	22.21	22.36	22.34	22.30
324	PO(OPr) ₃ ^{a,b}	17	29.46	29.48	28.92	28.56
325	PO(OPr) ₃ ^{a,b}	63	25.05	25.54	24.70	24.82
326	PO(OPr) ₃ ^{a,b}	88	23.04	23.44	22.65	22.74
327	PO(OPr) ₃ ^d	122	20.05	20.51	20.24	20.00
328	POCl(OEt)(NMe ₂) ^{a,b}	20	34.90	30.63	31.97	31.54
329	POCl ₂ (OEt) ^{a,b}	20	32.80	33.18	32.97	33.73
330	POCl ₂ (OMe) ^{a,c}	20	34.90	36.15	34.62	39.01
331	POCl ₂ (NEt ₂) ^{a,b}	20	35.70	30.64	37.19	31.49
332	POCl ₂ (NMe ₂) ^{a,b}	20	36.10	31.42	35.79	36.24
333	POF(OEt) ₂ ^{a,b}	20	25.90	29.71	31.25	29.19
334	POF ₂ (NMe ₂) ^{a,b}	20	25.10	29.61	25.58	25.54
335	PPh(OEt) ₂ ^{a,b}	20	31.95	30.65	31.07	32.06
336	PPh ₃ ^{a,b}	46	40.69	37.58	40.82	40.58
337	PPh ₃ ^{a,b}	58	39.45	36.53	40.23	39.52
338	PPh ₃ ^d	69	38.47	35.61	39.68	38.53
339	PPh ₃ ^{a,b}	78	37.60	34.85	39.20	37.68
340	PPh ₃ ^{a,b}	96	35.76	33.29	38.16	35.87
341	PPh ₃ ^{a,c}	107	34.65	32.34	37.47	34.73
342	Si(O(1-Me-Bu)) ₄ ^{a,b}	30	23.36	22.16	22.86	23.36
343	Si(O(2-Et-Bu)) ₄ ^{a,b}	30	25.17	23.04	24.63	25.16
344	Si(O(2-Et-Hex)) ₄ ^{a,b}	30	25.38	23.80	25.96	26.55
345	Si(OBu) ₄ ^{a,b}	30	22.85	21.88	22.78	22.82
346	Si(OEt) ₃ (3-OPh-Pr) ^{a,c}	20	29.83	27.30	32.73	29.72
347	Si(OEt) ₃ (3-OPh-Pr) ^{a,b}	30	28.94	26.45	30.91	28.92
348	Si(OEt) ₃ (3-OPh-Pr) ^{a,b}	40	28.04	25.59	29.02	28.09
349	Si(OEt) ₃ (3-OPh-Pr) ^{a,b}	50	27.15	24.74	27.12	27.23
350	Si(OEt) ₃ (3-OPh-Pr) ^{a,b}	60	26.25	23.88	25.26	26.36
351	Si(OEt) ₃ (3-OPh-Pr) ^d	70	25.36	23.03	23.49	25.48
352	Si(OEt) ₄ ^{a,b}	20	21.67	21.29	20.54	21.25
353	Si(OEt) ₄ ^{a,b}	29	20.75	20.48	19.61	20.37
354	Si(OHept) ₄ ^{a,b}	30	26.19	23.55	25.85	26.21
355	Si(OHex) ₄ ^{a,b}	30	25.44	23.05	25.07	25.25
356	Si(O ⁱ Bu) ₄ ^{a,b}	30	21.96	22.53	21.78	21.85

#	Compost	T (°C)	Exp.	MLR	BPNN	SVM
357	Si(O'Pent) ₄ ^{a,b}	30	22.80	23.33	23.07	23.89
358	Si(O'Pr) ₄ ^{a,c}	30	21.18	21.53	19.76	21.11
359	Si(OOct) ₄ ^{a,b}	30	27.15	23.86	26.55	26.95
360	Si(OPent) ₄ ^{a,b}	30	23.89	22.38	24.12	24.00
361	Si(O ^s Bu) ₄ ^{a,b}	30	22.29	21.70	21.63	22.67
362	SiCl ₂ Ph ₂ ^{a,b}	30	37.24	36.12	38.91	35.08
363	SiEt ₄ ^{a,b}	20	23.04	19.71	22.64	22.68
364	SiEt ₄ ^d	31	21.90	18.76	21.57	21.66
365	SiEt ₄ ^{a,b}	41	20.75	17.91	20.65	20.64
366	SiH(OEt) ₃ ^{a,b}	20	21.25	21.09	20.73	21.14
367	SiH(OEt) ₃ ^d	25	20.80	20.66	20.30	20.63
368	SiHEt ₂ Cl ^d	20	22.30	19.21	21.69	21.43
369	SiHEt ₂ Cl ^{a,b}	25	21.77	18.78	21.17	20.89
370	SiHEt ₃ ^{a,c}	20	20.71	18.10	20.52	20.84
371	SiHEt ₃ ^d	25	20.32	17.67	20.07	20.34
372	SiHEtCl ₂ ^d	20	21.70	19.89	21.64	21.76
373	SiHEtCl ₂ ^{a,b}	25	21.09	19.46	21.08	21.20
374	SiHMeCl ₂ ^d	0	21.43	16.82	22.05	21.17
375	SiHMeCl ₂ ^{a,b}	10	20.07	15.96	20.86	19.96
376	SiHMeCl ₂ ^{a,b}	20	18.70	15.11	19.76	18.72
377	SiHMeCl ₂ ^{a,b}	30	17.33	14.25	18.77	17.46
378	SiMe ₂ Br ₂ ^d	25	24.57	22.66	24.35	21.99
379	SiMe ₂ Br ₂ ^{a,b}	35	23.63	21.81	23.29	20.93
380	SiMe ₂ Br ₂ ^{a,b}	45	22.80	20.95	22.26	19.85
381	SiMe ₂ Cl ₂ ^{a,b}	0	22.29	20.79	21.71	22.25
382	SiMe ₂ Cl ₂ ^d	10	21.23	19.93	20.54	21.27
383	SiMe ₂ Cl ₂ ^{a,b}	20	20.16	19.08	19.49	20.25
384	SiMe ₂ Cl ₂ ^{a,b}	30	19.09	18.22	18.54	19.20
385	SiMe ₂ Cl ₂ ^{a,b}	40	18.03	17.37	17.71	18.13
386	SiMe ₂ Cl ₂ ^{a,b}	50	16.96	16.51	16.98	17.05
387	SiMe ₂ Cl ₂ ^{a,c}	60	15.89	15.66	16.36	15.97
388	SiMe ₂ Ph ₂ ^{a,b}	25	32.47	32.13	33.11	32.36
389	SiMe ₂ Ph ₂ ^d	35	31.59	31.28	31.65	31.42
390	SiMe ₂ Ph ₂ ^{a,b}	45	30.71	30.42	30.14	30.43
391	SiMe ₂ PhBr ^{a,b}	25	28.30	31.40	29.15	30.62
392	SiMe ₂ PhBr ^{a,b}	35	27.59	30.54	27.88	29.72
393	SiMe ₂ PhBr ^d	45	26.88	29.69	26.69	28.78
394	SiMe ₃ Br ^d	25	19.87	20.36	19.17	19.15
395	SiMe ₃ Br ^{a,b}	35	18.75	19.51	18.41	18.11
396	SiMe ₃ Br ^{a,b}	45	17.73	18.65	17.69	17.06
397	SiMe ₃ Cl ^{a,c}	0	19.61	19.90	19.30	19.67
398	SiMe ₃ Cl ^{a,b}	10	18.60	19.04	18.41	18.71
399	SiMe ₃ Cl ^{a,b}	20	17.58	18.19	17.62	17.72
400	SiMe ₃ Cl ^{a,b}	30	16.57	17.33	16.91	16.70
401	SiMe ₃ Cl ^{a,b}	40	15.56	16.48	16.29	15.67
402	SiMe ₃ Cl ^d	50	14.54	15.62	15.75	14.63
403	SiMe ₃ (OH) ^{a,b}	20	18.40	17.90	18.71	18.42
404	SiMe ₃ Ph ^{a,b}	25	24.29	25.74	23.42	24.18
405	SiMe ₃ Ph ^{a,b}	35	23.49	24.89	22.63	23.38
406	SiMe ₃ Ph ^d	45	22.72	24.03	21.93	22.57
407	SiMe ₄ ^{a,b}	20	12.85	17.19	13.96	13.37

#	Compost	T (°C)	Exp.	MLR	BPNN	SVM
408	SiMeCl ₃ ^{a,b}	0	22.46	20.44	21.37	22.35
409	SiMeCl ₃ ^{a,b}	10	21.35	19.59	20.18	21.33
410	SiMeCl ₃ ^{a,b}	20	20.24	18.73	19.11	20.28
411	SiMeCl ₃ ^{a,c}	30	19.13	17.88	18.14	19.20
412	SiMeCl ₃ ^d	40	18.02	17.02	17.29	18.11
413	SiMeCl ₃ ^{a,b}	50	16.91	16.17	16.55	17.02
414	SiMeCl ₃ ^{a,b}	60	15.80	15.31	15.91	15.94
415	SiMeEt ₂ Br ^{a,b}	25	23.55	22.49	25.37	23.56
416	SiMeEt ₂ Br ^d	35	22.71	21.63	24.30	22.51
417	SiMeEt ₂ Br ^{a,b}	45	21.87	20.78	23.24	21.44
418	SiPhCl ₃ ^d	20	31.09	35.90	32.87	31.20
419	SiPhCl ₃ ^{a,b}	40	29.15	34.19	29.85	29.33
420	SiPhCl ₃ ^{a,b}	60	27.21	32.48	27.11	27.32
421	SiPhCl ₃ ^{a,b}	80	25.27	30.77	24.77	25.25
422	SiPhCl ₃ ^{a,c}	100	23.33	29.06	22.87	23.22
423	SiPhCl ₃ ^{a,b}	120	21.39	27.35	21.37	21.30
424	SiPhCl ₃ ^{a,b}	140	19.45	25.64	20.20	19.56
425	SnPent ₄ ^{a,b}	25	24.10	26.14	26.65	28.22
426	SnBu ₂ Cl ₂ ^d	45	31.74	28.86	30.90	26.08
427	SnBu ₂ Cl ₂ ^{a,b}	100	28.10	24.16	27.05	26.75
428	SnBu ₃ (NCO) ^{a,b}	20	35.80	29.97	33.06	28.91
429	SnBu ₃ (NCO) ^{a,b}	30	33.10	29.11	31.47	27.98
430	SnBu ₃ (NCO) ^{a,b}	40	30.40	28.26	29.90	27.01
431	SnBu ₃ (NCO) ^{a,c}	50	27.65	27.40	28.39	26.03
432	SnBu ₃ (NCO) ^{a,b}	60	24.92	26.55	26.97	25.03
433	SnBu ₃ (NCO) ^d	70	22.20	25.69	25.66	24.03
434	SnBu ₃ (NCO) ^{a,b}	80	19.50	24.84	24.46	23.02
435	SnBu ₃ Cl ^d	20	31.60	28.25	29.61	28.60
436	SnBu ₃ Cl ^{a,b}	50	28.80	25.69	27.48	25.77
437	SnBu ₃ Cl ^{a,b}	100	24.40	21.42	24.17	20.91
438	SnBu ₄ ^{a,b}	20	26.50	25.70	26.27	26.98
439	SnBu ₄ ^d	50	24.30	23.14	24.22	24.34
440	SnBu ₄ ^{a,b}	100	20.50	18.87	21.03	19.81
441	SnBuCl ₃ ^{a,c}	20	32.60	31.11	33.52	27.22
442	SnBuCl ₃ ^{a,b}	50	30.20	28.54	30.77	24.46
443	SnBuCl ₃ ^d	100	26.10	24.27	25.66	25.74
444	SnEt ₃ (NCO) ^{a,b}	55	22.95	26.88	22.86	25.07
445	SnEt ₃ (NCO) ^d	60	22.60	26.45	22.46	24.53
446	SnEt ₃ (NCO) ^{a,b}	70	21.80	25.60	21.73	23.44
447	SnEt ₃ (NCO) ^{a,b}	80	21.05	24.74	21.07	22.36
448	SnEt ₄ ^{a,b}	20	26.20	23.55	25.25	25.87
449	SnEt ₄ ^{a,b}	40	24.27	21.84	23.40	24.11
450	SnEt ₄ ^{a,b}	60	22.38	20.13	21.58	22.27
451	SnEt ₄ ^d	80	20.53	18.42	19.84	20.42
452	SnEt ₄ ^{a,b}	100	18.73	16.71	18.24	18.62
453	SnEt ₄ ^{a,c}	120	16.97	15.00	16.85	16.92
454	SnEt ₄ ^{a,b}	140	15.26	13.29	15.70	15.37
455	Sn ⁱ Bu ₄ ^{a,b}	25	24.50	25.69	24.57	27.03
456	Sn ⁱ Pr ₄ ^{a,b}	25	25.70	26.02	24.36	25.81
457	SnMe ₃ (CH=CH ₂) ^{a,b}	25	21.30	25.54	20.89	22.96
458	SnMe ₃ CHex ^{a,b}	25	25.30	25.14	24.57	22.99

#	Compost	T (°C)	Exp.	MLR	BPNN	SVM
459	SnMe ₃ Et ^{a,b}	25	21.60	21.86	22.44	21.15
460	SnMe ₃ Pr ^{a,c}	25	21.40	22.90	22.57	22.78
461	SnMe ₃ Ph ^{a,b}	25	28.90	30.46	26.54	28.79
462	SnMe ₃ Pr ^d	25	22.20	22.73	23.12	21.13
463	SnMe ₄ ^d	0	21.65	23.22	19.47	21.22
464	SnMe ₄ ^{a,b}	20	19.56	21.51	18.19	19.45
465	SnMe ₄ ^{a,b}	40	17.54	19.80	16.99	17.57
466	SnMe ₄ ^{a,b}	60	15.52	18.09	15.91	15.63
467	SnMe ₄ ^{a,b}	80	13.58	16.38	14.96	13.71
468	SnPr ₄ ^{a,b}	25	24.50	24.41	25.41	26.23

^a Training set de MLR i SVM. ^b Training set de BPNN.

^c Cross-validation set de BPNN. ^d Prediction set.

3.3. Entalpia de formació de compostos organometàl·lics

L'entalpia de formació estàndard, o simplement entalpia de formació, $\Delta_f H^\circ$ és una propietat important ja que és una mesura l'estabilitat termodinàmica dels compostos químics. Tot i així, no es considera una propietat termodinàmica primària ja que, com a norma general, els seus valors no s'obtenen amb una mesura directa sino que es deriven de valors experimentals o teòrics d'entalpies de reacció.

La determinació experimental de l'entalpia de formació de compostos orgànics ha estat molt tractada i, actualment, es coneixen valors d'un gran nombre d'aquests compostos. El mètode més utilitzat en aquesta determinació és la combustió calorimètrica. D'altra banda, s'han desenvolupat diversos mètodes empírics que permeten estimar $\Delta_f H^\circ$ amb gran exactitud. Dintre de tots aquests, els més emprats són el mètodes additius, que consideren que el valor d'una propietat experimental és la suma de contribucions, determinades experimentalment, de cada fragment molecular o enllaç present en la molècula. Aquest mètode assumeix, doncs, que la contribució d'un determinat fragment a la propietat es manté constant sigui quina sigui la molècula en què es trobi. L'exemple més senzill d'aplicació de mètodes additius és el càlcul del pes molecular d'un compost a partir dels pesos atòmics dels elements que el formen.

Els mètode additiu més utilitzat per estimar entalpies de reacció de compostos orgànics és "l'additivitat de grups" desenvolupat per Benson i Buss.⁶⁸ Els paràmetres originals d'aquest mètode han estat subseqüentment ampliat i refinats posteriorment.⁶⁹ Una altra aproximació és el mètode de Laidler,⁷⁰ en aquest cas les contribucions són assignades als diferents enllaços presents en la molècula i, per tant, aquestes contribucions reflecteixen la força dels enllaços que descriuen. El refinament d'aquest mètode amb diverses famílies d'hidrocarburs⁷¹ fa que sigui matemàticament equivalent al de Benson, és a dir, tots dos condueixen al mateix valor d'entalpia de formació. Per últim, la revisió feta per Pedley⁷² dels mètodes additius inclou una nova manera d'estimar les entalpies de formació dels compostos orgànics. Aquest mètode, que es basa en l'additivitat d'entalpies d'enllaç, té en compte també entalpies de fragments moleculars, interaccions d'enllaç i correccions d'anell. Tots aquests mètodes proporcionen valors d'entalpia de formació molt pròxims als valors que es poden mesurar experimentalment.

Amb els compostos organometàl·lics la situació no és tan bona com en el cas anterior ja que, normalment, la determinació experimental de $\Delta_f H^\circ$ és difícil de realitzar. Un dels principals problemes que es troba és que la calorimetria de combustió en bomba estàtica, que és el mètode més emprat per fer aquestes determinacions experimentals, no és apropiada ja que la reacció que es produeix pot donar compostos difícils de caracteritzar correctament. Un exemple d'aquest problema seria la combustió d'un compost organometàl·lic que condueix a un òxid no estequiomètric. L'aplicació dels mètodes additius a compostos organometàl·lics és molt restringida degut a la manca, en la majoria de casos, dels valors corresponents als grups i enllaços que contenen els àtoms metàl·lics; l'altre inconvenient, com ja s'ha comentat, és la manca de valors experimentals per a aquest tipus de compostos i a la poca exactitud que presenten aquests valors quan són coneguts. Així doncs, els mètodes additius han estat poc aplicats a conjunts de compostos organometàl·lics.⁷³

Una altra manera d'estimar els valors de $\Delta_f H^\circ$ de compostos organometàl·lics és el "Mètode del Potencial Covalent" creat per Benson.⁷⁴ Aquest mètode relaciona les diferències d'entalpia de formació entre conjunts de compostos RY i RH, on R és l'espècie organometàl·lica i Y és el grup unit a R a través d'un àtom X; amb el potencial covalent de l'àtom X, V_x . Aquest potencial, V_x , es defineix com la relació entre el nombre d'electrons de valència de l'àtom X i el seu radi. Aquest mètode ha estat aplicat amb èxit a conjunts d'alquilsilans^{75,76} però la seva aplicació a un conjunt més complet de compostos organometàl·lics no condueix a les relacions lineals esperades enfront del potencial covalent.⁷⁷

Una altra manera senzilla d'estimar els valors de $\Delta_f H^\circ$ per a compostos organometàl·lics és l'anomenat "Mètode de Tina" desenvolupat pel grup de Simões.⁷⁸ Aquest mètode consisteix a establir relacions lineals entre l'entalpia de formació de diferents famílies de compostos ML_n , on M és un àtom metàl·lic o una espècie organometàl·lica i L és un lligand; amb les entalpies de formació de LH. Donat que les entalpies de formació dels compostos orgànics LH es poden conèixer amb exactitud, les correlacions establertes es poden emprar per trobar valors de $\Delta_f H^\circ$ desconeguts o, fins i tot, per avaluar valors experimentals trobats per a compostos organometàl·lics.

En l'actualitat, el desenvolupament de la química quàntica ha permès un important augment de l'exactitud en la predicció de dades termodinàmiques. Tot i així, l'aplicació d'aquests mètodes a compostos organometàl·lics es troba restringida per la capacitat computacional i la manca de dades experimentals fiables. Arribats a aquest punt, es fa necessari el desenvolupament de nous mètodes que permetin l'estimació d'entalpies de formació a partir de paràmetres senzills. Una possibilitat és, òbviament, la metodologia QSPR, que és capaç d'emprar descriptors numèrics, extrets de l'estructura molecular, per establir models que permeten l'estimació de propietats experimentals.⁷⁹ En el cas de propietats termodinàmiques, la metodologia QSPR ha estat utilitzada per predir les energies de Gibbs de compostos orgànics amb descriptors topològics basats en la teoria de grafs.^{80,81} Més recentment, s'han establert models lineals i no lineals, amb sis descriptors, per predir l'energia de Gibbs d'un conjunt de 177 compostos orgànics.⁸²

En aquest apartat es mostra l'aplicació de la metodologia QSPR per establir un model lineal que permeti la predicció de l'entalpia de formació en fase gas de compostos organometàl·lics. S'espera que els descriptors continguts pel model puguin ser fàcilment calculables per poder, així, establir una manera senzilla d'estimar aquesta propietat. Per comprovar la validesa d'aquesta metodologia, s'ha realitzat també un estudi amb compostos orgànics de naturalesa semblant als organometàl·lics, per tal de corroborar si els resultats que s'obtenen amb la metodologia QSPR són comparables a aquells que donen els mètodes additius per a aquest tipus de compostos.

Conjunts de dades

Els valors de $\Delta_f H^\circ$ en fase gas dels compostos organometàl·lics han estat extrets, majoritàriament, de l'*Organometallic Thermochemistry Database* del NIST⁸³ i el llibre de Cox i Pilcher.⁸⁴ En total, s'ha construït una base de dades que conté 132 valors d'entalpia de formació de compostos organometàl·lics: 5 d'alumini, 2 d'arsènic, 17 de bor, 3 de bismut, 5 de gal·li, 12 de germani, 22 de mercuri, 2 de fòsfor, 7 de plom, 5 d'antimoni, 6 de seleni, 6 de silici, 31 d'estany i 8 de tel·luri. Aquests compostos que formen el conjunt estudiat tenen la fórmula general: MR_nX_m , on R pot ser un grup alquil, aril, alquenil, alquinil i metal·laciicloalquil de cinc o sis baules, i X pot ser Cl, Br,

I i H. Els valors de $\Delta_f H^\circ$, que es troben en la Taula 3.40, van des de -514 kJ/mol del B(octil)₃ fins a 736.9 kJ/mol per al SnPh₃(C≡Ph), amb un valor mitjà de -23.21 kJ/mol.

La divisió del conjunt de dades en *training* i *prediction sets* proporciona subconjunts de 104 i 32 valors experimentals de $\Delta_f H^\circ$, respectivament. Cal assegurar que aquesta selecció de compostos proporciona subconjunts de dades que continguin tots els metalls i els substituents presents en l'estudi. Els compostos del *tset* s'utilitzaran per construir el model lineal que permet calcular l'entalpia de formació, la capacitat predictiva d'aquest model es comprovarà mitjançant la predicció de aquesta propietat pels compostos del *pset*.

En aquest cas, l'optimització de geometria dels compostos implicats en l'estudi s'ha realitzat principalment amb el mètode PM3; els compostos de bor s'han tractat amb el mètode AM1 ja que aquest element no es troba parametritzat pel mètode anterior.

El conjunt de 168 compostos orgànics triats per comprovar la bondat de la metodologia QSPR es compon de diferents alcans, haloalcans, alquilbenzens, halobenzens, alquens, alquins, cicloalcans i cicloalquens; és a dir, són compostos molt semblants als grups orgànics que es troben en el conjunt de compostos organometàl·lics. Els valors d'entalpia de formació per a aquests compostos, que s'han extret del llibre de Pedley⁷²; es troben recollits en la Taula 3.41, van des de -238.9 kJ/mol per a l'1-clorooctà fins a 251.7 kJ/mol per a l'1,2-diiodobenzè, amb un valor mitjà de -80.4 kJ/mol. De la manera habitual, s'ha dividit el conjunt de dades en *tset* i *pset*, obtenint dos subconjunts que contenen 133 i 35 valors experimentals de $\Delta_f H^\circ$, respectivament.

Compostos organometàl·lics

El nombre de descriptors moleculars calculats pel programa CODESSA per a aquest conjunt de compostos és de 1322. A aquest conjunt de descriptors s'hi van afegir els 31 descriptors externs propis dels metalls, extrets de la compilació d'Emsley,⁸⁵ Taula 3.42, conformant un total de 1353 descriptors.

Taula 3.42. Descriptors externs dels metalls.

Descriptor	Definició
Z	Nombre atòmic
M	Pes atòmic
PI	Potencial d'ionització
PI^{n+}	Potencial d'ionització iònic
AE	Afinitat electrònica
AE^{n+}	Afinitat electrònica iònica
χ_P	Electronegativitat de Pauling
χ_M	Electronegativitat de Mulliken
χ_S	Electronegativitat de Sanderson
χ_{A-R}	Electronegativitat d'Allred-Rochow
q^+	Càrrega iònica
e^-	Electrons de valència
e^-_{val}	Electrons de valència iònica
R_{cov}	Radi covalent
R_{at}	Radi atòmic
R_{ion}	Radi iònic
R_{Shan}	Radi iònic de Shannon
q^+/R_{ion}	Relació càrrega/radi iònic
q^+/R_{Shan}	Relació càrrega/radi iònic de Shannon
η	Duresa
η^{n+}	Duresa iònica
S	Tovesa
S^{n+}	Tovesa iònica
Z^*_S	Càrrega nuclear efectiva de Slater
Z^*_C	Càrrega nuclear efectiva de Clementi
Z^*_{FF}	Càrrega nuclear efectiva de Froesche-Fischer
d	Densitat
T_f	Temperatura de fusió
T_{eb}	Temperatura d'ebullició
V	Volum
$\Delta_{vap}H^\circ$	Entalpia de vaporització molar

Els descriptors inclosos en aquesta taula recullen la informació que caracteritza els metalls, dintre d'aquests descriptors es poden trobar alguns que contenen la informació sobre el metall en estat elemental i d'altres, els que s'han anomenat iònics, contenen la informació sobre el metall en l'estat d'oxidació formal que presenta dins de cada compost. En principi, sembla que aquests descriptors no haurien de ser necessaris ja que la naturalesa del metall ja es troba recollida en els descriptors moleculars calculats; en la pràctica, com es veurà més endavant, aquests descriptors són de vital importància. El mètode escollit per establir els models lineals és l'heurístic de CODESSA, aquest procés comporta alhora una reducció dràstica en el nombre de descriptor passant de 1353 a 172, 28 dels quals corresponen a descriptors externs dels metalls.

Els primers models lineals que es van desenvolupar per calcular $\Delta_f H^\circ$ per al conjunt de compostos organometàl·lics no presentaven resultats gaire esperançadors, amb $R^2 \approx 0.95$ i valors de *rmse* propers a 60 kJ/mol. Cal dir que aquests models es van deduir només a partir dels descriptors moleculars calculats per CODESSA, en els quals la naturalesa del metall ja s'hi troba implícita, atesa l'optimització prèvia de la molècula. Els resultats obtinguts van suggerir que la inclusió dels descriptors propis dels metalls, com a descriptors externs, podria millorar de manera significativa les correlacions, i de fet així ha estat. L'addició d'aquests descriptors permet deduir uns models clarament superiors, amb coeficients de determinació per sobre de 0.99 i *rmse* per sota dels 30 kJ/mol. Aquests models contenen, però, només descriptors del tipus constitucional, que reflecteixen la composició química de la molècula, sense fer cap referència a l'estructura electrònica o geomètrica. Així doncs, aquests models tenen un problema bastant greu: no són capaços de distingir entre els isòmers de cadena d'un mateix metall, entre propil- i isopropil-derivats, per exemple. Analitzant els descriptors involucrats en tots els models lineals, tant els deduïts de la manera habitual pel mètode heurístic com aquells derivats amb el mètode de la millor regressió multilíneal, també implementat a CODESSA, es va decidir incloure un descriptor topològic com a solució a aquest problema: l'índex de Kier & Hall d'ordre 0. Així, finalment es proposa el model de nou descriptors, obtingut per regressió multilíneal, que es mostra en la Taula 3.43.

Taula 3.43. Model obtingut per regressió lineal.

Descriptor	coeficient	sd	test-t	β
Terme independent	217.37	32.08	6.78	-
Pes molecular	1.61	0.11	14.82	0.64
Nombre d'enllaços senzills	-6.36	0.76	-8.31	-0.32
Nombre d'enllaços dobles	72.17	23.13	3.12	0.04
Nombre d'enllaços triples	169.81	14.28	11.89	0.15
Nombre d'enllaços aromàtics	20.57	0.96	21.43	0.56
Nombre d'halògens	-196.06	7.02	-27.91	-0.54
Índex de Kier & Hall d'ordre 0	-31.04	4.68	-6.63	-0.42
Electronegativitat de Mulliken	-34.92	5.17	-6.75	-0.09
Relació càrrega/radi iònic	-6.53	1.29	-5.06	-0.10

La correlació entre els descriptors que formen el model es troba en el rang 0.01-0.85, amb un valor mitjà de 0.20. D'acord amb els coeficients de regressió estandarditzats β , els descriptors més importants són el pes molecular, el nombre d'enllaços aromàtics i el nombre d'halògens. Els resultats obtinguts amb aquest model es troben en la Taula 3.44. El valor de R^2_{cv} obtingut és 0.983, molt semblant a l'obtingut per al *tset* i indicatiu de l'estabilitat del model. Els valors de $\Delta_f H^\circ$ calculats amb aquest model es troben en la Taula 3.40 i la seva representació gràfica enfront dels valors experimentals es pot veure en la Figura 3.17.

Taula 3.44. Paràmetres estadístics obtinguts amb el model lineal.

Conjunt	n	R^2	F	rmse
<i>tset</i>	104	0.988	8224.3	29.17
<i>pset</i>	32	0.990	2459.9	29.60

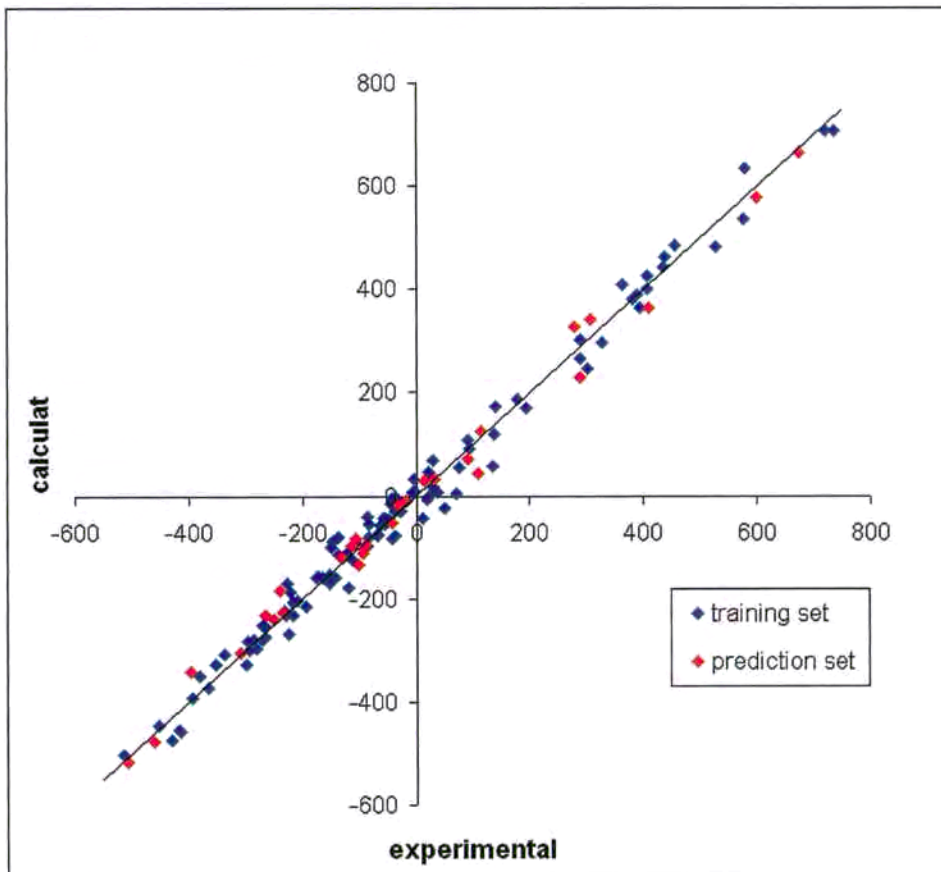


Figura 3.17. Representació gràfica de $\Delta_f H^\circ$ calculat vs. experimental amb el model lineal.

Dels nou descriptors que conté el model, sis són constitucionals: pes molecular, nombre d'enllaços senzills, dobles, triples i aromàtics, i nombre d'halògens. Tots aquests descriptors presenten signe positiu dins de la correlació excepte el nombre d'enllaços senzills i el nombre d'halògens. Els descriptors topològics donen informació sobre la connectivitat atòmica d'una molècula. En particular, el descriptor índex de Kier & Hall d'ordre 0, que pertany als ben coneguts índexs de connectivitat de valència "χ" de Kier i Hall,⁴³ conté informació relacionada amb la presència d'heteroàtoms i amb la hibridació dels àtoms en la molècula. Aquest descriptor, ${}^0\chi$ es calcula de la següent manera:

$${}^0\chi = \sum_i^n \left(\frac{1}{\delta_i} \right)^{1/2}$$

on n és el nombre total d'àtoms de la molècula (sense comptar hidrògens) i δ_i és la connectivitat atòmica d'ordre zero de cada àtom. Aquesta connectivitat atòmica s'obté:

$$\delta_i^v = \frac{Z_i^v - H_i}{Z_i - Z_i^v - 1}$$

on Z_i és el nombre total d'electrons de l'àtom i , Z_i^v és el nombre d'electrons de valència d'aquest mateix àtom i H_i és el nombre d'àtoms d'hidrogen units directament a l'àtom i . A títol d'exemple, es mostra el càlcul d'aquest descriptor per a dos isòmers de cadena diferents: $B(\text{Bu})_3$ i $B(\text{t-Bu})_3$. Aquest descriptor es calcula de la següent manera per a cadascuna d'aquestes molècules:

$${}^0\chi_{B\text{Bu}_3} = \left(\frac{1}{\delta_B^v} \right)^{1/2} + 3 \left[3 \left(\frac{1}{\delta_{\text{CH}_2}^v} \right)^{1/2} + \left(\frac{1}{\delta_{\text{CH}_3}^v} \right)^{1/2} \right]$$

$${}^0\chi_{B(\text{t-Bu})_3} = \left(\frac{1}{\delta_B^v} \right)^{1/2} + 3 \left[\left(\frac{1}{\delta_{\text{CH}}^v} \right)^{1/2} + \left(\frac{1}{\delta_{\text{CH}_2}^v} \right)^{1/2} + 2 \left(\frac{1}{\delta_{\text{CH}_3}^v} \right)^{1/2} \right]$$

La contribució de l'àtom de bor, que es manté constant, és:

$$\delta_B^v = \frac{(3-0)}{(5-3-1)} = 3$$

La diferència principal entre aquests compostos es deu als termes del substituent, que es calculen:

$$\delta_{CH}^v = \frac{(4-1)}{(6-4-1)} = 3 \quad \delta_{CH_2}^v = \frac{(4-2)}{(6-4-1)} = 2 \quad \delta_{CH_3}^v = \frac{(4-3)}{(6-4-1)} = 1$$

Donat que es coneixen totes les connectivitats atòmiques dintre de les molècules, el descriptor ${}^0\chi$ pren valors de 9.9413 i 10.4307, per a $B(Bu)_3$ i $B^t(Bu)_3$, respectivament. Aquest descriptor presenta signe negatiu dintre de la correlació.

El dos últims descriptors del model, que pertanyen al metall, són l'electronegativitat de Mulliken i la relació càrrega/radi iònic. L'electronegativitat de Mulliken es calcula com la mitjà del potencial d'ionització i l'afinitat electrònica de l'element. La càrrega de cada metall es correspon amb el seu estat d'oxidació formal en el compost organometàl·lic i el radi iònic és el del catió en aquest mateix estat d'oxidació. El valor dels radis iònics ha estat extret del llibre d'Emsley.⁸⁵ Aquests dos descriptors estan clarament relacionats a la naturalesa química dels enllaços M-C i M-X i, en conseqüència, a la seva energia d'enllaç. És conegut que la diferència d'electronegativitat entre els dos àtoms d'un enllaç es veu reflectida en la seva energia de dissociació. D'altra banda, la relació càrrega/radi d'un catió és una mesura del seu poder polaritzant i, conseqüentment, del grau de covalència de l'enllaç que es forma i de la seva energia de dissociació.

Els resultats obtinguts analitzant els subconjunts de compostos que contenen halògens o hidrogen com a àtoms terminals i els que no presenten aquestes característiques, són també bons, Taula 3.45.

Taula 3.45. Paràmetres estadístics obtinguts per subconjunts de compostos.

Conjunt	n	R ²	F	rmse
Halogenats o H terminal	40	0.991	4114.6	21.4
No halogenats o H terminal	92	0.987	7045.2	30.8

També es pot comprovar que les diferències entre valors experimentals i calculats per a compostos isomèrics de cadena són anàlogues, tal com es mostra a la Taula 3.46.

 Taula 3.46. Valors de $\Delta_f H^\circ$ per a compostos isomèrics.

Compost	experimental	calculat	residual
AlBu ₃	-216.0	-207.2	-8.8
Al'Bu ₃	-231.5	-222.4	-9.1
BBu ₃	-287.0	-280.0	-7.0
B'Bu ₃	-280.3	-295.2	14.9
BPr ₃	-236.0	-224.7	-11.3
B'Pr ₃	-251.4	-239.9	-11.5
BOct ₃	-514.6	-501.2	-13.4
B'Oct ₃	-507.9	-516.4	8.5
GaBu ₃	-226.9	-169.1	-57.8
Ga'Bu ₃	-239.5	-184.3	-55.2
HgBu ₂	-32.4	-18.2	-14.2
Hg'Bu ₂	-38.3	-28.3	-10.0
HgPr ₂	30.3	18.7	11.6
Hg'Pr ₂	37.0	8.6	28.4
HgBrPr	-51.4	-46.9	-4.5
HgBr'Pr	-53.8	-52.0	-1.8
HgClPr	-88.1	-92.7	4.6
HgCl'Pr	-86.9	-97.8	10.9
HgIPr	-3.7	11.0	-14.7
HgI'Pr	-5.8	6.0	-11.8
SnBu ₄	-217.4	-231.3	13.9
Sn'Bu ₄	-272.2	-251.6	-20.6
SnPr ₄	-142.9	-157.6	14.7
Sn'Pr ₄	-119.4	-177.8	58.4
SnMe ₃ Pr	-46.8	-47.0	0.2
SnMe ₃ 'Pr	-43.7	-52.1	8.4
TePent ₂	-137.0	-78.9	-58.1
Te'Pent ₂	-148.0	-89.1	-58.9
TePr ₂	-42.5	-5.2	-37.3
Te'Pr ₂	-46.0	-15.3	-30.7

Així doncs, el model proposat sembla ser un bon mètode per estimar l'entalpia de formació de compostos organometàl·lics dels elements i grups orgànics analitzats. Aquest model es podria utilitzar també, sense perdre de vista les limitacions que presenta, com a mètode per avaluar la qualitat de les mesures experimentals. Per exemple, es coneixen alguns valors experimentals de $\Delta_f H^\circ$ de diferents derivats alquílics MR_2 de zinc i cadmi.⁸³ Alguns d'aquests valors presenten uns errors experimentals elevats: $ZnMe_2$ (52.9 ± 1.3), $ZnEt_2$ (55.5 ± 3.9), $ZnPr_2$ (-17.2 ± 23.0), $ZnBu_2$ (-55.2 ± 23.4), $CdMe_2$ (104.8 ± 1.3) i $CdEt_2$ (105.3 ± 2.0). Tots els intents d'incloure aquests compostos dintre del model desenvolupat produeixen augments de *rmse* d'uns 10 kJ/mol, acostant l'error del *tset* als 40 kJ/mol. Per aquesta raó, pot dir-se que la inclusió d'aquests metalls en el model, requereix altres valors experimentals més acurats per a compostos que continguin aquests elements.

Compostos orgànics

Per tal de comprovar la capacitat de la metodologia QSPR en l'estimació d'entalpies de formació, s'ha construït un model lineal similar per a un conjunt de 168 compostos orgànics. Cal dir que els valors de $\Delta_f H^\circ$ per als compostos orgànics són més acurats que els que s'han utilitzat en l'estudi dels compostos organometàl·lics. Els compostos que conformen aquest conjunt de dades, que són semblants als substituents orgànics que presenten els compostos organometàl·lics analitzats, són alcans, haloalcans, alquilbenzens, halobenzens, alquens, alquins, cicloalcans i cicloalquens. Es va seguir la mateixa metodologia de treball que amb el conjunt de compostos organometàl·lics però, en aquest cas, com a descriptors moleculars només es van utilitzar els de les classes, constitucionals, topològics i geomètrics.

Utilitzant la regressió multilíneal que implementa el mètode heurístic de CODESSA s'obté el model que es mostra en la Taula 3.47 per als 133 compostos que formen el *tset*.

Taula 3.47. Model obtingut per regressió lineal.

Descriptor	coeficient	sd	test-t	β
Terme independent	162.40	14.43	11.25	-
Nombre d'enllaços senzills	-67.82	1.50	-45.25	-3.90
Nombre d'enllaços dobles	-51.10	3.06	-16.68	-0.21
Nombre d'anells de benzè	-504.55	12.59	-40.07	-1.90
Nombre d'anells	302.93	10.03	30.19	1.33
Nombre relatiu d'àtoms d'hidrogen	-193.38	21.26	-9.10	-0.21
Nombre d'halògens	-280.42	4.76	-58.94	-2.56
Índex de forma de Kier d'ordre 1	177.67	5.59	31.78	2.81
Índex de Randic d'ordre 2	-12.45	1.56	-7.96	-0.12
Índex de gravitació per a enllaços	0.11	0.01	8.33	0.21

La correlació entre els descriptors que formen el model es troba en el rang 0.02-0.76, amb un valor mitjà de 0.36. D'acord amb els coeficients de regressió estandarditzats β , els descriptors més importants són el pes molecular, el nombre d'enllaços senzills, el nombre d'halògens i l'índex de forma de Kier d'ordre 1. Els resultats obtinguts amb aquest model per als *tset* i *pset* es troben en la Taula 3.48. En aquesta ocasió, el valor de $R^2_{cv} = 0.995$, confirma també l'estabilitat del model. Els valors de $\Delta_f H^\circ$ experimentals i calculats amb aquest model es troben en la Taula 3.41, la representació gràfica d'aquests valors es mostra en la Figura 3.18.

Taula 3.48. Paràmetres estadístics obtinguts amb el model lineal.

Conjunt	n	R^2	F	rmse
<i>tset</i>	133	0.996	33564.1	6.78
<i>pset</i>	35	0.995	6653.1	8.11

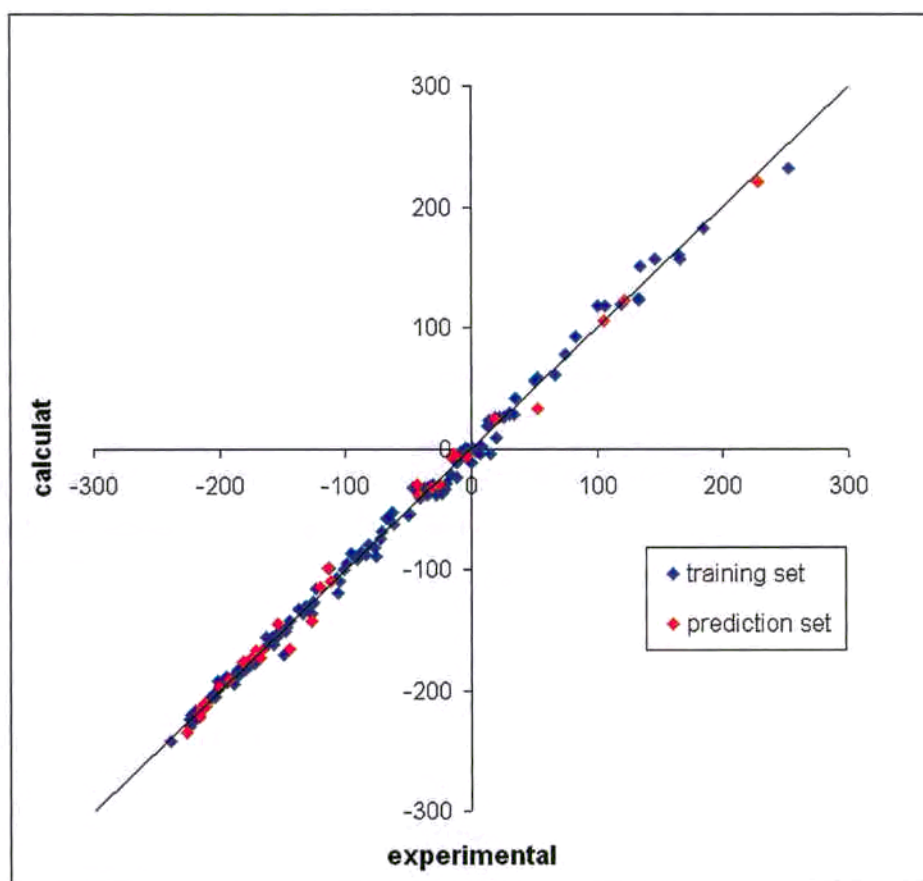


Figura 3.18. Representació gràfica de $\Delta_r H^\circ$ calculat vs. experimental amb el model lineal.

El model obtingut presenta sis descriptors constitucionals: nombre d'enllaços senzills i dobles, nombre d'anells i d'anells de benzè, nombre d'halògens i nombre relatiu d'àtoms d'hidrogen; dos topològics: l'índex de forma de Kier d'ordre 1 i l'índex de Randic d'ordre 2; i un descriptor geomètric: l'índex de gravitació per a enllaços. Els índexs de forma de Kier mesuren la forma de la molècula en base al nombre d'àtoms present en l'esquelet i a la seva ramificació.⁸⁶ Els índexs de connectivitat de Randic, que contenen també informació sobre la grandària i la ramificació de l'estructura, es calculen en base a la diferent coordinació de cada àtom en la molècula.⁴⁹

El descriptor geomètric índex de gravitació per a enllaços, que és un reflex de la forma de la molècula i de la distribució de la massa en el seu interior,⁸⁷ es calcula d'acord amb la llei de Newton:

$$G^b = \sum_{(i<j)}^{N_b} \frac{m_i m_j}{r_{ij}^2}$$

on m_i i m_j són els pesos atòmics dels àtom i i j , r_{ij} és la distància interatòmica entre aquests dos àtoms i N_b és el nombre d'enllaços de la molècula.

Els resultats obtinguts amb aquesta aproximació es poden comparar amb els deduïts mitjançant el mètode de Pedley, Taula 3.49. Els valors calculats de $\Delta_f H^\circ$ amb aquest mètode additiu es troben en la Taula 3.41.

Taula 3.49. Paràmetres estadístics obtinguts pel mètode de Pedley i el model lineal.

Conjunt	n	R ²	F	rmse
Pedley	168	0.998	81923.4	4.97
Model lineal	168	0.996	40475.7	7.01

Tots dos mètodes donen resultats semblants, la qual cosa els fa equivalents a efectes de càlcul per al conjunt de compostos tractat, encara que el mètode de Pedley dona uns errors més baixos. Si s'analitzen els valors que s'obtenen amb els dos mètodes per a dos subconjunts de compostos anàlegs: els isòmers dels alcans de fórmula C_8H_{18} i els policlorobenzens, Taula 3.50, es comprova que les diferències existents entre els dos mètodes de càlcul són molt petites, però que el mètode de Pedley no distingeix entre alguns isòmers dels policlorobenzens.

Taula 3.50. Valors de $\Delta_f H^\circ$ obtinguts per als diferents mètodes per a subconjunts de compostos.

Compost	Fórmula	Exp.	QSPR	Pedley
2,2,3,3-Tetrametilbutà	C ₈ H ₁₈	-226.0	-234.3	-224.1
2,2,3-Trimetilpentà	C ₈ H ₁₈	-219.9	-223.7	-222.5
2,2,4-Trimetilpentà	C ₈ H ₁₈	-223.9	-229.6	-223.7
2,3,3-Trimetilpentà	C ₈ H ₁₈	-216.2	-221.4	-219.5
2,3,4-Trimetilpentà	C ₈ H ₁₈	-217.2	-219.6	-219.6
3-Etil-3-metilpentà	C ₈ H ₁₈	-214.8	-213.5	-215.3
2,2-Dimetilhexà	C ₈ H ₁₈	-224.5	-223.3	-225.3
2,3-Dimetilhexà	C ₈ H ₁₈	-213.8	-215.1	-216.9
2,4-Dimetilhexà	C ₈ H ₁₈	-219.2	-216.8	-218.9
2,5-Dimetilhexà	C ₈ H ₁₈	-222.5	-219.4	-221.9
3,3-Dimetilhexà	C ₈ H ₁₈	-219.9	-218.4	-220.3
3,4-Dimetilhexà	C ₈ H ₁₈	-212.8	-212.2	-214.9
3-Etilhexà	C ₈ H ₁₈	-210.7	-208.3	-211.2
2-Metilheptà	C ₈ H ₁₈	-215.3	-213.4	-215.2
3-Metilheptà	C ₈ H ₁₈	-212.5	-210.5	-213.2
4-Metilheptà	C ₈ H ₁₈	-211.9	-210.8	-213.2
3-Etil-2-metilpentà	C ₈ H ₁₈	-211.0	-212.8	-214.9
<i>n</i> -octà	C ₈ H ₁₈	-208.5	-207.3	-208.5
1,2,3,4-Tetraclorobenzè	C ₆ H ₂ Cl ₄	-25.4	-29.9	-39.8
1,2,3,5-Tetraclorobenzè	C ₆ H ₂ Cl ₄	-34.9	-31.6	-39.8
1,2,4,5-Tetraclorobenzè	C ₆ H ₂ Cl ₄	-32.6	-31.5	-39.7
1,2,3-Triclorobenzè	C ₆ H ₃ Cl ₃	3.8	-1.4	-9.1
1,2,4-Triclorobenzè	C ₆ H ₃ Cl ₃	-8.1	-3.0	-9.2
1,3,5-Triclorobenzè	C ₆ H ₃ Cl ₃	-13.4	-4.9	-9.2
1,2-Diclorobenzè	C ₆ H ₄ Cl ₂	30.2	28.0	21.4
1,3-Diclorobenzè	C ₆ H ₄ Cl ₂	25.7	26.2	21.4
1,4-Diclorobenzè	C ₆ H ₄ Cl ₂	22.5	26.3	18.2

La Taula 3.51, recull els resultats globals d'aquests subconjunts i es pot veure que el mètode lineal presenta, en relació al mètode de Pedley, un resultat una mica millors amb els policlorobenzens i una mica pitjors amb els isòmers C₈H₁₈.

Taula 3.51. Paràmetres estadístics obtinguts amb els subconjunts d'isòmers.

Conjunt	Mètode	n	R ²	F	rmse
Isòmers C ₈ H ₁₈	Pedley	18	0.914	169.0	1.44
	Model lineal	18	0.829	77.4	3.10
Policlorobenzens	Pedley	9	0.952	137.8	6.12
	Model lineal	9	0.968	211.1	4.80
Tots	Pedley	27	0.999	21842.1	3.49
	Model lineal	27	0.999	20108.7	3.74

En resum, pot concloure's que els resultats que s'obtenen amb el model lineal per als compostos organometàl·lics són prou bons. Els descriptors que el formen són fàcils i ràpids d'obtenir, la qual cosa fa que el càlcul de la propietat estudiada esdevingui, també, molt senzill. Per tal de mantenir aquesta facilitat en el càlcul, que es pot fer amb una calculadora, no s'han volgut aplicar mètodes no lineals de tipus híbrid. Amb aquests mètodes, probablement, s'obtindrien nivells d'error més baixos però, alhora, es perdria la senzillesa i generalitat del mètode.

S'ha pogut comprovar que la metodologia QSPR és adequada per estimar entalpies de formació, ja que permet deduir un model lineal amb descriptors semblants útil per a compostos orgànics anàlegs. Els resultats que s'obtenen amb aquest model lineal són de la mateixa bondat que els trobats amb el mètode additiu de Pedley, model àmpliament emprat i acceptat en el camp de la Química orgànica.

Taula 3.40. Valors experimentals i calculats de $\Delta_f H^\circ$ per als compostos organometàl·lics (kJ/mol).

#	Compost	Experimental	QSPR
1	Al ⁱ Bu ₃ ^a	-231.5 ^c	-222.4
2	AlBu ₃ ^a	-216.0 ^d	-207.2
3	AlEt ₃ ^a	-114.1 ^c	-96.6
4	AlMe ₃ ^b	-86.5 ^c	-41.3
5	AlPr ₃ ^a	-153.3 ^d	-151.9
6	AsMe ₃ ^a	12.5 ^e	-44.7
7	AsPh ₃ ^b	408.4 ^c	365.1
8	B(3-Me-Bu) ₃ ^a	-381.8 ^c	-350.5
9	B(4-tol)Cl ₂ ^a	-294.4 ^e	-299.2
10	B ⁱ Bu ₃ ^a	-280.3 ^e	-295.2
11	B ⁱ Pr ₃ ^b	-251.4 ^c	-239.9
12	B ^o Oct ₃ ^a	-507.9 ^e	-516.4
13	BBu ₂ Br ^a	-300.4 ^c	-327.2
14	BBu ₂ Cl ^a	-365.7 ^c	-373.0
15	BBu ₂ I ^a	-225.5 ^e	-269.3
16	BBu ₃ ^a	-287.0 ^e	-280.0
17	BCy ₃ ^b	-397.9 ^e	-341.9
18	BEt ₃ ^a	-152.7 ^c	-169.4
19	BHex ₃ ^a	-395.0 ^e	-390.6
20	BHpt ₃ ^a	-454.8 ^e	-445.9
21	BMe ₃ ^a	-122.6 ^e	-114.1
22	BOct ₃ ^b	-514.6 ^e	-501.2
23	BPhCl ₂ ^a	-266.1 ^e	-274.1
24	BPr ₃ ^b	-236.0 ^e	-224.7
25	BiBu ₃ ^a	71.2 ^d	3.3
26	BiMe ₃ ^a	194.4 ^c	169.2
27	BiPh ₃ ^b	600.6 ^c	579.0
28	BiPr ₃ ^a	133.9 ^d	58.6
29	Ga ⁱ Bu ₃ ^b	-239.5 ^c	-184.3
30	GaBu ₃ ^a	-226.9 ^c	-169.1
31	GaEt ₃ ^a	-62.3 ^c	-58.5
32	GaMe ₃ ^a	-35.9 ^c	-3.2
33	GaPr ₃ ^a	-125.5 ^d	-113.8
34	GeBu ₄ ^b	-310.0 ^d	-304.8
35	GeBz ₄ ^a	388.2 ^c	389.1
36	GeEt ₃ H ^a	-111.9 ^c	-123.8
37	GeEt ₄ ^a	-164.9 ^c	-157.3
38	GeMe ₃ Br ^a	-222.2 ^c	-186.1
39	GeMe ₃ Cl ^b	-266.1 ^c	-231.9
40	GeMe ₄ ^b	-107.5 ^c	-83.6
41	GePh ₂ (CH ₂) ₄ ^a	138.9 ^c	174.0
42	GePh ₃ (C≡CPh) ^a	578.9 ^c	633.9
43	GePh ₃ (CH=CH ₂) ^a	362.6 ^c	408.9
44	GePh ₄ ^a	438.6 ^c	462.8
45	GePr ₄ ^a	-229.7 ^e	-231.0
46	Hg(C≡CPh) ₂ ^b	720.5 ^c	707.8
47	Hg(CH ₂ - ⁱ Bu) ₂ ^a	-82.7 ^c	-78.6
48	Hg ⁱ Bu ₂ ^a	-38.3 ^c	-28.3

#	Compost	Experimental	QSPR
49	Hg'Pr ₂ ^a	37.0 ^c	8.6
50	HgBr'Pr ^a	-53.8 ^c	-54.0
51	HgBrEt ^a	-30.3 ^c	-28.5
52	HgBrMe ^b	-18.6 ^c	-10.1
53	HgBrPr ^a	-51.4 ^c	-46.9
54	HgBu ₂ ^b	-32.4 ^e	-18.2
55	HgBz ₂ ^b	278.0 ^c	328.7
56	HgCl'Pr ^a	-86.9 ^c	-97.8
57	HgClEt ^a	-67.7 ^c	-74.3
58	HgClMe ^a	-55.1 ^c	-42.1
59	HgClPr ^a	-88.1 ^c	-92.7
60	HgEt ₂ ^a	75.0 ^c	55.5
61	HgI'Pr ^a	-5.8 ^c	6.0
62	HgIEt ^b	14.3 ^c	29.5
63	HgIMe ^a	21.6 ^c	47.9
64	HgIPr ^a	-3.7 ^c	11.0
65	HgMe ₂ ^a	92.4 ^c	92.4
66	HgPh ₂ ^a	394.2 ^c	365.6
67	HgPr ₂ ^a	30.3 ^c	18.7
68	PbEt ₄ ^b	109.6 ^c	45.1
69	PbMe ₄ ^b	136.1 ^c	118.9
70	PbPh ₂ Br ₂ ^a	177.7 ^c	187.1
71	PbPh ₂ I ₂ ^a	288.2 ^c	303.0
72	PbPh ₃ Br ^a	406.3 ^c	426.2
73	PbPh ₃ I ^a	455.4 ^c	484.1
74	PbPh ₄ ^a	674.1 ^c	665.2
75	PMe ₃ ^b	-94.1 ^c	-111.6
76	PPh ₃ ^a	328.4 ^c	298.2
77	SbBu ₃ ^b	-100.5 ^d	-133.4
78	SbEt ₃ ^a	48.7 ^c	-22.8
79	SbMe ₃ ^b	32.2 ^c	32.5
80	SbPh ₃ ^a	435.4 ^c	442.3
81	SbPr ₃ ^a	-37.7 ^d	-78.1
82	Se'Pr ₂ ^a	-108 ^c	-91.7
83	SeBu ₂ ^b	-131.8 ^c	-118.5
84	SeEt ₂ ^a	-51.8 ^c	-52.0
85	SeMe ₂ ^a	17.8 ^c	-7.9
86	SePent ₂ ^a	-172.7 ^c	-155.3
87	SePh ₂ ^a	289.7 ^c	265.3
88	SiEt ₄ ^a	-265.7 ^d	-253.7
89	SiMe ₂ Cl ₂ ^b	-461.1 ^c	-476.5
90	SiMe ₃ Br ^a	-297.5 ^e	-282.4
91	SiMe ₃ Cl ^a	-354.0 ^e	-328.2
92	SiMeCl ₂ H ^a	-415.0 ^e	-456.2
93	SiPh ₂ Cl ₂ ^a	-208.8 ^e	-203.3
94	Sn'Bu ₄ ^a	-272.2 ^c	-251.6
95	Sn'Pr ₄ ^b	-119.4 ^c	-177.8
96	SnBu ₃ Br ^a	-270.6 ^c	-278.5
97	SnBu ₄ ^a	-217.4 ^c	-231.3
98	SnEt ₃ Cl ^a	-193.3 ^d	-213.7
99	SnEt ₃ H ^a	-54.7 ^c	-55.9

#	Compost	Experimental	QSPR
100	SnEt ₄ ^a	-42.0 ^c	-83.8
101	SnEtCl ₃ ^a	-429.3 ^d	-473.4
102	SnEtPh ₃ ^a	380.0 ^f	381.2
103	SnMe ₂ Cl ₂ ^a	-337.0 ^g	-306.7
104	SnMe ₂ I ₂ ^a	-150.0 ^g	-99.2
105	SnMe ₃ (CH=CH ₂) ^b	91.5 ^c	72.5
106	SnMe ₃ ^f Pr ^b	-43.7 ^c	-52.1
107	SnMe ₃ ^f Bu ^a	-67.0 ^d	-77.2
108	SnMe ₃ Br ^a	-138.1 ^c	-112.6
109	SnMe ₃ Bz ^a	90.4 ^e	108.0
110	SnMe ₃ Cl ^a	-174.9 ^d	-158.4
111	SnMe ₃ Et ^a	-26.3 ^c	-28.6
112	SnMe ₃ H ^a	24.9 ^c	1.3
113	SnMe ₃ I ^a	-82.4 ^c	-54.7
114	SnMe ₃ Ph ^a	113.6 ^e	126.5
115	SnMe ₃ Pr ^a	-46.8 ^d	-47.0
116	SnMe ₄ ^a	-17.6 ^d	-10.1
117	SnMeCl ₃ ^a	-417.1 ^d	-454.9
118	SnMePh ₃ ^a	406.0 ^f	399.6
119	SnPh ₂ (CH ₂) ₄ ^a	301.8 ^c	247.5
120	SnPh ₂ (CH ₂) ₅ ^b	289.0 ^c	229.1
121	SnPh ₃ (C≡CPh) ^a	736.9 ^c	707.3
122	SnPh ₃ (CH=CH ₂) ^a	528.6 ^c	482.3
123	SnPh ₄ ^a	575.4 ^c	536.2
124	SnPr ₄ ^a	-142.9 ^c	-157.6
125	Te ^f Pent ₂ ^a	-148.0 ^c	-89.1
126	Te ^f Pr ₂ ^a	-46.0 ^c	-15.3
127	TeBu ₂ ^a	-56.0 ^c	-44.7
128	TeEt ₂ ^a	-4.4 ^c	31.7
129	TeMe ₂ ^a	28.6 ^c	68.5
130	TePent ₂ ^a	-137.0 ^c	-78.9
131	TePh ₂ ^b	306.8 ^c	341.7
132	TePr ₂ ^a	-42.5 ^c	-5.2

^a Training set. ^b Prediction set.

^c Referència 83.

^d Benson, S. W.; Francis, J. T.; Tsotsis, T. T. *J. Phys. Chem.* **1988**, 92, 4515.

^e Referència 84.

^f Chambers, D. B.; Glockling, F. *Inorg. Chim. Acta* **1970**, 4, 150.

^g De Ryck, P. H.; Verdock, L.; van der Kelen, G. P. *Bull. Soc. Chim. Belg.* **1985**, 94, 621.

Taula 3.41. Valors experimentals i calculats de $\Delta_f H^\circ$ per als compostos orgànics (kJ/mol).

#	Compost	Experimental	QSPR	Pedley
1	(+/-)-2-Bromobutà ^b	-120.3	-114.2	-118.2
2	(+/-)-2-Cloro-3-metilbutà ^a	-185.1	-187.1	-188.9
3	(+/-)-2-Clorobutà ^a	-161.1	-160.9	-164.4
4	(E)-2-Butè ^a	-11.4	-11.5	-11.4
5	(E)-2-Pentè ^b	-31.9	-32.5	-32.2
6	1,1,1-Tricloroetà ^b	-144.4	-164.9	-146.8
7	1,1,2,2-Tetracloroetà ^a	-149.3	-170.5	-179.5
8	1,1,2-Tricloroetà ^a	-151.3	-149.3	-161.2
9	1,1-Dicloroetà ^a	-127.7	-135.8	-135.2
10	1,1-Dimetilciclohexà ^a	-180.8	-176.8	-186.7
11	1,2,3,4-Tetraclorobenzè ^b	-25.4	-29.9	-39.8
12	1,2,3,5-Tetraclorobenzè ^a	-34.9	-31.6	-39.8
13	1,2,3-Triclorobenzè ^a	3.8	-1.4	-9.1
14	1,2,3-Tricloropropà ^a	-182.9	-182.8	-197.7
15	1,2,3-Trimetilbenzè ^a	-9.5	-3.0	-13.8
16	1,2,4,5-Tetraclorobenzè ^a	-32.6	-31.5	-39.7
17	1,2,4,5-Tetrametilbenzè ^a	-47.1	-32.0	-45.9
18	1,2,4-Triclorobenzè ^a	-8.1	-3.0	-9.2
19	1,2,4-Trimetilbenzè ^b	-13.9	-4.5	-13.8
20	1,2-Dibromo-2-metilpropà ^a	-113.3	-98.4	-109.0
21	1,2-Dibromobutà ^a	-91.6	-91.4	-97.7
22	1,2-Dibromociclohexà ^b	-114.1	-98.8	-112.7
23	1,2-Dibromoetà ^a	-37.5	-32.4	-46.9
24	1,2-Dibromoheptà ^a	-157.9	-162.2	-160.0
25	1,2-Dibromopropà ^a	-71.6	-68.2	-79.0
26	1,2-Diclorobenzè ^a	30.2	28.0	21.4
27	1,2-Dicloroetà ^a	-126.4	-126.5	-139.9
28	1,2-Dicloropropà ^a	-162.8	-162.1	-171.7
29	1,2-Diiodobenzè ^a	251.7	232.3	247.1
30	1,2-Diiodobutà ^a	12.3	19.2	18.2
31	1,2-Diiodoetà ^a	75.0	78.1	67.0
32	1,2-Diiodopropà ^a	35.6	42.3	37.0
33	1,2-Dimetilbenzè ^a	19.1	25.9	18.4
34	1,3,5-Triclorobenzè ^b	-13.4	-4.9	-9.2
35	1,3,5-Trimetilbenzè ^b	-15.9	-6.4	-13.7
36	1,3-Ciclohexadiè ^a	106.3	118.7	106.3
37	1,3-Ciclopentadiè ^a	134.3	151.0	134.3
38	1,3-Diclorobenzè ^a	25.7	26.2	21.4
39	1,3-Diclorobutà ^a	-195.0	-188.8	-192.4
40	1,3-Dicloropropà ^a	-159.2	-156.4	-160.7
41	1,3-Dimetilbenzè ^a	17.3	24.2	18.4
42	1,4-Dibromobutà ^a	-87.8	-88.2	-88.5
43	1,4-Diclorobenzè ^a	22.5	26.3	18.2
44	1,4-Diclorobutà ^a	-183.4	-182.5	-181.4
45	1,4-Dimetilbenzè ^b	18.1	24.4	18.4
46	1-Bromobutà ^a	-107.1	-107.8	-107.0
47	1-Bromoheptà ^b	-167.8	-172.8	-169.2
48	1-Bromohexà ^a	-148.3	-151.3	-148.5
49	1-Bromoetà ^a	-189.3	-194.2	-190.1

#	Compost	Experimental	QSPR	Pedley
50	1-Bromopentà ^a	-128.9	-129.7	-127.7
51	1-Bromopropà ^a	-87.0	-85.5	-86.2
52	1-Butè ^a	0.1	-11.7	-0.2
53	1-Butí ^a	165.2	156.4	165.3
54	1-Cloro-2-etylbenzè ^a	-6.7	-5.8	-0.8
55	1-Cloro-2-metilpropà ^a	-159.3	-160.6	-160.1
56	1-Cloro-3-metilbutà ^a	-179.7	-182.9	-180.9
57	1-Cloro-4-etylbenzè ^b	-3.6	-7.0	-0.9
58	1-Clorobutà ^a	-154.4	-154.9	-153.4
59	1-Clorooctà ^b	-238.9	-241.3	-236.5
60	1-Cloropentà ^a	-174.9	-176.8	-174.2
61	1-Cloropropà ^a	-131.9	-132.6	-132.6
62	1-Etil-2-metilbenzè ^a	1.3	0.8	-2.4
63	1-Etil-3-metilbenzè ^a	-1.8	-0.6	-2.4
64	1-Etil-4-metilbenzè ^a	-3.2	-0.5	-2.4
65	1-Etilciclohexè ^a	-63.4	-53.0	-63.7
66	1-Etilciclopentè ^a	-19.8	-28.0	-24.9
67	1-Iodo-2-metilbenzè ^a	132.8	124.3	132.8
68	1-Iodo-3-metilbenzè ^a	133.6	122.8	132.7
69	1-Iodo-4-metilbenzè ^b	121.9	123.0	132.8
70	1-Iodopropà ^a	-30.0	-30.2	-29.2
71	1-Metilciclohexè ^b	-43.3	-29.8	-43.0
72	1-Metilciclopentè ^a	-3.8	-3.4	-5.8
73	1-Pentè ^a	-21.1	-32.8	-21.0
74	2,2,3,3-Tetrametilbutà ^a	-226.0	-234.3	-224.1
75	2,2,3-Trimetilbutà ^a	-204.4	-205.7	-203.8
76	2,2,3-Trimetilpentà ^a	-219.9	-223.7	-222.5
77	2,2,4-Trimetilpentà ^a	-223.9	-229.6	-223.7
78	2,2-Dicloropropà ^a	-173.2	-177.7	-175.9
79	2,2-Dimetilbutà ^a	-185.9	-182.1	-183.8
80	2,2-Dimetilhexà ^a	-224.5	-223.3	-225.3
81	2,2-Dimetilpentà ^a	-205.7	-202.8	-204.5
82	2,2-Dimetilpropà ^b	-168.0	-167.6	-168.0
83	2,3,3-Trimetilpentà ^b	-216.2	-221.4	-219.5
84	2,3,4-Trimetilpentà ^b	-217.2	-219.6	-219.6
85	2,3-Dibromo-2-metilbutà ^a	-137.6	-133.0	-133.4
86	2,3-Dibromobutà ^a	-102.0	-100.0	-107.9
87	2,3-Dimetilbutà ^a	-178.1	-176.7	-177.4
88	2,3-Dimetilhexà ^a	-213.8	-215.1	-216.9
89	2,3-Dimetilpentà ^a	-198.7	-194.3	-196.1
90	2,4-Dimetilhexà ^a	-219.2	-216.8	-218.9
91	2,4-Dimetilpentà ^a	-201.6	-199.2	-200.1
92	2,5-Dimetilhexà ^a	-222.5	-219.4	-221.9
93	2-Bromo-2-metilpropà ^a	-132.4	-129.9	-132.4
94	2-Bromopropà ^a	-99.4	-95.4	-99.4
95	2-Butí ^a	145.7	157.2	145.7
96	2-Cloro-2-metilpropà ^b	-182.2	-176.2	-183.0
97	2-Cloro-3-metilbutà ^a	-202.2	-192.6	-198.7
98	2-Cloropropà ^a	-144.9	-142.1	-145.6
99	2-Clorohehexà ^a	-204.3	-204.7	-206.0
100	2-Iodo-2-metilpropà ^a	-72.1	-74.6	-72.0

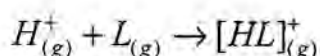
#	Compost	Experimental	QSPR	Pedley
101	2-Iodopropà ^a	-40.3	-40.0	-40.5
102	2-Metil-1-butè ^a	-35.2	-38.3	-35.5
103	2-Metil-2-butè ^b	-41.7	-38.0	-42.0
104	2-Metilbutà ^a	-153.6	-152.5	-152.9
105	2-Metilheptà ^a	-215.3	-213.4	-215.2
106	2-Metilhexà ^a	-194.5	-193.0	-194.4
107	2-Metilpentà ^b	-174.6	-172.7	-173.7
108	2-Metilpropà ^a	-134.2	-136.5	-134.1
109	2-Metilpropè ^a	-16.9	-20.7	-17.5
110	3,3-Dimetilhexà ^a	-219.9	-218.4	-220.3
111	3,3-Dimetilpentà ^b	-201.0	-197.4	-199.5
112	3,4-Dimetilhexà ^a	-212.8	-212.2	-214.9
113	3-Etil-2-metilpentà ^b	-211.0	-212.8	-214.9
114	3-Etil-3-metilpentà ^b	-214.8	-213.5	-215.3
115	3-Etilhexà ^a	-210.7	-208.3	-211.2
116	3-Etilpentà ^a	-189.5	-187.5	-190.4
117	3-Metil-1-butè ^a	-27.5	-38.5	-27.5
118	3-Metilciclopentè ^a	7.4	-3.7	4.5
119	3-Metilheptà ^a	-212.5	-210.5	-213.2
120	3-Metilhexà ^a	-191.3	-190.1	-192.4
121	3-Metilpentà ^b	-171.9	-169.5	-171.6
122	4-Etenilciclohexè ^a	66.4	61.0	70.0
123	4-Metilciclopentè ^a	14.6	-3.7	4.5
124	4-Metilheptà ^b	-211.9	-210.8	-213.2
125	Benzè ^a	82.6	92.9	82.6
126	Bromobenzè ^b	105.4	105.5	105.4
127	Bromoetà ^a	-61.9	-63.0	-65.4
128	Bromometà ^a	-35.4	-32.1	-35.4
129	Butà ^a	-125.7	-127.3	-125.5
130	Butilbenzè ^a	-11.8	-23.4	-11.8
131	Butilciclohexà ^b	-213.7	-213.4	-215.1
132	Clorobenzè ^a	52.0	58.5	52.0
133	Clorociclohexà ^a	-163.7	-155.4	-164.3
134	Cloroetà ^b	-112.1	-110.1	-111.9
135	Clorometà ^a	-81.9	-79.6	-81.9
136	Ciclohexa-1,4-diè ^a	100.4	118.8	100.4
137	Ciclohexà ^a	-123.3	-115.8	-123.3
138	Ciclohexè ^a	-4.9	0.3	-4.9
139	Ciclopentà ^a	-76.4	-89.1	-76.4
140	Ciclopentè ^a	34.0	29.1	34.0
141	Diclorometà ^a	-95.4	-86.5	-97.1
142	Diiodometà ^a	119.5	119.8	119.5
143	Età ^a	-83.8	-88.5	-83.8
144	Etè ^b	52.5	33.8	52.5
145	Ethenilciclohexà ^a	-48.9	-53.5	-48.3
146	Etilbenzè ^a	30.0	29.8	29.8
147	Etilciclohexà ^b	-171.5	-166.5	-173.5
148	Etilciclopentà ^b	-126.9	-142.5	-126.6
149	Etí ^a	228.2	220.8	226.6
150	Heptà ^a	-187.6	-186.9	-187.8
151	Hexametilbenzè ^a	-77.4	-81.6	-110.1

#	Compost	Experimental	QSPR	Pedley
152	Hexà ^b	-166.9	-166.7	-166.9
153	Iodobenzè ^a	164.9	160.1	164.9
154	Iodociclohexà ^a	-50.0	-54.8	-59.2
155	Iodoetà ^a	-8.1	-7.7	-8.4
156	Iodometà ^a	14.4	23.2	14.3
157	Metilbenzè ^a	50.5	56.8	50.6
158	Metilciclohexà ^b	-154.6	-144.5	-152.7
159	Metilciclopentà ^a	-106.2	-119.5	-105.9
160	Octà ^a	-208.5	-207.3	-208.5
161	Pentametilbenzè ^a	-67.2	-57.1	-78.0
162	Pentà ^a	-146.9	-146.8	-146.2
163	Propà ^a	-104.7	-109.5	-104.7
164	Propè ^a	20.0	8.6	20.5
165	Propilbenzè ^a	7.9	2.3	9.0
166	Propilciclohexà ^b	-192.3	-190.3	-194.3
167	Propí ^a	184.9	182.3	186.1
168	<i>tert</i> -Butilbenzè ^a	-23.0	-37.0	-33.6

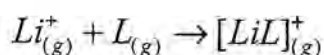
^a Training set. ^b Prediction set.

3.4. Basicitat catiònica de compostos orgànics

Les interaccions àcid-base són, sense cap mena de dubte, un dels processos més importants en química i bioquímica. Els estudis quantitius de molècules en fase gas proporcionen les acideses i basicitats, lliures de tota influència de les molècules de solvent i dels contraions corresponents. Una de les reaccions més estudiades en fase gas és l'equilibri de transferència de protó, de manera que es disposa de gran quantitat de valors de l'afinitat (PA) i la basicitat (GB) protòniques en fase gasosa per a diferents compostos orgànics, L.⁸⁸ Aquestes magnituds venen definides en termes de la hipotètica reacció en fase gas



on PA i GB són, amb signe negatiu, l'increment d'entalpia i d'energia lliure de Gibbs, respectivament, de la reacció anterior. Després del protó, el catió liti va ser el primer metall estudiat. Anàlogament a la reacció de transferència de protó, la basicitat i l'afinitat catiòniques front el liti (LCB i LCA respectivament), es defineixen com el valor negatiu de l'energia lliure de Gibbs i de l'entalpia de la reacció de formació del compost de coordinació $[LiL]^+$.



Cal dir que les característiques de coordinació del catió liti són molt diferents a les del protó. Aquest últim s'addiciona a la base mitjançant un enllaç σ covalent i polar, amb conseqüència d'una gran transferència de càrrega; mentre que els enllaços que es formen amb el catió liti són produïts, eminentment, per interaccions electrostàtiques de tipus ió-dipol. Com a resultat d'aquesta interacció diferent, els valors de LCB són molt més petits que els de GB i cobreixen un rang molt menor en l'escala d'energies. D'altra banda, el catió liti pot formar quelats en els quals el metall actua com a pont entre dos o més àtoms de la base. Els diferents tipus d'enllaç que s'estableixen entre el protó i el catió liti amb les bases dona lloc a grans variacions en l'ordre de basicitat.⁸⁹

La majoria de dades disponibles sobre afinitats, acideses i basicitats del catió liti en fase gasosa s'han obtingut d'experiments en els quals s'ha determinat la constant

d'equilibri de la reacció de formació del compost $[\text{LiL}]^+$, on L és una base orgànica. Els valors de LCB es determinen amb tècniques experimentals variades, sent l'espectrometria de masses una de les més utilitzades. També es poden obtenir valors de LCB mitjançant ressonància iònica de ciclotró (FT-ICR),⁹⁰ espectrometria de masses a elevada pressió (HPMS),⁹¹ dissociació unimolecular (mètode cinètic de Cooks),^{92,93} i dissociació per col·lisió induïda (CID).⁹⁴ També s'han obtingut valors de LCB a partir de mètodes teòrics⁸⁹ a diferents nivells de càlcul, *ab initio* (G2 i G2MP2) i DFT (B3LYP/6-311+G**) amb els quals s'obté una bona correlació entre valors experimentals i calculats. Per últim, hi ha un treball de l'any 2004, on s'aplica la metodologia QSPR de la mateixa manera, construint models lineals emprant un gran nombre dels valors experimentals utilitzats aquí.⁹⁵

La basicitat catiònica de compostos orgànics sembla una propietat adequada per ser estudiada mitjançant la metodologia QSPR ja que el seu valor numèric (LCB) depèn directament de l'estructura molecular i perquè no hi ha molècules de solvent a tenir en compte. Així doncs, és d'esperar que es puguin obtenir models que proporcionin bones prediccions per aquesta propietat tot i la complexitat de les interaccions entre el catió liti i les bases orgàniques. En aquest capítol, s'establiran models lineals i no lineals que permetin el càlcul de LCB només a partir de l'estructura de les bases orgàniques.

Conjunt de dades

El conjunt de dades a estudiar consta de 229 molècules de les quals es coneix el seu valor de LCB. La majoria d'aquests valors, uns dos-cents, han estat extrets del treball de Burk i col·laboradors.⁸⁹ Aquests valors pertanyen a bases oxigenades: alcohols, èters, grups carbonílics, sulfòxids, sulfones i compostos amb grups P=O; bases nitrogenades: cianurs, amines i amides; bases sulfurades: tiols i tioèters; bases fosforilades: fosfines i òxids de fosfina; i alguns compostos inorgànics: aigua, amoníac, diòxid de sofre i POCl_3 . Els conjunt de valors de LCB es completa amb compostos determinats més recentment: tres alquilbenzens (etil-, n-butil- i heptil-),⁹⁶ el benzoat de metil i els tres isòmers del dimetilftalat,⁹⁷ quinze aminoàcids⁹⁸ i quatre bases aromàtiques amb anells fusionats: naftalè, azulè, antracè i fenantrè.⁹⁹ Els valors experimentals de LCB es cobreixen un rang de 74.8 a 221.33 kJ/mol, amb un valor mitjà de 145.6 kJ/mol. La taula 3.52 conté els valors experimentals i els calculats de

LCB per a tots els compostos. Aquesta taula conté, també, la classificació dels compostos estudiats en els diferents subconjunts de treball, així, per al model lineal s'han generat els subconjunts *training set* (*tset*) i *prediction set* (*pset*) amb 185 i 44 compostos, respectivament; mentre que per als models no lineals que utilitzen xarxes neuronals BPNN s'han generat els subconjunts *tset*, *pset* i *cvset* amb 166, 44 i 19 compostos cadascun.

Model lineal

Els models lineals s'establiran mitjançant el mètode heurístic que implementa CODESSA. El mètode heurístic realitza un pretractament dels descriptors amb el qual és capaç de reduir el seu nombre inicial, en aquest cas, el nombre de descriptors es veu rebaixat dels 808 inicials a 180.

El model de set descriptors obtingut amb la regressió multilinear per als 185 compostos que formen el *tset* es mostra en la Taula 3.53. El nivell de significació, associat al *tset-t* de Student, mostra l'existència de correlació entre LCB i els descriptors triats. La correlació entre els set descriptors es troba en el rang 0.02-0.63, amb un valor mitjà de 0.22. El model presenta un valor VIF mitjà de 2.03, que indica poca col·linealitat entre descriptors.

Taula 3.53. Model obtingut per regressió lineal.

Descriptor	coeficient	<i>sd</i>	β	test- <i>t</i>	VIF
Terme independent	101.52	4.30	-	23.61	-
Nombre d'àtoms de nitrogen	10.53	1.66	0.27	6.36	1.83
Nombre relatiu d'àtoms de sofre	-113.85	24.79	-0.16	-4.59	1.18
Nombre relatiu d'àtoms de fluor	-74.55	7.97	-0.34	-9.35	1.36
Contingut d'informació estructural d'ordre 0	4.99	0.69	0.32	7.28	1.96
Àrea superficial positiva fraccionada, FPSA-2	13.26	1.86	0.34	7.13	2.30
HASA-2/TMSA	-490.69	101.08	-0.29	-4.85	3.54
HACA-2	10.67	1.14	0.42	9.37	2.06

El model conté tres descriptors constitucionals: nombre d'àtoms de nitrogen i nombre relatiu d'àtoms de sofre i fluor; un descriptor topològic: contingut d'informació estructural d'ordre zero; i tres descriptors electrostàtics: FPSA-2, HASA-2/TMSA i HACA-2. Els descriptors constitucionals són, potser, els més fàcilment interpretables.

El nombre d'àtoms de nitrogen té, lògicament, un coeficient positiu, com més àtoms de nitrogen tingui la base més fàcilment es pot formar un quelat i, a més, segons la teoria de Pearson, el liti s'enllaçarà més favorablement a àtoms durs com el nitrogen. Per contra, el nombre relatiu d'àtoms de sofre té signe negatiu, seguint també la teoria de Pearson. El nombre relatiu d'àtoms de fluor presenta signe negatiu, és lògic pensar que com més elevat sigui el nombre d'àtoms d'aquest tipus, menys densitat electrònica quedarà en els heteroàtoms per poder formar enllaços per donació al catió liti. Els descriptors topològics descriuen la connectivitat dels àtoms dintre de la molècula. Una classe específica d'aquests descriptors són els índex de complexitat molecular (que es basen en la teoria d'informació de Shannon),²⁹ als quals pertany el contingut d'informació estructural d'ordre zero. Els tres descriptors electrostàtics pertanyen al tipus CPSA i, tots tres, es calculen amb les càrregues parcials obtingudes pels mètodes quàntics de MOPAC. El descriptor àrea superficial positiva fraccionada FPSA-2, s'obté:

$$FPSA - 2 = \frac{\sum_A q_A \sum_A S_A}{TMSA}$$

on q_A són les càrregues atòmiques parcials positives, S_A són les superfícies atòmiques accessibles al solvent carregades positivament i $TMSA$ és la superfície total de la molècula. Els altres dos descriptors pertanyen a la subclasse HB-CPSA i indiquen interaccions per enllaç d'hidrogen; tots dos es refereixen a les característiques acceptores de la superfície carregada dels àtoms de la molècula. El descriptor HASA-2/TMSA es calcula :

$$HASA - 2 = \sum_A \frac{q_A \sqrt{S_A}}{\sqrt{S_{TOT}}}$$

on q_A és la càrrega parcial sobre l'àtom acceptor de l'enllaç d'hidrogen, S_A és la superfície d'aquest àtom, S_{TOT} és la superfície molecular accessible al solvent i $TMSA$ és la superfície total de la molècula.

El descriptor HACA-2 es calcula de manera molt similar:

$$HACA-2 = \sum_A \frac{q_A \sqrt{S_A}}{\sqrt{S_{TOT}}}$$

on q_A és la càrrega parcial sobre l'àtom acceptor de l'enllaç d'hidrogen, S_A és la superfície d'aquest àtom i S_{TOT} és la superfície molecular accessible al solvent. En aquest cas, només es consideren àtoms acceptors a aquells àtoms que presenten càrregues parcials per sobre d'un valor llindar. Aquests tres descriptors es relacionen amb les interaccions intermoleculars i, en aquest cas, reflecteixen les interaccions ió-dipol que s'estableixen entre el catió liti i la base. Segons els coeficients de regressió estandarditzats, el descriptor més important dintre d'aquest model és HACA-2, seguit per FPSA-2 i el nombre relatiu d'àtoms de fluor.

Els resultats que s'obtenen per als dos subconjunts de compostos *tset* i *pset* es troben en la Taula 3.54, el valor de $R^2_{cv}=0.806$, demostra l'estabilitat del model establert. La Figura 3.19 mostra la representació gràfica dels valors calculats vs. valors experimentals de LCB per a aquests dos conjunts.

Taula 3.54. Paràmetres estadístics obtinguts amb el model lineal.

Conjunt	n	R ²	F	rmse
<i>tset</i>	185	0.825	863.3	13.21
<i>pset</i>	44	0.878	274.4	10.45

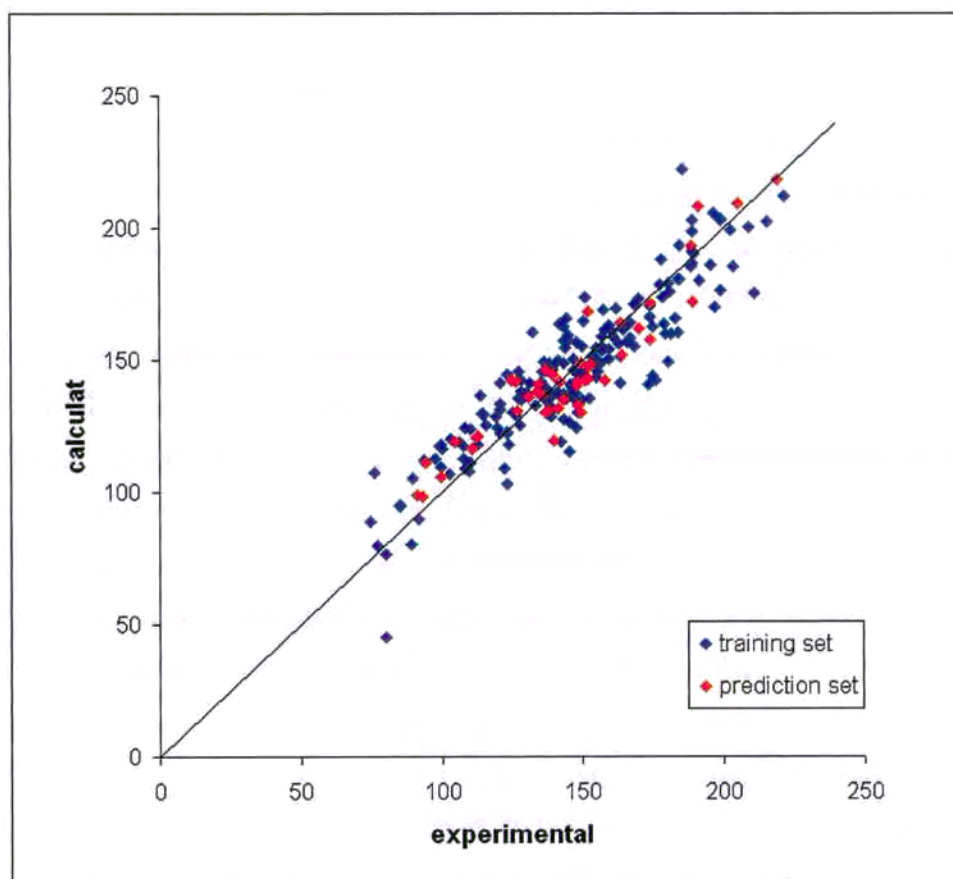


Figura 3.19. Representació gràfica de LCB calculat vs. experimental amb el model lineal.

Els enllaços que es formen entre el catió liti i les bases són majoritàriament deguts a interaccions electrostàtiques ió-dipol, en la formació aquests enllaços no hi ha una gran transferència de càrrega i el catió liti conserva entre un 80 i un 90% de la càrrega positiva. Apart d'aquest tipus d'interacció electrostàtica, el catió liti presenta altres interaccions específiques amb les bases, en gran mesura degudes a la possibilitat de formació de quelats, que poden modificar de manera espectacular el valor de LCB. Aquests quelats poden ser clàssics: el catió liti es coordina a un àtom d'oxigen o nitrogen; o no clàssics: on àtoms de fluor o clor intervenen en la formació d'anells.⁸⁹ Altres tipus de coordinació són possibles dependent de quina base entri en joc, per exemple, l'efecte "escorpí" que es dona amb els alquilbenzens⁹⁶ o la formació d'enllaços π entre catió i bases que contenen anells aromàtics fusionats.⁹⁹ De quelats clàssics, a través d'àtoms d'oxigen o nitrogen, s'han trobat per a la majoria dels compostos del conjunt de compostos estudiats. Aquest efecte també és present en els

adductes de dimetilftalat, així l'*o*-dimetilftalat presenta el valor de LCB més elevat dels tres isòmers ja que permet la formació del quelat amb els dos grups carbonílics contigus.⁹⁷ Pel que fa als aminoàcids, els valors més alts de LCB es troben en compostos que contenen àtoms d'oxigen en la cadena lateral. L'àcid aspàrtic, per exemple, presenta un valor de LCB molt elevat degut a la formació d'un anell de sis baules quan interactua amb el catió liti. Aquest valor de LCB disminueix quan es formen anells de set baules (àcid glutàmic) o quan la interacció es dona amb cadenes laterals que contenen substituents aromàtics (fenilalanina). D'aquesta manera, la glicina, que no presenta cadena lateral, té el valor de LCB més baix de tots.⁹⁸ La interacció d'alquilbenzens amb el catió liti segueix clarament el, ja comentat, efecte "escorpi" i els adductes més estables (amb LCB més grans) corresponen a aquells compostos en els que el substituent alquílic es dirigeix cap al sistema π , afavorint la interacció d'aquest últim amb el metall. La formació d'aquest tipus de complexos condueix a un augment significatiu en el LCB (entre 20 i 30 kJ/mol) enfront als alquilbenzens en els que el substituent alquílic no està dirigit cap al catió.⁹⁶ Les bases que contenen anells aromàtics fusionats formen complexos amb interaccions π amb el catió liti. Estudis teòrics han trobat que la força d'aquesta interacció decreix a mesura que el nombre d'anells fusionats creix. Així, es pot comprovar que l'estabilitat dels compostos en els que el catió liti s'enllaça als anells terminals és més elevada que la que es dona en compostos on el metall interacciona amb els anells més interns.⁹⁹

Tots aquests comentaris mostren la complexitat de les interaccions de les bases orgàniques amb el catió liti i demostren que subtils variacions en l'estructura molecular de les bases provoquen canvis dramàtics en els valors de LCB. Tot i així, els resultats que s'obtenen aquí amb la regressió multilíneal són prou bons i, com es veurà més endavant, són semblants als que s'obtenen treballant amb mètodes teòrics més complexos.

Per tal d'arrodonir aquest estudi, es van construir models lineals per quatre subconjunts amb compostos que, partint dels 229 inicials, contenen almenys un àtom d'oxigen (141 compostos), un àtom de nitrogen (87), un de sofre (41) i un de fluor (35); òbviament, els compostos de cada família poden contenir altres heteroàtoms apart del característic de cada conjunt. La Taula 3.55 mostra els paràmetres estadístics per a aquests quatre nous models així com els descriptors que els formen.

Taula 3.55. Models obtinguts per les diferents famílies de compostos amb regressió lineal.

Família	n	R ²	F	rmse	R ² _{cv}	Descriptors (classe)
Oxigen	141	0.919	112.2	6.4	0.893	Nombre relatiu d'àtoms de fluor (a) Contingut d'informació (ordre 0) (b) Àrea superficial parcial positiva, PPSA-3 (d) Àrea superficial carregada acceptora d'enllaç H, HACA-2 (d) Nombre de setis H-donadors (c) Contingut mitjà d'informació d'enllaç (ordre 1) (b) Atracció e-n mínima per un carboni (e)
Nitrogen	87	0.902	81.1	7.3	0.872	Àrea superficial parcial positiva, PPSA-1 (e) Dipol de la molècula (e) Nombre d'enllaços triples (a) Àrea superficial parcial negativa fraccionada, FNSA-3 (d) Energia d'estat atòmic màxima d'un carboni (e) Índex de reactivitat electròfila mínim d'un carboni (e) Contingut mitjà d'informació d'enllaç (ordre 1) (b).
Sofre	41	0.981	230.4	4.8	0.960	Càrrega positiva superficial total, PPSA-2 (d) Ordre d'enllaç σ - π màxim (d) Àrea superficial acceptora d'enllaç H, HASA-1 (d) Nombre relatiu d'àtoms de sofre (a) Àrea superficial relativa carregada negativament, RNCS (d) Àrea superficial parcial positiva normalitzada, WPSA-3 (d) Contingut d'informació (ordre 1) (b).
Fluor	35	0.960	93.6	6.6	0.920	Nombre relatiu d'àtoms de fluor (a) Contingut d'informació (ordre 0) (b) Nombre d'àtoms de nitrogen (a) Índex de reactivitat nucleòfila mitjà d'un carboni (e) Moment principal d'inèrcia C (f) Moment Dipolar de la molècula (e) Nombre d'enllaços dobles (a).

^a Constitucional. ^b Topològic. ^c Electrostàtic. ^d CPSA. ^e Quàntic. ^f Geomètric.

Com es pot comprovar els resultats que s'obtenen en tots els casos són millors que amb el conjunt de compostos complet, amb clares millores en els coeficients de determinació i els errors. Els descriptors d'aquests quatre nous models són molt semblants als que conté el model construït amb tots els compostos. Aquests models contenen, majoritàriament, descriptors constitucionals, electrostàtics i topològics, sent l'única excepció el model per la família del fluor, on no apareixen descriptors electrostàtics. Així doncs, es pot dir que els descriptors d'aquests models contenen el mateix tipus d'informació que el model desenvolupat pel *training set* i que tots ells són capaços d'explicar prou bé les característiques bàsiques de les molècules en fase gas.

Models no lineals

Per tal d'establir models no lineals s'han fet servir els dos mètodes implementats en el programa ADAPT. El primer és l'anomenat híbrid lineal-no lineal, que utilitza els descriptors triats amb el mètode lineal per construir la xarxa neuronal. El segon mètode és no lineal amb BPNN, que fa una selecció no lineal de descriptors amb algorismes genètics i calcula la propietat amb la xarxa neuronal. Totes les xarxes neuronals construïdes consten de tres capes i la seva arquitectura optimitzada és 7-5-1. A partir de l'arquitectura es pot calcular el nombre de paràmetres ajustables, en aquest cas 46; dividint aquest nombre pel nombre de compostos del *tset* d'ADAPT, 166, s'obté un valor de 3.6, superior al mínim acceptable: 2.0. Un cop ensinistrades les xarxes neuronals, s'obtenen els valors de LCB calculat que es troben en la Taula 3.52.

Mètode híbrid

Els paràmetres estadístics que s'obtenen pels tres subconjunts de compostos, Taula 3.56, amb aquest model no lineal són clarament millors als que s'obtenen amb el model lineal.

Taula 3.56. Paràmetres estadístics obtinguts amb el model híbrid.

Conjunt	n	R²	F	rmse
<i>tset</i>	166	0.905	1559.1	9.42
<i>cvset</i>	19	0.880	124.5	11.53
<i>pset</i>	44	0.885	323.5	9.80

La Figura 3.20 mostra la representació gràfica dels valors calculats vs. valors experimentals de LCB per *training*, *cross-validation* i *prediction sets*.

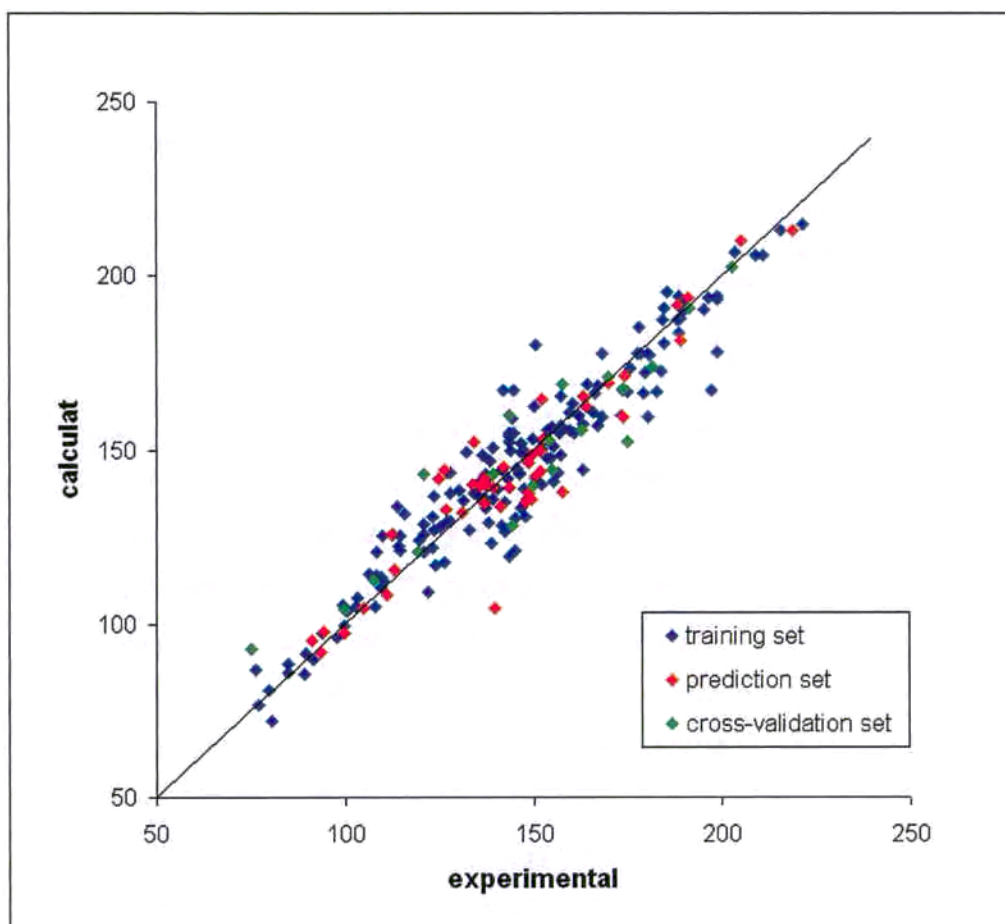


Figura 3.20. Representació gràfica de LCB calculat vs. experimental amb el model híbrid.

Xarxes neuronals BPNN

Aquest mètode consta d'una selecció de descriptors no lineal, que es fa mitjançant un algorisme genètic, aplicada a una xarxa neuronal d'arquitectura 7-5-1. Les rutines de selecció objectiva implementades per ADAPT redueixen el nombre de descriptors inicials de 180 a 102. A partir d'aquests descriptors s'obtenen deu models, de set descriptors, completament no lineals. El millor d'aquests models conté els descriptors que es troben en la Taula 3.57.

Taula 3.57. Descriptors continguts en el model BPNN.

Descriptor	Classe
Nombre d'àtoms d'hidrogen	Constitucional
Nombre d'àtoms d'oxigen	Constitucional
Nombre d'àtoms de nitrogen	Constitucional
Àrea carregada donadora d'enllaç d'hidrogen, HDCA-2	CPSA
Energia HOMO-1	Quàntic
Moment Dipolar	Quàntic
Interacció electrostàtica total/nombre d'àtoms	Quàntic

Aquest model conté tres descriptors constitucionals, tres quàntics i un electrostàtic del tipus CPSA. Els descriptors constitucionals, que són molt anàlegs als que s'obtenen amb el model lineal, indiquen la capacitat de la base per poder enllaçar el catió liti amb els seus heteroàtoms. El descriptor HDCA-2 és també molt similar als descriptors CPSA que es poden trobar en el model lineal. Aquest descriptor, que descriu interaccions d'enllaç d'hidrogen de tipus donador, i el dipol de la molècula defineixen les interaccions ió-dipol entre el catió liti i la base.

Els resultats que s'obtenen amb el model desenvolupat amb BPNN, Taula 3.58, són encara millors que els que s'han obtingut amb el mètode híbrid. Els valors de LCB calculats per amb aquest mètode es poden trobar a la Taula 3.52 i s'han representat gràficament en la Figura 3.21.

Taula 3.58. Paràmetres estadístics obtinguts amb el model BPNN.

Conjunt	n	R ²	F	rmse
<i>tset</i>	166	0.954	3413.7	6.54
<i>cvset</i>	19	0.877	121.0	11.03
<i>pset</i>	44	0.914	448.8	8.61

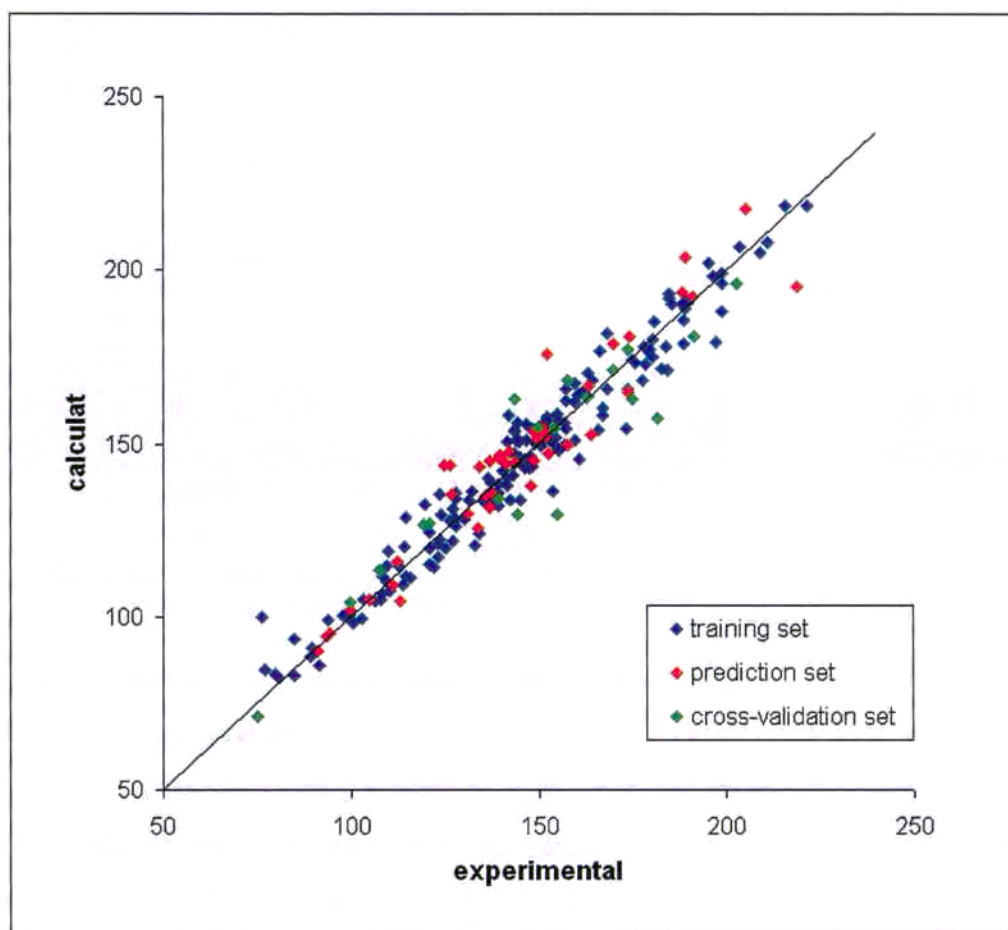


Figura 3.21. Representació gràfica de LCB calculat vs. experimental amb el model BPNN.

El valor de *rmse* que presenta el model complet és de 7.37. La Taula 3.59 mostra els 13 compostos que presenten valors de residual més gran que dues vegades el *rmse*; entre ells n'hi ha cinc que presenten residuals tres vegades més grans que aquest valor i que, per tant, es podrien considerar *outliers*. Aquests compostos són el triptòfan, la 1,8-nafitiridina, el carbonat de dimetil ((MeO)₂CO), el dimetiltereftalat i el diòxid de sofre. No sembla que l'error comès en aquests compostos es degui a què el model establert funcioni malament per a un determinat tipus de compostos amb similitud estructural.

Taula 3.59. Prediccions amb valors més grans que dues vegades el *rmse**.

Compost	Experimental	BPNN	residual
Dimetiltrefalal	152.00	176.19	-24,19
SO ₂	76.15	100.02	23.87
HCO ₂ Bu	143.51	162.85	-19.34
Pirimidina(1,3)	124.68	143.88	-19.20
PhOMe	126.36	143.95	-17.59
CF ₃ CONH ₂	141.84	158.44	-16.60
(cPr) ₂ CO	160.67	145.69	14.98
(CH ₂) ₃ OSO ₂	153.55	136.11	17.44
Dimetilftalat	196.98	179.27	17.71
Piridazina(1,2)	173.22	154.56	18.66
Triptòfan	218.82	195.4	23.42
1,8-naftiridina	181.59	157.31	24.28
(MeO) ₂ CO	154.81	129.7	25.11

* Els compostos ombrejats tenen residuals més grans que tres vegades el *rmse*

Comparació dels resultats

Com ja s'ha comentat anteriorment, s'han trobat en la bibliografia diversos intents de calcular LCB amb mètodes teòrics. En un d'ells s'utilitzen mètodes *ab initio* a nivell G2 i G2MP2; i mètodes DFT per calcular LCB per dos conjunts de 36 i 66 compostos.⁸⁹ També s'han trobat altres estudis on s'utilitzen mètodes DFT a nivell B3LYP/6-311+G** per calcular LCB per alguns dels compostos aquí estudiats.^{96,99}

En el treball més recent, s'aplica la metodologia QSPR de la mateixa manera, emprant també 205 dels valors experimentals utilitzats aquí.⁹⁵ El model lineal desenvolupat és molt semblant al que s'ha establert aquí: els sis descriptors que el formen són semblants i presenta resultats acceptables, del mateix ordre als que s'obtenen aquí amb el mètode lineal; tot i així, aquests resultats són clarament inferiors als que s'obtenen amb els mètodes no lineals.

Per comparar la validesa dels models proposats en aquest treball, també es van comparar els resultats obtinguts amb els que es van trobar en la bibliografia, de nivell de càlcul superior. A la Taula 3.60 es troben els paràmetres estadístics obtinguts amb cada mètode de càlcul pels dos subconjunts de compostos trobats en la bibliografia, de 36, 66 i 205 compostos. Com es pot comprovar, amb el conjunt de 36 compostos els mètodes de càlcul *ab initio* i DFT donen els millors resultats, encara que el model BPNN s'hi acostava bastant. En canvi, amb el conjunt de 66 compostos els models QSPR construïts donen millors resultats que els mètodes DFT.

Taula 3.60. Estadística amb diferents mètodes pel càlcul de LCB.

Mètode de càlcul	Ref.	n	R ²	rmse
G2	89	36	0.974	5.14
G2MP2	89	36	0.974	5.21
DFT	89	36	0.975	6.53
Model lineal	-	36	0.832	9.32
Model híbrid	-	36	0.906	7.14
Model BPNN	-	36	0.935	5.57
DFT	89	66	0.922	11.43
Model lineal	-	66	0.843	10.09
Model híbrid	-	66	0.875	9,15
Model BPNN	-	66	0.942	6.20
QSPR ^a	95	205	0.801	8.78
Model lineal	-	205	0.795	11.64
Model híbrid	-	205	0.880	9.12
Model BPNN	-	205	0.932	7.05

^a Descriptors: nombre relatiu d'àtoms S, energia HOMO-1, FPSA-1, HASA-2, càrrega atòmica neta mínima, component carregat del moment dipolar.

En resum, pot concloure's que ha estat possible establir models lineals i no lineals amb xarxes neuronals BPNN que permeten el càlcul de LCB per a un conjunt de 229 compostos que conté substàncies de naturalesa química molt variada. Aquests models contenen set descriptors calculats directament de l'estructura molecular dels compostos químics.

Els resultats que s'obtenen d'aplicar la metodologia QSPR al càlcul de LCB són bons, sobretot tenint en compte la complexitat de les possibles interaccions que es poden produir entre el catió liti i les diferents bases. Encara que els dos mètodes emprats donen resultats satisfactoris, els valors de LCB que s'obtenen amb xarxes neuronals són més pròxims als experimentals.

La comparació dels resultats que s'obtenen amb el models QSPR amb els que s'obtenen amb altres càlculs de nivell teòric més complex (*ab initio* o DFT), amb els mateixos conjunts de compostos, demostra que la metodologia QSPR pot establir models per calcular LCB amb resultats iguals o fins i tot millors i de manera molt més ràpida.

Taula 3.52. Valors experimentals i calculats de LCB (kJ/mol).

#	Compost	Experimental	MLR	Híbrid	BPNN
1	(C ₆ H ₅ O) ₃ PO ^{a,c}	189.12	198.95	187.72	191.28
2	(CF ₃) ₂ CHOMe ^{a,c}	109.62	110.99	110.34	114.63
3	(CF ₃) ₂ CO ^{a,c}	79.91	76.50	80.69	83.32
4	(CF ₃) ₃ CNH ₂ ^{a,c}	99.58	109.75	99.66	100.56
5	(CF ₃) ₃ COH ^{a,c}	84.94	95.30	88.37	93.75
6	(CF ₃ CH ₂) ₂ O ^{a,c}	122.17	109.25	109.40	114.38
7	(CF ₃ CO) ₂ CH ₂ ^{a,c}	114.22	129.85	122.43	120.32
8	(CF ₃ CO) ₂ NH ^{a,c}	147.28	140.31	138.91	142.93
9	(CH ₂) ₃ OSO ₂ ^{a,c}	153.55	151.45	155.81	136.11
10	(CF ₃) ₂ S ^{a,c}	80.33	45.21	72.07	82.71
11	(CH ₂) ₄ S ^{a,c}	107.95	117.58	105.07	104.85
12	(CH ₂) ₅ S ^{a,c}	108.37	124.42	113.92	111.58
13	(CHF ₂) ₂ CO ^{a,c}	102.93	106.82	104.83	99.59
14	(EtO) ₃ PO ^{a,c}	188.70	186.55	183.50	185.87
15	(ⁱ OPr) ₂ PO(H) ^{a,c}	184.51	180.59	180.60	192.08
16	(Me ₂ N) ₃ PO ^{a,c}	198.74	203.09	193.19	199.09
17	(MeO) ₂ PO(H) ^{a,c}	177.82	188.32	185.30	178.25
18	(MeO) ₂ PO(Me) ^{a,c}	184.10	180.60	187.49	171.47
19	(MeO) ₂ SO ₂ ^{a,c}	142.26	136.73	135.02	147.47
20	(MeO) ₃ PO ^{a,c}	182.84	166.10	166.80	171.67
21	(MeO)MePO(Ph) ^{a,c}	188.70	203.30	194.10	178.89
22	(MeS)SO ₂ (Me) ^{a,c}	152.30	148.32	150.17	157.75
23	((CF ₃) ₂ CF) ₂ CO ^{a,c}	91.63	90.10	89.86	86.02
24	1,2,4-triazol ^{a,c}	136.82	140.16	136.36	135.05
25	1,2-dimetilimidazol ^{a,c}	174.89	162.84	166.59	174.37
26	1,3,5-trimetilpirazol ^{a,c}	160.25	161.09	163.13	164.03
27	1,4-dimetilpirazol ^{a,c}	154.81	154.53	156.73	158.81
28	1,5-dimetilpirazol ^{a,c}	157.32	154.76	157.02	162.28
29	Cianur d'1-adamantil ^{a,c}	159.41	154.31	155.76	161.87
30	1-metilimidazol ^{a,c}	168.20	155.19	159.51	165.87
31	1-metilpirazol ^{a,c}	143.51	148.84	152.27	153.88
32	2,4,5-trimetilimidazol ^{a,c}	178.24	174.11	177.54	173.11
33	2,5-dimetiltetrahidrofuran ^{a,c}	146.44	139.64	143.30	145.72
34	2,6-difluoropiridina ^{a,c}	138.91	128.76	123.30	135.76
35	2-fluoropiridina ^{a,c}	146.86	135.72	133.45	144.69
36	3(5)-metilpirazol ^{a,c}	146.86	150.35	152.05	150.89
37	3,4,5-trimetilpirazol ^{a,c}	161.92	169.64	160.03	165.61
38	3-cloropiridina ^{a,c}	132.21	160.83	149.53	136.25
39	4-(CF ₃)piridina ^{a,c}	123.43	144.64	136.71	122.25
40	4-(dimetilamino)piridina ^{a,c}	175.73	142.29	173.50	173.46
41	4-CF ₃ PhOPO(Ph) ₂ ^{a,c}	185.35	222.31	195.29	190.20
42	4-MeC ₆ H ₄ SO ₂ (Me) ^{a,c}	168.20	171.21	177.65	182.08
43	4-NO ₂ C ₂ H ₄ SO ₂ (Me) ^{a,c}	150.62	174.10	180.11	155.48
44	Alanina ^{a,c}	180.75	176.16	177.16	185.40
45	Antracè ^{a,c}	141.50	132.13	134.04	140.20
46	Àcid aspàrtic ^{a,c}	215.48	202.71	213.35	218.76
47	Azulè ^{a,c}	145.10	126.58	130.22	133.61
48	BrCN ^{a,c}	123.01	102.97	122.02	117.28
49	C ₅ H ₁₁ CHO ^{a,c}	143.93	154.78	149.71	151.15
50	C ₆ H ₁₃ CHO ^{a,c}	144.77	159.51	154.77	150.93
51	C ₆ H ₅ OPO(Ph) ₂ ^{a,c}	196.23	205.69	193.77	198.27
52	C ₆ H ₆ ^{a,c}	112.55	118.27	125.65	114.66

#	Compost	Experimental	MLR	Híbrid	BPNN
53	c-C ₆ H ₁₁ CH ₂ OH ^{a,c}	143.51	157.19	153.94	144.10
54	CCl ₃ CH ₂ OH ^{a,c}	127.19	145.49	143.24	122.08
55	CCl ₃ CHO ^{a,c}	113.80	136.64	133.69	109.04
56	CF ₂ O ^{a,c}	76.99	79.56	76.82	84.81
57	CF ₃ (CH ₂) ₃ NH ₂ ^{a,c}	155.23	149.97	140.91	147.93
58	CF ₃ CH ₂ OCH=CH ₂ ^{a,c}	114.64	129.46	125.27	128.58
59	CF ₃ CH ₂ OMe ^{a,c}	123.85	118.09	116.92	129.55
60	CF ₃ CN ^{a,c}	89.12	80.50	85.65	88.63
61	CF ₃ CO ₂ Et ^{a,c}	128.03	135.04	129.53	135.92
62	CF ₃ CO ₂ Me ^{a,c}	120.92	122.60	120.57	124.28
63	CF ₃ COCH ₂ COMe ^{a,c}	147.70	157.08	149.18	146.56
64	CF ₃ CONH ₂ ^{a,c}	141.84	164.01	167.04	158.44
65	CF ₃ CONMe ₂ ^{a,c}	166.10	164.05	166.23	176.81
66	CF ₃ COSMe ^{a,c}	125.10	130.37	128.00	119.91
67	CH ₂ (CN) ₂ ^{a,c}	110.04	111.42	125.26	118.79
68	CH ₂ ClCN ^{a,c}	123.01	122.93	130.73	121.16
69	CH ₂ ClPO(OEt) ₂ ^{a,c}	184.51	193.88	190.67	193.26
70	CH ₂ FCN ^{a,c}	109.62	108.03	113.37	110.06
71	CHCl ₂ CN ^{a,c}	115.90	125.45	131.50	111.15
72	Cl ₃ PO ^{a,c}	145.18	115.18	121.06	145.95
73	(cPr) ₂ CO ^{a,c}	160.67	159.23	154.96	145.69
74	cPrCOMe ^{a,c}	156.48	148.70	143.37	149.24
75	Dimetilacetamida ^{a,c}	179.08	163.84	166.12	176.69
76	Dimetilisofalal ^{a,c}	157.11	169.09	165.55	165.90
77	Dimetilfitalat ^{a,c}	196.98	170.10	166.97	179.27
78	Et ₂ CO ^{a,c}	153.55	151.55	147.47	150.23
79	Et ₂ O ^{a,c}	139.33	131.84	135.83	135.76
80	Et ₂ S ^{a,c}	110.46	123.82	112.35	107.58
81	Et ₃ PO ^{a,c}	195.39	186.15	190.38	202.22
82	EtCHO ^{a,c}	137.24	139.21	133.39	136.33
83	EtCN ^{a,c}	147.70	124.55	130.97	145.82
84	Etilbenzè ^{a,c}	130.21	137.57	138.34	128.43
85	EtOH ^{a,c}	127.19	135.66	129.48	131.39
86	EtSH ^{a,c}	89.54	105.20	91.39	91.13
87	Àcid glutàmic ^{a,c}	221.33	212.46	214.79	218.78
88	Sulfat de glicol ^{a,c}	138.07	130.09	129.27	135.21
89	Sulfit de glicol ^{a,c}	148.95	144.49	147.97	150.39
90	H ₂ CO ^{a,c}	106.27	118.95	114.08	104.56
91	H ₂ O ^{a,c}	103.34	120.25	107.43	104.83
92	HCN ^{a,c}	108.37	108.80	120.56	106.79
93	HCO ₂ -n-Pr ^{a,c}	143.51	157.58	154.76	153.66
94	HCONH ₂ ^{a,c}	157.32	154.24	148.48	154.64
95	HCONHMe ^{a,c}	165.69	157.07	160.89	154.05
96	ⁱ BuOH ^{a,c}	135.98	145.57	141.52	135.20
97	ⁱ BuSH ^{a,c}	99.16	117.81	105.33	100.00
98	ⁱ BuSMe ^{a,c}	114.64	129.91	121.05	111.53
99	Imidazol ^{a,c}	159.41	150.47	155.92	150.88
100	ⁱ Pr ₂ CO ^{a,c}	156.90	159.15	154.91	149.99
101	ⁱ Pr ₂ S ^{a,c}	120.92	131.39	125.86	119.74
102	ⁱ PrOH ^{a,c}	135.14	141.83	136.57	133.21
103	ⁱ PrSH ^{a,c}	93.72	112.12	97.28	99.25
104	Isoleucina ^{a,c}	189.54	190.64	189.82	190.71
105	Leucina ^{a,c}	189.12	191.56	192.11	189.11

#	Compost	Experimental	MLR	Híbrid	BPNN
106	Me ₂ NCN ^{a,c}	163.18	141.28	144.37	163.60
107	Me ₂ NH ^{a,c}	134.31	137.50	137.61	124.21
108	Me ₂ O ^{a,c}	123.43	122.44	127.16	135.57
109	Me ₂ S ^{a,c}	97.91	112.86	95.98	100.25
110	Me ₂ SO ₂ ^{a,c}	155.23	145.90	150.88	157.13
111	Me ₃ SiOMe ^{a,c}	144.77	157.62	166.99	150.68
112	MeC(OH)=CHCOMe ^{a,c}	180.33	179.06	177.48	180.35
113	MeCHO ^{a,c}	133.05	132.82	126.83	120.44
114	MeCN ^{a,c}	142.26	118.91	126.41	133.64
115	MeCO ₂ Me ^{a,c}	147.28	137.85	138.64	144.83
116	MeCOEt ^{a,c}	150.62	146.69	141.74	149.19
117	MeCONH ₂ ^{a,c}	166.94	164.43	168.30	158.39
118	MeCONHMe ^{a,c}	173.64	166.22	167.02	165.66
119	MeCOSMe ^{a,c}	141.42	137.97	128.36	137.85
120	MeOCH ₂ CN ^{a,c}	137.24	134.65	142.86	138.59
121	MeOCONMe ₂ ^{a,c}	166.94	158.47	157.13	160.33
122	MeSCH ₂ CN ^{a,c}	143.51	127.57	119.18	147.13
123	MeSH ^{a,c}	84.94	94.81	85.75	82.88
124	Metionina ^{a,c}	210.87	175.39	206.05	208.04
125	Benzoat de metil ^{a,c}	154.60	149.08	147.78	151.95
126	Naftalè ^{a,c}	127.70	125.47	129.43	128.20
127	Bu ₂ O ^{a,c}	152.72	148.89	152.40	152.97
128	Bu ₂ S ^{a,c}	128.03	142.71	143.27	126.25
129	BuCHO ^{a,c}	141.42	149.83	144.50	150.63
130	BuOH ^{a,c}	137.24	146.28	141.16	135.52
131	BuSH ^{a,c}	100.42	116.26	103.91	98.37
132	Butilbenzè ^{a,c}	136.40	149.63	148.72	140.06
133	NCCONMe ₂ ^{a,c}	144.35	165.90	158.99	156.00
134	neo-C ₃ H ₉ OH ^{a,c}	138.49	148.87	146.75	133.78
135	NH ₃ ^{a,c}	126.36	129.04	117.74	127.62
136	Heptilbenzè ^{a,c}	150.08	164.73	162.44	153.73
137	Cianur d'octil ^{a,c}	156.90	153.81	156.76	156.08
138	Pr ₂ O ^{a,c}	145.60	140.86	144.64	144.00
139	Pr ₂ S ^{a,c}	120.92	133.68	128.53	115.20
140	PrCN ^{a,c}	148.11	130.46	135.70	154.33
141	PrOH ^{a,c}	131.38	141.39	135.46	133.89
142	Ph ₂ SO ^{a,c}	183.68	160.58	172.56	178.02
143	Ph ₃ PO ^{a,c}	198.74	203.62	194.31	196.38
144	PhCH ₂ CN ^{a,c}	146.86	146.47	149.17	155.53
145	PhCH ₂ OH ^{a,c}	149.79	155.48	153.20	150.64
146	PhSO(Me) ^{a,c}	179.49	160.22	172.29	177.14
147	PhSO ₂ (Me) ^{a,c}	164.43	161.54	168.85	168.23
148	<i>p</i> -MeC ₆ H ₄ COMe ^{a,c}	159.41	163.35	160.90	167.49
149	Prolina ^{a,c}	198.74	176.60	178.00	188.14
150	Pirazina(1,4) ^{a,c}	119.66	124.69	124.15	132.41
151	Pirazol ^{a,c}	140.58	137.98	138.97	142.11
152	Piridazina(1,2) ^{a,c}	173.22	140.70	159.84	154.56
153	Piridina ^{a,c}	146.44	138.08	143.02	143.06
154	Serina ^{a,c}	203.34	185.80	206.73	207.08
155	SO ₂ ^{a,c}	76.15	107.29	86.62	100.02
156	^t Bu ₂ S ^{a,c}	128.03	138.50	137.46	134.30
157	^t BuCN ^{a,c}	152.30	135.40	140.04	156.36
158	^t BuOEt ^{a,c}	148.11	142.38	146.83	143.63

#	Compost	Experimental	MLR	Híbrid	BPNN
159	^t BuOH ^{a,c}	139.33	147.31	143.06	132.17
160	^t BuOMe ^{a,c}	143.09	137.44	141.74	140.88
161	Sulfona de tetrametilè ^{a,c}	163.18	151.85	157.25	170.56
162	Sulfòxid de tetrametilè ^{a,c}	180.33	149.22	159.31	175.37
163	Tetrametilguanidina ^{a,c}	177.40	178.85	177.50	168.51
164	Tetrazol ^{a,c}	139.33	136.90	150.59	138.32
165	Treonina ^{a,c}	208.78	200.29	206.00	205.16
166	Valina ^{a,c}	188.28	185.85	187.24	190.31
167	(CF ₃) ₃ CCO ₂ Et ^{a,d}	144.35	135.89	128.16	129.62
168	(MeO) ₂ CO ^{a,d}	154.81	143.59	144.36	129.70
169	1,8-naftiridina ^{a,d}	181.59	160.02	173.95	157.31
170	3-(dimetilamino)piridina ^{a,d}	169.66	161.50	171.02	171.21
171	C ₆ H ₅ SO ₂ (Me) ^{a,d}	157.74	161.55	168.89	168.24
172	CF ₃ CCH ^{a,d}	74.89	88.93	92.79	71.04
173	CF ₃ CO ₂ CH ₂ CF ₃ ^{a,d}	107.53	112.72	112.53	113.55
174	ClCO ₂ Me ^{a,d}	120.92	141.24	142.91	126.82
175	Dimetilsulfòxid ^{a,d}	174.89	143.93	152.38	163.08
176	HCO ₂ Bu ^{a,d}	143.51	162.88	160.02	162.85
177	Isoforona ^{a,d}	173.64	170.83	167.39	177.08
178	Me ₃ PO ^{a,d}	191.21	180.18	190.69	180.89
179	MeOH ^{a,d}	119.24	128.65	120.78	126.62
180	Cianur d'heptil ^{a,d}	153.97	149.57	152.88	154.51
181	Fenilalanina ^{a,d}	202.51	199.62	202.69	196.11
182	^s BuOH ^{a,d}	139.33	146.42	142.96	134.02
183	^t BuCO ₂ Et ^{a,d}	162.76	156.31	155.51	163.17
184	^t BuSH ^{a,d}	99.58	117.94	104.08	104.28
185	Valeronitril ^{a,d}	149.79	135.17	139.78	154.80
186	(4-FC ₆ H ₄ O)PO(Ph) ₂ ^b	190.79	208.70	193.87	192.63
187	(CF ₃) ₂ CHOH ^b	99.58	105.72	97.56	102.15
188	(CH ₂) ₃ OSO ^b	164.01	151.97	162.48	152.85
189	(EtO) ₂ PO(Me) ^b	188.28	193.53	191.78	193.65
190	1,2,3-triazol ^b	134.31	140.90	152.23	143.57
191	1,3,4,5,-tetrametilpirazol ^b	163.18	164.54	165.60	166.84
192	1,4-dioxà ^b	126.78	130.73	132.83	135.40
193	2-metil-tetrahidrofuran ^b	143.51	135.21	139.01	144.91
194	3-metilpiridina ^b	152.72	148.62	152.99	147.42
195	4-metilpirazol ^b	149.37	148.75	148.56	151.49
196	C ₅ F ₅ N ^b	93.30	98.46	91.96	94.44
197	CCl ₃ CN ^b	112.13	120.59	125.80	116.07
198	CF ₃ CH ₂ OH ^b	110.88	116.60	108.21	109.09
199	CF ₃ CHO ^b	91.21	99.08	95.36	90.28
200	CF ₃ COMe ^b	112.97	121.33	115.64	104.58
201	Cisteïna ^b	189.12	172.31	181.50	203.77
202	Dimetilformamida ^b	173.64	158.16	159.49	165.16
203	Dimetiltereftalat ^b	152.00	168.36	164.79	176.19
204	EtCO ₂ Me ^b	151.88	143.34	143.77	151.95
205	EtSMe ^b	104.60	118.89	104.38	104.98
206	Glicina ^b	174.05	171.98	171.56	181.12
207	HCO ₂ Et ^b	141.84	141.97	145.14	147.73
208	HCO ₂ Me ^b	135.56	136.44	139.78	135.15
209	^t Pr ₂ O ^b	148.53	142.18	146.56	145.33
210	^t PrCN ^b	149.37	130.31	135.64	152.64
211	Isooxazol ^b	137.65	130.65	139.66	135.38

#	Compost	Experimental	MLR	Híbrid	BPNN
212	Me ₂ CO ^b	147.70	140.52	134.78	138.01
213	Me ₃ N ^b	133.89	138.07	140.09	125.71
214	MeCO ₂ Et ^b	150.62	142.55	142.54	154.68
215	MeCOOH ^b	136.82	146.23	141.68	131.68
216	MeNH ₂ ^b	130.96	136.19	132.24	130.12
217	MeOSO ₂ (Me) ^b	151.88	147.52	149.78	153.95
218	n-PrCHO ^b	139.33	144.77	139.12	146.12
219	n-PrSH ^b	94.14	111.30	97.59	95.09
220	Ph ₂ SO ₂ ^b	169.87	162.05	169.03	178.79
221	PhCHO ^b	157.74	142.77	137.90	149.77
222	PhCN ^b	148.53	133.24	137.73	153.42
223	Fenantrè ^b	141.08	131.79	133.72	144.24
224	PhOMe ^b	126.36	142.26	144.22	143.95
225	Pirimidina(1,3) ^b	124.68	142.44	141.92	143.88
226	Tetrahidrofuran ^b	136.82	130.50	134.46	144.95
227	Tiazol ^b	139.75	119.66	104.51	146.66
228	Triptòfan ^b	218.82	218.37	213.03	195.40
229	Tirosina ^b	205.02	209.81	210.40	217.91

^a Training set de CODESSA. ^b Prediction set de CODESSA i ADAPT.

^c Training set d'ADAPT. ^d Cross-validation set d'ADAPT.

3.5. Bibliografia

1. Sigman, M. E.; Rives, S. S. *J. Chem. Inf. Comput. Sci.* **1994**, 34, 617.
2. Cundari, T. R.; Moody, E. W. *J. Chem. Inf. Comput. Sci.* **1997**, 37, 871.
3. Cundari, T. R.; Moody, E. W. *J. Mol. Struct.(Theochem)* **1998**, 425, 43.
4. Eng, G.; Brinckman, F. E.; Olson, G. J.; Tierney, E. J.; Bellama, J. M. *App. Organometallic Chem.* **1991**, 5, 33.
5. Ownby, D. R.; Newman, M. C. *QSAR Comb. Sci.* **2003**, 22, 241.
6. Song, X.; Zapata, A.; Eng, G. *J. Organomet. Chem.* **2006**, 691, 1756.
7. Pavan, M.; Worth, A.P. EUROPEAN COMMISSION, DIRECTORATE GENERAL, JOINT RESEARCH CENTRE, EUR 22355 EN, **2006**.
8. Pavan, M.; Worth, A.P.; Netzeva, T. EUROPEAN COMMISSION, DIRECTORATE GENERAL, JOINT RESEARCH CENTRE, EUR 22327 EN, **2006**.
9. Tetko, I. V.; Solov'ev, V. P.; Antonov, A. V.; Yao, X.; Doucet, J.-P.; Fan, B.T.; Hoonakker, F.; Fourches, D.; Jost, P.; Lachiche, L.; Varnek, A. *J. Chem. Inf. Model.* **2006**, 46, 808.
10. Burello, E.; Rothenberg, G. *Adv. Synth. Catal.* **2003**, 345, 1334.
11. Burello, E.; Farriseng, D.; Rothenberg, G. *Adv. Synth. Catal.* **2004**, 346, 1844.
12. Occhipinti, G.; Björsvik, H.-R.; Jensen, V. R. *J. Am. Chem. Soc.* **2006**, 128, 6952.
13. Prieto, O.; Woodward, S. *J. Organomet. Chem.* **2006**, 691, 1515.
14. Chen, J.; Jiwu, W.; Mingzong, L.; You, T. *Journal of Molecular Catalysis A: Chemical* **2006**, 258, 191.
15. Ianni, J. C.; Annamalai, V.; Phuan, P.-W.; Panda, M.; Kozlowski, M. C. *Angew. Chem. Int. Ed.* **2006**, 45, 5502.
16. Chen, J.; Jiwu, W.; Daliang, L.; You, T. *J Mol Model* **2007**, 13, 91-98.
17. Andrussow, L. *Landolt-Börnstein, Zahlenwerte und Funktionen aus Physik, Chemie, Astronomie, Geophysik und Technik*, 6th Ed., vol.2, part 5a, Springer-Verlag, Berlin, Goettingen, Heidelberg, **1969**.
18. Touloukian, Y. S. *Thermophysical properties of matter*, TPRC data series, Vol. **11** Viscosity, Plenum Press: New York, **1975**.
19. Stephan, K.; Lucas, K. *Viscosity of dense fluids*, Plenum Press: New York, **1979**.

20. Daubert, T. E.; Danner, R. P. *Physical and thermodynamic properties of pure chemicals*, Hemisphere Publ.Corp.: New York, **1989**.
21. Viswanath, D. S.; Natarajan, G. *Databook on the viscosity of liquids*, Hemisphere Publ. Corp.: New York, **1989**.
22. Lide, R. D., Ed. *CRC Handbook of chemistry and physics*, 75th Ed., CRC Press: Boca Raton, Ann Arbor, London, Tokyo, **1994**.
23. Barthel, J.; Neueder, R.; Meier, R. *Viscosity of non-aqueous solutions*, DECHEMA Chemistry Data Series Vol. XII, Electrolyte Data Collection, Parts 3, 3a, and 3b, Kreysa, G.; Ed.; DECHEMA: Frankfurt/M. **1997**.
24. Schoff, C. K. *Concentration dependence of the viscosity of dilute polymer solutions: Huggins and Schulz-Blaschke constants*, Polymer Handbook, Brandrup, J., Immergut, E. H.; Grulke, E. A; Eds.; 4th Ed., J. Wiley & Sons, Inc.: New York, **1999**.
25. Wohlfarth, C.; Wohlfahrt, B. *Landolt-Börnstein, Viscosity of Pure Organic Liquids and Binary Liquid Mixtures*, vol 18, subvolume A, Springer-Verlag: Berlin, Heidelberg, **2001**.
26. Škubla, P. *Collect. Czech. Chem. Commun.* **1985**, 50, 1907.
27. Joback, K. G.; Reid, R. C. *Chem. Eng. Commun.* **1987**, 57, 233.
28. Suzuki, T.; Ohtaguchi, K.; Koide, K. *Comput. Chem. Eng.* **1996**, 20, 161.
29. Suzuki, T.; Ebert, R.-U.; Schüürmann, G. *J. Chem. Inf. Comput. Sci.* **1997**, 37, 1122.
30. Kapur, G. S.; Ecker, A.; Meusinger, R. *Energy & Fuels* **2001**, 15, 943.
31. Toropov, A. A.; Nesterov, I. V.; Nabiev, O. M. *Journal of Molecular Structure (Theochem)* **2003**, 637, 37.
32. Ivanciuc, O.; Ivanciuc, T.; Filip, P. A.; Cabrol-Brass, D. *J. Chem. Inf. Comput. Sci.* **1999**, 39, 515.
33. Katritzky, A. R.; Chen, K.; Wang, Y.; Karelson, M.; Lučić, B.; Trinajstić, N.; Suzuki, T.; Schüürmann, G. *J. Phys. Org. Chem.* **2000**, 13, 80.
34. Lučić, B.; Bašić, I.; Nadramija, D.; Miličević, A.; Trinajstić, N.; Suzuki, T.; Petrukhin, R.; Karelson, M.; Katritzky, A. R. *ARKIVOC* **2002**, iv, 45.
35. Cocchi, M.; De Benedetti, P. G.; Seeber, R.; Tassi, L.; Ulrici, A. *J. Chem. Inf. Comput. Sci.* **1999**, 39, 1190.
36. Kauffman, G. W.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **2001**, 41, 408.
37. Artemenko, N. V.; Baskin, I. I.; Palyulin, V. A.; Zefirov, N. S. *Russ. Chem. Bull., Int. Ed.* **2003**, 52, 20.

38. Chalk, A. J.; Beck, B.; Clark, T. A. *J. Chem. Inf. Comput. Sci.* **2001**, 41, 1053.
39. Suzuki, T.; Ebert, R.-U.; Schüürmann, G. *J. Chem. Inf. Comput. Sci.* **2001**, 41, 776.
40. Halberstam, N. M.; Baskin, I. I.; Palyulin, V. A.; Zefirov, N. S. *Dokl. Chem* **2002**, 384, 140.
41. Zefirov, N. S.; Kirpichenok, M. A.; Izmailov, F. F.; Trofimov, M. I. *Dokl. Akad. SSSR* **1987**, 296, 883.
42. Sanderson, R. T. en: *Encyclopedia of Chemistry*, 3^a Ed., Hampel, C. A.; Hawley, G. G.; Eds.; Van Nostrand Reihold: New York, **1973**, 385.
43. Kier, L. B.; Hall, L. H. *Eur. J. Med. Chem.* **1977**, 12, 307.
44. Shannon, C.; Weaver, W. *The Mathematical Theory of Communication*, University of Illinois Press: Urbana (ILL), **1949**.
45. Bonchev, D. *Information Theoretic Indices for Characterization of Chemical Structures*, Research Studies Press: Chichester (UK), **1983**.
46. Osmialowski, K.; Halkiewicz, J.; Kaliszan, R. *J. Chromatogr.* **1986**, 63, 361.
47. Csizmadia, I. G. *Theory and Practice of MO Calculations on Organic Molecules*, Elsevier: Amsterdam, **1976**.
48. Sannigrahi, A. B. *Adv. Quant. Chem.* **1992**, 23, 301.
49. Randic, M. *J. Am. Chem. Soc.* **1975**, 97, 6609.
50. Girifalco, L. A.; Good, R. J. *J. Phys. Chem.* **1957**, 61, 904.
51. Masterson, W. L.; Slowinski, E. J. *Chemical Principles*, W. B. Saunders: Philadelphia, **1973**.
52. LeGrand, D. J.; Gaines, G. L. *Colloid Interface Sci.* **1975**, 51, 338.
53. Adamson, A. W. *Physical chemistry of surfaces*, 5th Ed., J. Wiley & Sons: New York, **1990**.
54. Schäfer, K.; Lax, E. *Landolt-Börnstein, Zahlenwerte und Funktionen aus Physik, Chemie, Astronomie, Geophysik und Technik*, 6th Ed., vol.2, part 3, Springer-Verlag, Berlin, Goettingen, Heidelberg, **1969**.
55. Jasper, J. J. *Phys. Chem. Ref. Data* **1972**, 1, 841.
56. McLure, I. A.; Pegg, I. L.; Soares, V. A. M. *Colloid Sci.* **1983**, 4, 283.
57. Wu, S. *Surface and interfacial tension of polymers, oligomers, plasticizers, and organic pigments. Polymer Handbook*, 3rd Ed.; Brandrup, I.; Immergut, E. H.; Eds.; J. Wiley & Sons: New York, Chichester, Brisbane, Toronto, Singapore, **1989**.
58. Yaws, C. L. *Thermodynamic and physical property data*, Gulf. Publ. Comp., **1992**.

59. Wohlfarth, C.; Wohlfahrt, B. *Landolt-Börnstein, Surface Tension of Pure Liquids and Binary Liquid Mixtures*, vol 16, Springer-Verlag: Berlin, Heidelberg, **2001**.
60. Stanton, D.; Jurs, P. C. *Anal. Chem.* **1990**, 62, 2323.
61. Stanton, D.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **1992**, 32, 109.
62. Bonchev, D. *J. Chem. Inf. Comput. Sci.* **2000**, 40, 934.
63. Liu, S.; Cai, S.; Cao, S.; Li, S. *J. Chem. Inf. Comput. Sci.* **2000**, 40, 1337.
64. Shamsipur, M.; Ghavami, R.; Hemmateenejad, B.; Sharghi, H. *QSAR Comb. Sci.* **2004**, 23, 734.
65. Wang, J.; Du, H.; Liu, H.; Yao, X.; Hu, Z.; Fan, B. T. *Talanta* **2007**, 73, 147.
66. Osmialowski, K.; Halkiewicz, J.; Radecki, A.; Kaliszan, R. *J. Chromatogr.* **1985**, 346, 53.
67. Clementi, E. *Computational Aspects of Large Chemical Systems*, Springer Verlag: New York, **1980**.
68. Benson, S. W.; Buss, J. H. *J. Chem. Phys.* **1958**, 29, 546.
69. Domalsky, E. S.; Hearing, E. D.; Hearing Jr. V. J. *NIST Estimation of the Chemical Thermodynamic Properties for Organic Compounds at 298.15 K*; NIST Standard Reference Database 18; National Institute of Standards and Technology: Gaithersburg, **1994**.
70. Laidler, K. J. *Can. J. Chem.* **1956**, 34, 626.
71. Leal, J. P. *J. Phys. Chem. Ref. Data* **2006**, 35, 55.
72. Pedley, J. B. *Thermochemical Data and Structures of Organic Compounds*. Vol I. TRC, **1994**.
73. Pilcher, G.; Skinner, H. A. *Thermochemistry of Organometallic Compounds*; En: *The Chemistry of the Metal-Metal Bond*; Hartley, F. R.; Patai, S.; Eds.; Wiley: New York, **1982**; capitol 2.
74. Luo, Y.-R.; Benson, S. W. *Acc. Chem. Res.* **1992**, 25, 375.
75. Luo, Y.-R.; Benson, S. W. *J. Phys. Chem.* **1989**, 93, 4643.
76. Luo, Y.-R.; Benson, S. W. *J. Phys. Chem.* **1989**, 93, 3791.
77. Becerra, R.; Walsh, R. *J. Phys. Chem. A* **1997**, 101, 8959.
78. Dias, A. R.; Simoes, J. A. M.; Teixeira, C.; Airoidi, C.; Chagas, A. P. *J. Organomet. Chem.* **1987**, 335, 71.
79. Katritzky, A. R.; Lobanov, V. S.; Karelson, M. *J. Chem. Inf. Comput. Sci.* **2000**, 40, 1.

80. Ivanciuc, O.; Ivanciuc, T.; Cabrol-Brass, D.; Balaban, A. T. *J. Chem Inf. Comput. Sci.* **2000**, 40, 631.
81. Ivanciuc, O.; Ivanciuc, T.; Klein, D. J.; Seitz, W. A.; Balaban, A. T. *J. Chem. Inf. Comput. Sci.* **2001**, 41, 536.
82. Yan, A. *J. Chem. Inf. Model.* **2006**, 46, 2299.
83. Simões, J. A. M. *Organometallic Thermochemistry Data*; En: *NIST Chemistry WebBook; NIST Standard Reference Database Number 69*; Linstrom, P. J.; Mallard, W. G.; Eds.; National Institute of Standards and Technology: Gaithersburg, June **2005** (<http://webbook.nist.gov>).
84. Cox, J. D.; Pilcher, G. *Thermodynamic of Organic and Organometallic Compounds*, Academic Press, **1970**.
85. Emsley, J. *The Elements*, Clarendon Press, **1989**.
86. Kier, L. B. in: *Computational Chemical Graph Theory*, Rouvray, D. H.; Ed.; Nova Science Publishers: New York, **1990**.
87. Katritzky, A. R.; Mu, L.; Lobanov, V. S.; Karelson, M. *J. Phys. Chem.* **1996**, 100, 10400.
88. Hunter, E. P. L.; Lias, S. G. *J. Phys. Chem. Ref. Data* **1998**, 27, 413.
89. Burk, P.; Koppel, I. A.; Koppel, I.; Kurg, R.; Gal, J.-F.; Maria, P.-C.; Herreros, M.; Notario, R.; Abboud, J.-L. M.; Anvia, F.; Taft, R. W. *J. Phys. Chem. A* **2000**, 104, 2824.
90. Marshall, A. G. *Acc. Chem. Res.* **1996**, 29, 307.
91. Castleman, A. W.; Keesee, R. G. *Chem. Rev.* **1986**, 86, 589.
92. Cooks, R. G.; Patrick, J. S.; Kotiaho, T.; McLukey, S. A. *Mass Spectrom. Rev.* **1994**, 13, 287.
93. Cooks, R., P. S. H. Wong *Acc. Chem. Res.* **1998**, 31, 379.
94. Rodgers, M. T.; Armentout, P. B. *Mass Spectrom. Rev.* **2000**, 19, 215.
95. Talmm, K.; Fara, D. C.; Katritzky, A. R.; Burk, P.; Karelson, M. *J. Phys. Chem. A* **2004**, 108, 4812.
96. Mo, O.; Yañez, M.; Gal, J.-F.; Maria, P. C.; Decouzon, M. *Chem. Eur. J.* **2003**, 9, 4330.
97. Gal, J.-F.; Maria, P.-C.; Decouzon, M. *Int. J. Mass. Espectrom.* **2002**, 217, 75.
98. Feng, W. Y.; Gronert, S.; Lebrilla, C. *J. Phys. Chem. A* **2003**, 107, 405.
99. Gal, J.-F.; Maria, P. C.; Decouzon, M.; Mo, O.; Yañez, M.; Abboud, J.-L. M. *J. Am. Chem. Soc.* **2003**, 125, 10394.