

UCH-FC  
MAG-Q  
N 415  
C.1



**PREDICCIÓN DE LA ADSORCIÓN DE  
HERBICIDAS EN SUELOS DERIVADOS DE  
CENIZAS VOLCÁNICAS MEDIANTE LA  
CREACIÓN DE MODELOS (Q)SAR:  
IMPLICANCIAS PARA LA TOMA DE DECISIONES  
EN EL MARCO AMBIENTAL CHILENO**

**Tesis**

**Entregada A La  
Universidad de Chile  
En Cumplimiento Parcial De Los Requisitos  
Para Optar Al Grado De**

**Magíster en Ciencias Químicas**

**Facultad de Ciencias**

**Por**

**ÁNGELO JAVIER NEIRA ALBORNOZ**

**Marzo, 2017**

**Directora de Tesis: Dra. Lizethly Cáceres Jensen**

**Co-Directora de Tesis: Dra. Marcela Urzúa Acevedo**

**FACULTAD DE CIENCIAS**  
**UNIVERSIDAD DE CHILE**  
**INFORME DE APROBACIÓN**  
**TESIS DE MAGÍSTER**

Se informa a la Escuela de Postgrado de la Facultad de Ciencias que la Tesis de Magíster presentada por el candidato

**ÁNGELO JAVIER NEIRA ALBORNOZ**

Ha sido aprobada por la comisión de Evaluación de la tesis como requisito para optar al grado de Magíster en Ciencias Químicas, en el examen de Defensa Privada de Tesis rendido el día 26 de septiembre de 2016.

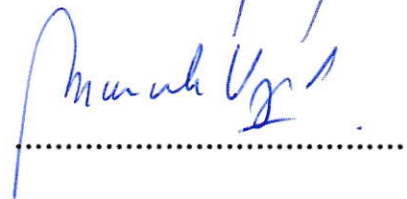
**Directora de Tesis:**

**Dra. Lizethly Cáceres Jensen.**



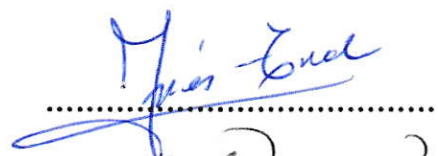
**Co-Directora de Tesis:**

**Dra. Marcela Urzúa Acevedo.**



**Comisión de Evaluación de la Tesis:**

**Prof. María Inés Toral (Presidenta)**



**MSc. Inés Ahumada Torres.**



**Dr. Waldo Quiroz Venegas.**



“Comprendo que aun cuando pueda a través de la ciencia, captar los fenómenos y enumerarlos, no por ello puedo aprehender el mundo”

Albert Camus, *El mito de Sísifo*.

Dedico esta tesis a los lectores que, con ojos críticos, la analicen y utilicen humildemente para mejorar poco a poco la comprensión de nuestros suelos.

## AGRADECIMIENTOS

En mi afán indestructible de clasificarlo todo, dedico estos agradecimientos en función de la cercanía que han tenido los mencionados con mi trabajo de tesis.

En primer lugar, al Laboratorio de Físicoquímica & Analítica de la UMCE, su gente y experiencias. A los profesores Lizethly y Jorge, por su apoyo, dedicación y orientación. A los estudiantes que allí convivimos, por el humor, las distracciones y conversaciones agradables. Por todos ustedes y su profesión convertida en estilo de vida, solo puedo sentir admiración.

A mi co-directora de tesis y la comisión evaluadora, que me han apoyado mejorando la comprensión y estructuración de este trabajo. Adicionalmente, a la profesora María Inés Toral por su ayuda frente a las deficiencias administrativas de mi querida universidad.

A mi familia y amigos que, como siempre, han sabido calmar mis exacerbaciones y avivar mis pasiones en los momentos indicados. A todos, gracias por acompañarme en esta gloria efímera, una más de tantas otras.

Finalmente, espero que el espíritu de esta tesis sirva como medicina para quienes padecen de una historia ambiental conflictiva, particularmente donde nuestro escudo nacional se jacta más de la fuerza que de la razón. Espero que vean en ella un llamado a la cooperación, la mesura y el respeto como condiciones necesarias para el desarrollo sostenible, y motivar así su rebelión contra este mundo competitivo y perturbado.

## ÍNDICE DE MATERIAS

<b>LISTA DE TABLAS .....</b>	<b>VII</b>
<b>LISTA DE FIGURAS.....</b>	<b>VIII</b>
<b>LISTA DE SÍMBOLOS Y ABREVIATURAS.....</b>	<b>IX</b>
<b>RESUMEN.....</b>	<b>X</b>
<b>ABSTRACT.....</b>	<b>XII</b>
<b>1. INTRODUCCIÓN .....</b>	<b>1</b>
<b>1.1. Herbicidas e impacto ambiental.....</b>	<b>1</b>
<b>1.2. Suelos derivados de cenizas volcánicas .....</b>	<b>2</b>
<b>1.3. Contexto regulatorio y modelos (Q)SAR .....</b>	<b>4</b>
<b>2. ESTADO DEL ARTE.....</b>	<b>8</b>
<b>2.1. Predicción del coeficiente de adsorción de herbicidas mediante modelos (Q)SAR .....</b>	<b>8</b>
2.1.1. Generación de un set de datos de calidad.....	10
2.1.2. Selección de un <i>endpoint</i> representativo .....	11
2.1.3. Descriptores empleados en modelos (Q)SAR en suelos .....	13
2.1.4. Tratamiento estadístico de los datos y descriptores.....	15
2.1.5. Sets de entrenamiento y de prueba .....	16
2.1.6. Ecuación de regresión .....	17
2.1.7. Validaciones interna y externa .....	17
2.1.8. Dominio de aplicabilidad .....	19
2.1.9. Interpretación mecanística.....	21
<b>2.2. Propuestas para facilitar la validación de modelos (Q)SAR con propósitos regulatorios.....</b>	<b>21</b>
2.2.1. Transferibilidad y criterios OCDE como condición necesaria para la validación de modelos (Q)SAR.....	21
2.2.2. Contextualización y validación empírica mediante la aplicación del ciclo PDCA en modelos (Q)SAR.....	23
2.2.3. Esquema de creación de modelos (Q)SAR con propósitos regulatorios.....	25
<b>3. HIPÓTESIS Y OBJETIVOS.....</b>	<b>28</b>

<b>3.1. Propuesta de investigación .....</b>	<b>28</b>
<b>3.2. Hipótesis .....</b>	<b>28</b>
<b>3.3. Objetivo general .....</b>	<b>29</b>
<b>3.4. Objetivos específicos.....</b>	<b>29</b>
<b>4. MATERIALES Y MÉTODOS.....</b>	<b>31</b>
<b>4.1. Herbicidas y suelos .....</b>	<b>31</b>
<b>4.2. Generación de modelos (Q)SAR en SDCV .....</b>	<b>32</b>
4.2.1. Selección de un <i>endpoint</i> representativo .....	32
4.2.2. Predicción de valores de <i>endpoint</i> empleando modelos (Q)SAR para SCP .....	32
4.2.3. Cálculo de descriptores para generar modelos (Q)SAR en SDCV .....	33
4.2.4. Análisis exploratorio de la matriz de descriptores.....	33
4.2.5. Descriptores de la interacción herbicida-SDCV .....	34
4.2.6. Generación de bases de datos para el análisis de los valores de <i>endpoint</i> de herbicidas en SDCV .....	34
4.2.7. Calidad estadística de las bases de datos .....	35
4.2.8. División en sets de entrenamiento y de prueba .....	35
4.2.9. Selección de las mejores combinaciones de descriptores.....	36
4.2.10. Generación de la ecuación de regresión lineal .....	36
<b>4.3. Validación de modelos (Q)SAR.....</b>	<b>36</b>
<b>4.4. Transferibilidad y criterios OCDE .....</b>	<b>37</b>
<b>4.5. Dominio de Aplicabilidad e interpretación mecanística .....</b>	<b>38</b>
4.5.1. Detección de valores atípicos y puntos de influencia.....	38
4.5.2. Interpretación y aplicabilidad del modelo en condiciones ambientales .....	38
<b>5. RESULTADOS Y DISCUSIÓN.....</b>	<b>40</b>
<b>5.1. Selección de un <i>endpoint</i> representativo .....</b>	<b>40</b>
<b>5.2. Ajuste de los valores de <i>endpoint</i> de herbicidas en SDCV empleando modelos (Q)SAR para SCP .....</b>	<b>41</b>
5.2.1. Modelo de Bintein y Devillers .....	41
5.2.2. Modelo de Sabljic y col.....	45
5.2.3. Modelo de Wen y col. ....	48
<b>5.3. Análisis exploratorio de los descriptores propuestos .....</b>	<b>50</b>
<b>5.4. Generación de descriptores para la interacción herbicida-SDCV .....</b>	<b>53</b>
<b>5.5. Análisis de los valores de <i>endpoint</i> de herbicidas en SDCV.....</b>	<b>55</b>
5.5.1. Variabilidad de los valores de <i>endpoint</i> en función del herbicida.....	55

5.5.2. Variabilidad de los valores de <i>endpoint</i> en función de SDCV .....	60
5.5.3. Variabilidad de los valores de <i>endpoint</i> de herbicidas no ionizables en SDCV bajo la aplicación de $\log K_f$ y $\log K_{oc}$ .....	61
<b>5.6. Tratamiento estadístico de las bases de datos empleadas en la generación de modelos (Q)SAR para adsorción de herbicidas en SDCV .....</b>	<b>63</b>
<b>5.7. División de los valores de <i>endpoint</i> entre set de entrenamiento y set de prueba .....</b>	<b>66</b>
<b>5.8. Creación y validación de modelos (Q)SAR para predecir la adsorción de herbicidas en SDCV .....</b>	<b>66</b>
5.8.1. Calidad de modelos (Q)SAR para predecir adsorción de herbicidas no ionizables .....	66
5.8.2. Calidad de modelos (Q)SAR para predecir adsorción de herbicidas ácidos .....	70
5.8.3. Calidad de modelos (Q)SAR para predecir n de herbicidas no ionizables .....	72
<b>5.9. Transferibilidad y criterios OCDE en la generación de modelos QSAR para la predicción de <math>\log K_f</math> y n de herbicidas en SDCV .....</b>	<b>74</b>
<b>5.10. Dominio de aplicabilidad e interpretación mecanística de los modelos QSAR creados para predecir adsorción de herbicidas en SDCV.....</b>	<b>75</b>
5.10.1. Detección de valores atípicos y puntos de influencia.....	75
5.10.2. Interpretación mecanística de los modelos (Q)SAR generados para la adsorción de herbicidas no ionizables en SDCV .....	75
5.10.3. Interpretación mecanística de los modelos (Q)SAR generados para la adsorción de herbicidas ácidos en SDCV .....	83
5.10.4. Interpretación mecanística de los modelos (Q)SAR generados para predecir n de herbicidas no ionizables en SDCV .....	86
<b>5.11. Aplicabilidad de los modelos (Q)SAR generados para predecir la adsorción de herbicidas en SDCV.....</b>	<b>87</b>
5.11.1. Aplicabilidad en condiciones ambientales .....	87
5.11.2. Aplicabilidad en el contexto chileno .....	90
<b>5.12. Discusión metodológica .....</b>	<b>92</b>
<b>6. CONCLUSIONES .....</b>	<b>97</b>
<b>7. BIBLIOGRAFÍA.....</b>	<b>99</b>
<b>ANEXO A: DEFICIENCIAS EN LA CREACIÓN DE MODELOS (Q)SAR.....</b>	<b>107</b>
<b>ANEXO B: BASE TEÓRICO/CONCEPTUAL PARA LA GENERACIÓN Y VALIDACIÓN DE DATOS Y <i>ENDPOINT</i> .....</b>	<b>113</b>
<b>ANEXO C: PROPIEDADES Y CARACTERIZACIÓN DE HERBICIDAS Y SUELOS EMPLEADOS EN ESTE TRABAJO .....</b>	<b>116</b>

<b>ANEXO D: VALIDACIÓN DE LOS VALORES EMPÍRICOS EMPLEADOS EN EL SET DE DATOS .....</b>	<b>120</b>
<b>ANEXO E: DESCRIPCIÓN DE LA METODOLOGÍA Y SOFTWARE USADOS PARA LA CREACIÓN DE MODELOS (Q)SAR PARA PREDECIR COEFICIENTES DE ADSORCIÓN DE HERBICIDAS EN SDCV .....</b>	<b>125</b>
<b>ANEXO F: GENERACIÓN DE DESCRIPTORES PARA LA INTERACCIÓN HERBICIDA-SDCV .....</b>	<b>134</b>
<b>ANEXO G: CALIDAD ESTADÍSTICA DE LOS MODELOS (Q)SAR GENERADOS .....</b>	<b>146</b>



## LISTA DE TABLAS

<b>Tabla 1.</b> Efecto del cambio de escala y la normalización al contenido de CO en la variabilidad de los valores de endpoint .....	<b>62</b>
<b>Tabla 2.</b> Análisis de la reducción de descriptores en la generación de modelos QSAR para la predicción de la adsorción de herbicidas en SDCV.....	<b>65</b>
<b>Tabla A1.</b> Origen y tratamiento de los datos usados en la generación de modelos (Q)SAR .....	<b>108</b>
<b>Tabla A2.</b> Modelos (Q)SAR para la adsorción de contaminantes orgánicos en suelos.....	<b>110</b>
<b>Tabla A3.</b> Problemas recurrentes y aspectos a evaluar en la realización de modelos (Q)SAR para adsorción de contaminantes en suelos.....	<b>112</b>
<b>Tabla C1.</b> Herbicidas usados en el presente estudio .....	<b>117</b>
<b>Tabla C2.</b> Propiedades fisicoquímicas principales de SDCV usados en el presente estudio .....	<b>118</b>
<b>Tabla C3.</b> Composición mineralógica de los SDCV .....	<b>119</b>
<b>Tabla E1.</b> Parámetros de calidad estadística para la bondad del ajuste de modelos (Q)SAR.....	<b>130</b>
<b>Tabla E2.</b> Parámetros de calidad estadística para la robustez del modelo (Q)SAR.....	<b>131</b>
<b>Tabla E3.</b> Parámetros de calidad estadística para la predictividad del modelo (Q)SAR.....	<b>132</b>
<b>Tabla E4.</b> Parámetros de calidad estadística para la comparación entre modelos (Q)SAR.....	<b>133</b>
<b>Tabla F1.</b> Principales factores que afectan el proceso de adsorción de herbicidas en SDCV .....	<b>135</b>
<b>Tabla G1.</b> Calidad estadística de los modelos (Q)SAR para predecir $\log K_f$ o $K_f$ de herbicidas no ionizables. ....	<b>147</b>
<b>Tabla G2.</b> Calidad estadística de los modelos (Q)SAR para predecir $\log K_f$ o $K_f$ de herbicidas ácidos .....	<b>148</b>
<b>Tabla G3.</b> Calidad estadística de los modelos (Q)SAR para predecir n de herbicidas no ionizables .....	<b>149</b>

## LISTA DE FIGURAS

<b>Figura 1.</b> Esquema de generación de modelos (Q)SAR para la adsorción de herbicidas en SDCV y su contexto socioambiental.....	<b>7</b>
<b>Figura 2.</b> Esquema general de la creación de modelos (Q)SAR para la predicción de la adsorción de contaminantes orgánicos en suelos .....	<b>9</b>
<b>Figura 3.</b> Esquema para realizar modelos (Q)SAR para predecir adsorción de herbicidas en SDCV...	<b>26</b>
<b>Figura 4.</b> Ciclo de mejora continua propuesto para modelos (Q)SAR para predecir adsorción de herbicidas en SDCV .....	<b>27</b>
<b>Figura 5.</b> Ajuste comparativo del modelo de Bintein y Devillers y los valores de $\log K_f$ del presente estudio.....	<b>44</b>
<b>Figura 6.</b> Ajuste comparativo del modelo de Sabljic y col. en función de $K_{ow}$ y los valores de $\log K_f$ del presente estudio.....	<b>47</b>
<b>Figura 7.</b> Ajuste comparativo del modelo de Wen y col. y los valores de $\log K_f$ del presente estudio	<b>51</b>
<b>Figura 8.</b> Valores de <i>endpoint</i> ( $\log K_f$ y $n$ ) en función de cada herbicida y SDCV.....	<b>56</b>
<b>Figura 9.</b> Valores de $K_f$ y $n$ de todos los herbicidas en SDCV en función del tipo de herbicida .....	<b>57</b>
<b>Figura 10.</b> Ajuste de las ecuaciones 26, 27, 28 y 30 para adsorción de herbicidas no ionizables (intervalo de aplicación $\log K_f$ : 0,66 - 2,02).....	<b>76</b>
<b>Figura 11.</b> Ajuste de las ecuaciones 31, 32, 34 y 35 para adsorción de herbicidas ácidos (intervalo de aplicación $\log K_f$ : 0,05 - 1,72).....	<b>77</b>
<b>Figura 12.</b> Ajuste de las ecuaciones 36 y 37 para $n$ en herbicidas no ionizables (intervalo de aplicación $n$ : 0,28 - 1,57) .....	<b>78</b>
<b>Figura 13.</b> Ajuste de las ecuaciones 26, 27 y 30 para la adsorción de herbicidas ácidos y básicos.....	<b>81</b>
<b>Figura 14.</b> Uso complementario de modelos (Q)SAR generados en este trabajo y su contexto en el ciclo PDCA .....	<b>95</b>

## LISTA DE SÍMBOLOS Y ABREVIATURAS

### ABREVIATURAS HERBICIDAS:

2,4-D:	Ácido 2,4-diclorofenoxiacético
AT:	Atrazina
CS:	Clorsulfurón
DI:	Diurón
GFS:	Glifosato
MBT:	Metabentiazurón
MC:	Metolacloro
MMS:	Metsulfurón-metil
NS:	Nicosulfurón
PC:	Picloram
SM:	Simazina
TF:	Trifluralín

### ABREVIATURAS DE SUELOS:

COLL:	Collipulli
DIG:	Diguillín
FRE:	Freire
FRU:	Frutillar
MET:	Metreco
NBR:	Nueva Braunau
OSN:	Osorno
STB:	Santa Bárbara
TCO:	Temuco
RAL:	Ralún

### SIGLAS:

ACP:	Análisis de Componentes Principales
CIC:	Capacidad de Intercambio Catiónico
CO:	Carbono Orgánico
MBD:	Modelo de Bintein y Devillers
MO:	Materia Orgánica
MS:	Modelo de Sabljic y col.
MW:	Modelo de Wen y col.
OCDE:	Organización para la Cooperación y el Desarrollo Económico
(Q)SAR:	Relación (Cuantitativa) Estructura-Actividad
SCP:	Suelos de Carga Permanente
SDCV:	Suelos Derivados de Cenizas Volcánicas
SU:	Sulfonilureas

### SÍMBOLOS:

$K_d$ :	Coefficiente de partición suelo/solución
$K_f$ :	Coefficiente de adsorción de <i>Freundlich</i>
$K_{oc}$ :	Coefficiente de partición suelo/solución normalizado al contenido de CO
$K_{ow}$ :	Coefficiente de partición octanol/agua
$PIE_{suelo}$ :	Punto isoeléctrico del suelo
$S_w$ :	Solubilidad en agua

## RESUMEN

Los modelos (Q)SAR para predecir adsorción de herbicidas en suelos han sido propuestos por la OCDE como apoyo a Estudios de Impacto Ambiental, optimizando tiempo y costo asociados a la toma racional de decisiones en materia medioambiental. Sin embargo, este tipo de modelos aun no se ha implementado en países con suelos derivados de cenizas volcánicas (SDCV), como Chile. Este trabajo tuvo como propósito la creación de modelos (Q)SAR para predecir la adsorción de herbicidas ionizables y no ionizables en SDCV, utilizando datos de calidad obtenidos a partir de (i) estudios cinéticos e isothermas de adsorción usando la misma metodología en todos los casos, (ii) descarte de variables confundentes y (iii) representatividad del *endpoint* seleccionado, en este caso  $\log K_f$  y  $n$ . La metodología implicó (i) revisión bibliográfica de modelos ya existentes, (ii) contextualización del sistema herbicida-SDCV, (iii) generación de descriptores moleculares, de SDCV y de la interacción herbicida-SDCV, (iv) validación estadística de bases de datos, (v) uso de regresión lineal múltiple para la creación de modelos (Q)SAR, (vi) validación estadística del modelo según bondad del ajuste, robustez, predictividad y transferibilidad, (vii) dominio de aplicabilidad e interpretación mecanística de los modelos construidos, y (viii) aplicabilidad de estos modelos en el marco legal

chileno. A partir de la revisión bibliográfica, se detectaron deficiencias en la generación de modelos (Q)SAR para predecir la adsorción en suelos de carga permanente, que fueron relacionadas con (i) predicción en condiciones locales de cada suelo, (ii) interacciones herbicida-suelo, (iii) precisión de la predicción sin evaluar su confianza, (iv) calidad estadística, (v) transferibilidad y (vi) aplicabilidad en condiciones reales. Las deficiencias antes mencionadas fueron consideradas durante la generación de modelos (Q)SAR, con el objetivo de evitar dichos errores. Adicionalmente, se propuso el ciclo PDCA como apoyo a la contextualización, generación, evaluación y validación de estos modelos (Q)SAR, con la finalidad de mejorar su evaluación en condiciones reales (propósitos regulatorios) y fomentar su mejora continua en la detección de falencias (perfectibilidad). Los modelos (Q)SAR creados en este trabajo para predecir  $\log K_f$  de herbicidas no ionizables en SDCV fueron los únicos estadísticamente confiables, pero presentaron alta incerteza en sus coeficientes, limitando su aplicación con propósitos regulatorios. La interpretación de dichos modelos mostró una adsorción favorecida en SDCV con menor contenido de arcilla y  $Fe_{DCB}$ , y herbicidas con mayor densidad de átomos, de lo cual se desprende la relevancia de herbicidas y SDCV en el proceso de adsorción. Finalmente, la única forma de aplicar modelos (Q)SAR con propósitos regulatorios en Chile es la participación del Estado, particularmente el Ministerio del Medio Ambiente, en el financiamiento, creación y aplicación de estos modelos como complemento en Estudios de Impacto Ambiental, debido a la complejidad de evaluar sin líneas de base la relevancia del proceso de adsorción en condiciones reales y su correspondencia con la metodología empleada en la generación de datos para el modelo (Q)SAR.

## ABSTRACT

The (Q)SAR models to predict sorption of herbicides on soils have been proposed by OCDE as support for Environmental Impact Assessment, optimizing cost and time associated to rational decision making in environmental matters. However, this kind of models has not yet been implemented in countries with volcanic ash derived soils (VADS), like Chile. This study aimed the creation of (Q)SAR models to predict sorption of ionizable and non-ionizable herbicides on VADS, using quality data obtained from (i) kinetics and isotherm sorption studies using the same methodology in all cases, (ii) discarding of confounding variables and (iii) representativeness of selected endpoint, in this case  $\log K_f$  and  $n$ . The methodology involved (i) literature review of existing models, (ii) contextualization of herbicide-VADS system, (iii) generation of molecular, VADS and herbicide-VADS interaction descriptors, (iv) statistical validation of databases, (v) use of multiple linear regression for (Q)SAR model creation, (vi) statistical validation of model through goodness-of-fit, robustness, predictivity and transferability, (vii) applicability domain and mechanistic interpretation of built models, and (viii) applicability of these models in the Chilean legal framework. From the literature review, deficiencies were detected in the generation of (Q)SAR models to predict sorption on permanent charge soils,

which was related to (i) prediction on local conditions for each soil, (ii) herbicide-soil interactions, (iii) prediction accuracy without evaluate their confidence, (iv) statistical quality, (v) transferability and (vi) applicability in real conditions. The aforementioned deficiencies were considered during the generation of (Q)SAR models, in order to avoid such errors. In addition, it was proposed the PDCA cycle as support for contextualization, generation, evaluation and validation of these (Q)SAR models, in order to improve their evaluation in real conditions (regulatory purposes) and promote their continuous improvement in the detection of shortcomings (perfectibility). The (Q)SAR models created in this work to predict  $\log K_f$  for non-ionizable herbicides on VADS were the only statistically reliable, but showed high uncertainty in their coefficients, limiting their application with regulatory purposes. The interpretation of these models showed a favored sorption in VADS with lower clay and  $Fe_{DCB}$  content, and herbicides with higher atom density, whereof it follows the relevance of herbicides and VADS in sorption process. Finally, the only way to apply (Q)SAR models with regulatory purposes in Chile is the participation of the State, particularly the *Ministerio del Medio Ambiente*, in financing, creation and application of these models as a complement in Environmental Impact Assessment, due to the complexity of assessing without baselines the relevance of sorption process in real conditions and their correspondence with the methodology used in the generation of data for the (Q)SAR model.

# 1. INTRODUCCIÓN

## 1.1. Herbicidas e impacto ambiental

Los herbicidas son agroquímicos usados con la finalidad de mejorar la productividad silvoagrícola. Actúan a nivel metabólico, afectando negativamente a malezas. Su interés de estudio radica en que la mayoría de los herbicidas aplicados al suelo son ionizables, por lo cual son frecuentemente detectados en aguas superficiales y subterráneas a nivel mundial [1]. Por otra parte, los herbicidas no ionizables persisten por largos períodos de tiempo en el ambiente y pueden ser acumulados hasta altos niveles tróficos [2].

Los herbicidas pueden ser selectivos o no selectivos, lo que se relaciona con sus dosis de aplicación. Por ejemplo, las sulfonilureas (SU) son herbicidas selectivos usados en bajas dosis de aplicación para el control de malezas en cultivos de cereales, mientras que Glifosato (GFS) es un herbicida no selectivo usado en altas dosis de aplicación para el control de todo tipo de cultivos.

La ventaja del uso de herbicidas selectivos es su reducido impacto al ambiente y organismos no blanco. Sin embargo, los herbicidas no selectivos son más baratos, son útiles bajo numerosas condiciones de cultivo, sirven para controlar todo tipo de



malezas y no exigen condiciones de uso o cultivos específicos al momento de su aplicación.

En Chile, la regulación en el uso de plaguicidas le corresponde al Servicio Agrícola y Ganadero (SAG). Si bien, el uso de plaguicidas en el país ha disminuido a través de los años [3, 4], el uso de herbicidas ha presentado un aumento en términos porcentuales, alcanzando un 20% del total de plaguicidas vendidos según su último registro [4].

En países desarrollados, como Estados Unidos, el uso de herbicidas específicos ha ido en aumento de manera complementaria a tratamientos como el control integrado de plagas y rotación de cultivos [5], siendo las SU uno de los herbicidas más prometedores debido a su baja toxicidad, rápida degradación, bajas dosis de aplicación y gran actividad herbicida [6]. Sin embargo, en Chile aun se prefieren herbicidas no específicos, como GFS, e incluso no ionizables, como Diurón (DI) [4].

El comportamiento y destino ambiental de herbicidas en el ambiente es función de procesos edáficos, tales como: adsorción, degradación, transporte, etc. El proceso de adsorción es importante al momento de evaluar la toxicidad y calidad del ambiente, pues da cuenta de la capacidad depuradora del suelo, así como la potencial lixiviación de herbicidas hacia la napa freática, afectando a su volatilización, biodisponibilidad, bioactividad, fitotoxicidad y degradación [7].

## **1.2. Suelos derivados de cenizas volcánicas**

Los suelos derivados de cenizas volcánicas (SDCV) son suelos de carga variable, denominados así debido al elevado contenido de óxidos de Fe, que les confiere una

densidad de carga superficial dependiente del pH [8]. Caso contrario ocurre con los suelos de carga permanente (SCP), que presentan densidad de carga independiente del pH.

Los SDCV son propios de zonas volcánicas, representan el 0,84% de la superficie terrestre [9] y el 70% de la actividad agrícola de Chile. Entre sus características destacan: alto contenido de materia orgánica (MO), alto contenido de óxidos de Fe y Al, alta capacidad de intercambio catiónico (CIC) y alta superficie específica proporcionada por la presencia de minerales amorfos, tales como Alofán [10].

El mayor impacto de los óxidos de Fe en SDCV radica en la posibilidad de presentar cargas superficiales positivas incluso en condiciones de densidad de carga total negativa, afectando la interacción herbicida-SDCV [8, 11]. En particular, en SDCV aumenta la adsorción de herbicidas ácidos respecto a SCP debido a interacciones atractivas entre la forma aniónica de dichos herbicidas y los óxidos de Fe de carga positiva, compensando parte de la repulsión electrostática producida en minerales y carbono orgánico (CO) [7].

De esta forma, la dinámica de herbicidas en SDCV depende del grado de ionización de herbicidas y del punto isoeléctrico del suelo ( $PIE_{\text{suelo}}$ ). En este sentido, la retención de aniones en SDCV sugiere una menor lixiviación y por lo tanto un menor impacto ambiental por contaminación de napas subterráneas asociado al uso de herbicidas ácidos en nuestro país, en comparación a países con SCP.

Con base en lo anterior, la realización de estudios de impacto ambiental de herbicidas en SDCV puede estimarse a través del proceso de adsorción y debería

considerar las diferentes propiedades herbicida-suelo para su aplicación con propósitos regulatorios.

### **1.3. Contexto regulatorio y modelos (Q)SAR**

En legislación ambiental, existen dos principios que promueven la aplicación de estudios de impacto ambiental [12]:

- (i) Principio precautorio, aplicado cuando no existe claridad respecto al posible impacto ambiental de una actividad humana.
- (ii) Principio preventivo, usado cuando el impacto ambiental es un hecho, pero se desconoce su magnitud.

En ambos casos, la cuantificación del impacto ambiental es llevada a cabo a través de metodologías científicas, aplicación de instrumentos de gestión y políticas ambientales, con la finalidad de asegurar una reacción temprana frente a posibles daños ambientales sin limitar el desarrollo socioeconómico del país.

Sin embargo, la aplicación de lo mencionado anteriormente resulta costosa y requiere gran consumo de tiempo, lo que ralentiza la toma de decisiones. De este modo, si bien la aplicación de ambos principios de la legislación ambiental sirve como apoyo al desarrollo sostenible, también podría ser vista como una limitante para el desarrollo económico.

Una forma de generar medidas confiables de manera rápida y económica para poder evaluar de forma oportuna el impacto ambiental producto del uso de herbicidas en SDCV consiste en la elaboración de modelos mediante Relación (Cuantitativa) Estructura-Actividad ((Q)SAR).

Los modelos (Q)SAR son relaciones matemáticas entre valores de actividad física, química, biológica o ambiental, denominados *endpoints*, y propiedades o descriptores moleculares, generalmente aplicados bajo mecanismos de acción similares [13].

En términos generales, la generación de modelos (Q)SAR para predecir el coeficiente de adsorción de herbicidas en suelos requiere (i) datos empíricos de *endpoint*, (ii) descriptores moleculares y/o propiedades fisicoquímicas de herbicidas y suelos, y (iii) técnicas estadísticas para establecer y validar las relaciones entre *endpoint* y descriptores.

Los objetivos principales de un modelo (Q)SAR son (i) predecir y clasificar la actividad, (ii) interpretar mecanismos de acción, y (iii) optimizar tiempo y costo vinculados a la obtención de dichos datos [14-18]. Además, su aplicación con propósitos regulatorios permite evaluar la reducción de impactos antropogénicos y apoyar la síntesis y aplicación de herbicidas de menor impacto ambiental [19].

Para aplicar estos modelos, la Organización para la Cooperación y el Desarrollo Económico (OCDE) propone cinco principios [13, 20]:

- (i) Un *endpoint* definido con base en los objetivos del modelo y la matriz a la cual es aplicado.
- (ii) Un algoritmo no ambiguo para la generación de descriptores y del modelo.
- (iii) Un dominio de aplicabilidad definido, para conocer los límites y usos del modelo.
- (iv) Una medición apropiada de la bondad del ajuste, robustez y predictividad del modelo, garantizando con ello su calidad estadística.

- (v) Si es posible, dar una interpretación mecanística que avale su sentido físico y químico, que fomente la generación de conocimiento.

Los principios OCDE han sido adoptados por la Unión Europea y Estados Unidos en la generación de modelos (Q)SAR para predecir la adsorción de plaguicidas en SCP [21]. Sin embargo, aun no se han generado modelos (Q)SAR en SDCV y la aplicación de los modelos asociados a SCP para la predicción de un *endpoint* en SDCV conllevaría una alta incerteza debido a las diferentes interacciones herbicida-suelo en ambos tipos de suelos.

La generación de modelos (Q)SAR confiables para la adsorción de herbicidas en SDCV, permite su incorporación en el marco regulatorio, representando una herramienta útil para optimizar la toma de decisiones frente a posibles impactos ambientales, facilitando el cumplimiento de los principios precautorio y preventivo sin limitar el desarrollo económico del país (Figura 1).



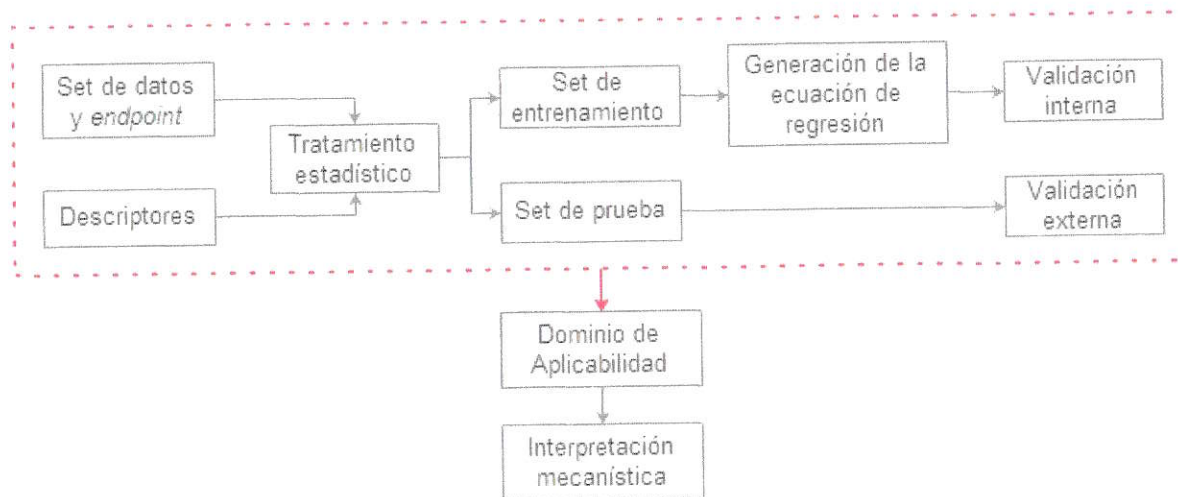
**Figura 1.** Esquema de generación de modelos (Q)SAR para la adsorción de herbicidas en SDCV y su contexto socioambiental.

## **2. ESTADO DEL ARTE**

### **2.1. Predicción del coeficiente de adsorción de herbicidas mediante modelos (Q)SAR**

En la figura 2 se presenta una esquema general para la realización de modelos (Q)SAR en suelos, a partir de una revisión bibliográfica de modelos (Q)SAR para la predicción de la adsorción de contaminantes orgánicos en suelos, restringida a publicaciones en las áreas “Soil chemistry”, “Environmental chemistry” y “Computational chemistry”, utilizando las palabras claves “Soil sorption”, “QSAR”, “QSPR” y “Ash derived volcanic soils” en el servidor “Web of Science” [22], abarcando publicaciones desde 1967 hasta la actualidad.

A continuación, se presenta el análisis de cada paso mostrado en la figura 2 para la creación de modelos (Q)SAR en suelos, enfocado en las deficiencias de los modelos ya existentes, con el fin de proponer una metodología para la generación de modelos (Q)SAR para predecir adsorción de herbicidas en SDCV con propósitos regulatorios.



**Figura 2.** Esquema general de la creación de modelos (Q)SAR para la predicción de la adsorción de contaminantes orgánicos en suelos.



### 2.1.1. Generación de un set de datos de calidad

En general, los valores de coeficientes de adsorción de contaminantes en suelos, sedimentos y/o lodos seleccionados para la creación de modelos (Q)SAR provienen de bases de datos, *handbooks* u otros trabajos, sin un tratamiento previo asociado (Tabla A1), lo que se traduce en las siguientes deficiencias:

- (i) Uso de valores promedio para la adsorción de un contaminante en diferentes distintos adsorbentes (suelos, sedimentos y/o lodos)
- (ii) Uso de diversas metodologías para el cálculo del *endpoint* (estudios *batch*, *centrifugación*, etc.)
- (iii) Uso de diferentes condiciones experimentales (pH, CIC, relación suelo/solución, etc.)
- (iv) Incerteza respecto a la existencia de equilibrio químico o la linealidad de la isoterma.

La falta de un tratamiento previo de los datos se basa en el supuesto de que ya son confiables, pues han sido usados previamente por otros autores. Sin embargo, numerosos modelos (Q)SAR han sido generados a partir de las mismas bases de datos, cumpliendo los criterios estadísticos necesarios para su validación a pesar del variado origen de estos datos (Tabla A1) y la diversidad de ecuaciones de regresión asociadas (Tabla A2), lo que vuelve compleja la interpretación de dichos modelos.

A raíz de lo anterior, se propone considerar los siguientes tres aspectos como necesarios para obtener valores de *endpoint* de calidad:

- (i) Determinar los valores de *endpoint* bajo condiciones experimentales uniformes, describiendo la metodología para asegurar su reproducibilidad.
- (ii) Confirmar la existencia de equilibrio químico mediante estudios cinéticos y definir el intervalo de concentraciones bajo el cual es aplicable el *endpoint*, con base en las isothermas de adsorción.
- (iii) Analizar posibles variables confundentes, como variación del pH y presencia de degradación durante los estudios de cinética y adsorción, suelo contaminado previo al estudio experimental; que afecten a la determinación de los valores de *endpoint* o la interpretación del modelo.
- (iv) No promediar valores de *endpoint* de un mismo contaminante entre diferentes suelos, sedimentos ni lodos.

Lo anterior asegura que los coeficientes de adsorción de contaminantes en suelos sean válidos (condición de equilibrio químico), comparables entre sí (mismas condiciones experimentales) y aplicables con propósitos regulatorios (sensibles al tipo de suelo).

### **2.1.2. Selección de un *endpoint* representativo**

Existen tres tipos de *endpoint* a usar a partir del set de datos validado:

- (i) Coeficiente de partición suelo/solución ( $K_d$ ), empleado para cuantificar el proceso de adsorción lineal de diversos contaminantes orgánicos e inorgánicos en suelos [23, 24].
- (ii) Coeficiente de adsorción de *Freundlich* ( $K_f$ ), usado cuando la adsorción sigue el modelo de *Freundlich* [11, 25-28], y considera un parámetro

adimensional ( $n$ ) que corrige la baja representatividad de  $K_d$  en isothermas no lineales [29, 30].

- (iii) Coeficiente de partición normalizado al contenido de CO ( $K_{oc}$ ), que permite reducir la variabilidad inter-suelo cuando el proceso de adsorción en suelos ocurre principalmente en la MO, lo que se suele cumplir para compuestos apolares de alta lipofilia [31].

El *endpoint* seleccionado debe representar a todos los contaminantes y suelos contemplados en el set de datos. De este modo, adicionalmente a la obtención de datos de calidad, se requiere un análisis de las propiedades fisicoquímicas de cada suelo y contaminante para determinar si el set de datos debe ser usado como  $K_d$ ,  $K_f$  o  $K_{oc}$  (Anexo B). Sin embargo, ningún artículo analiza la representatividad del *endpoint* seleccionado, lo que tiene las siguientes consecuencias:

- (i) Uso de  $K_{oc}$  en suelos con bajo contenido de CO y/o disociación del herbicida [32-42].
- (ii) Uso de  $K_d$  sin corroborar linealidad de la isoterma de adsorción [43].
- (iii) Uso de  $K_f$  sin considerar el impacto del parámetro  $n$  (Anexo B).

Lo anterior incumple el primer principio OCDE e impide la aplicación de dichos modelos en condiciones reales. Esto último es potenciado por el amplio uso de  $K_{oc}$  (Tabla A2), utilizado como paso previo para promediar coeficientes de adsorción entre suelos, impidiendo la generación de modelos (Q)SAR sensibles a las condiciones de suelo locales.

Por lo tanto, para establecer un *endpoint* representativo de la adsorción de herbicidas en SDCV, se propone contextualizar el proceso de adsorción en función de las

propiedades del sistema herbicida-SDCV, y las condiciones ambientales bajo las cuales se aplicará el modelo con propósitos regulatorios.

### **2.1.3. Descriptores empleados en modelos (Q)SAR en suelos**

Los primeros modelos (Q)SAR enfocados a adsorción en suelos consideraron contaminantes de alta lipofilia, generalmente compuestos aromáticos clorados, en suelos y sedimentos cuya adsorción ocurriese principalmente en la fracción orgánica, usando como descriptores la solubilidad en agua ( $S_w$ ) y el coeficiente de partición octanol/agua ( $K_{ow}$ ) [31, 44-49].

Posteriormente, comenzó el uso de descriptores moleculares teóricos, calculados mediante software o servidor web [15, 32, 50]. Estos descriptores son de rápida obtención y reducen el error empírico de  $S_w$  y  $K_{ow}$ , pero no contemplan propiedades del suelo y generan errores en la predicción de plaguicidas polares e ionizables [2, 16, 34, 40, 42, 51].

Entre las propiedades de SDCV que afectan al proceso de adsorción de compuestos ionizables se encuentra el pH, relacionado al grado de ionización de herbicidas y la carga superficial total de dichos suelos [52]. Además, CIC y contenido de CO, arcilla y óxidos de Fe y/o Al podrían predecir en gran medida la adsorción en SDCV [53].

Por lo tanto, un modelo (Q)SAR para predecir coeficiente de adsorción de herbicidas en SDCV debería considerar óxidos de hierro, evaluar la relevancia del contenido de CO en el proceso de adsorción y considerar el pH.

La mayoría de modelos (Q)SAR existentes a la fecha se basa exclusivamente en descriptores moleculares de los contaminantes para explicar la variabilidad del

*endpoint* (Tabla A2), lo que puede ser asociado al promediado de datos entre diferentes suelos, sedimentos y/o lodos, sin considerar el efecto del suelo en el proceso de adsorción, dejando la variabilidad únicamente a efectos moleculares. Sin embargo, esto implica un sesgo interpretativo y falencias al momento de aplicar estos modelos con propósitos regulatorios, debido a la imposibilidad de (i) adecuar la predicción a las condiciones locales de cada suelo y (ii) contemplar las posibles interacciones entre herbicidas y suelos asociadas a mecanismos de adsorción.

Además, se recurre al uso de descriptores moleculares teóricos que representan propiedades empíricas, tales como  $^1\chi$ ,  $\log D$  y  $\log P$  o  $\log P_n$  (Tabla A2), calculados mediante software para representar a  $K_{ow}$ . Estos descriptores no tienen error experimental asociado, aumentando la precisión de los modelos (Q)SAR, pero no son contrastados con los valores empíricos que representan, ni con la precisión de los valores observados de *endpoint*, generando incertezas respecto a la confianza de las predicciones, limitando así su aplicabilidad con propósitos regulatorios.

En contraposición a la idea de generar predicciones sin error, la incerteza en ciencias ambientales suele ser elevada debido a la complejidad de las matrices ambientales, como el suelo. En consecuencia, aumentar la precisión de modelos predictivos sin considerar la variabilidad intrínseca de estas matrices podría disminuir la confianza de los valores predichos. Esto, a su vez, puede repercutir gravemente a nivel socioambiental (p. ej. subestimación del impacto ambiental en algún suelo particular) o socioeconómico (p. ej. sobreestimar el impacto ambiental y limitar el desarrollo económico).

Lo anterior se haría evidente al contrastar los valores predichos con mediciones en condiciones reales. Sin embargo, ningún modelo (Q)SAR analizado ha sido probado en tales condiciones, a pesar de ser generados con propósitos regulatorios.

En este sentido, se propone la incorporación de descriptores que incluyan de manera intrínseca la variabilidad del sistema contaminante-suelo, así como la dependencia espacial, temporal e inter-suelo, en vez de usar descriptores teóricos con bajo o nulo error pero poco representativos.

Lo anterior supone el uso de descriptores de naturaleza empírica, permitiendo que la aplicación de modelos con fines regulatorios quede en manos del sector productivo y público de nuestro país, quienes continuamente deben cuantificar las propiedades fisicoquímicas del suelo para la realización de estudios de línea base, es decir, describir y caracterizar las áreas potencialmente afectadas por actividad natural y/o antropogénica.

#### **2.1.4. Tratamiento estadístico de los datos y descriptores**

El tratamiento estadístico posee la finalidad de minimizar la incerteza del modelo y mejorar su aplicabilidad [54].

Además de la calidad de los valores de *endpoint*, se evalúa su distribución en todo el intervalo de valores [20] para evitar la presencia de puntos de influencia en la generación de la ecuación de regresión del modelo (Q)SAR para la adsorción de herbicidas en SDCV.

Respecto a los descriptores propuestos y definidos para la creación del modelo (Q)SAR para la adsorción de herbicidas en SDCV, el tratamiento estadístico consiste

en reducir su cantidad, prefiriendo descriptores con baja correlación entre sí y alta correlación con el *endpoint*, lo que puede ser analizado a través de la matriz de correlación.

Como apoyo a la reducción de descriptores, en este trabajo se propone el Análisis de Componentes Principales (ACP) para evaluar la presencia de información redundante y la posible unificación de los descriptores propuestos en nuevos descriptores que engloben mayor información y representen mayor variabilidad del *endpoint*.

#### **2.1.5. Sets de entrenamiento y de prueba**

Para validar un modelo (Q)SAR, sus datos deben primero ser distribuidos entre un set de entrenamiento, cuyos datos son usados para generar el modelo y probar su robustez y bondad del ajuste, y un set de prueba, cuyos datos permiten evaluar la predictividad del modelo. Sin embargo, solo algunos modelos (Q)SAR para predecir la adsorción de contaminantes en suelos consideran set de prueba, mientras que el número de datos destinado a cada set es variable (Tabla A2).

Con la idea de estandarizar la creación de los sets de entrenamiento y de prueba, Tropsha y col. (2010) exigen un mínimo de 30 datos para generar y validar un modelo (Q)SAR, los que son divididos entre [20]:

- (i) Set de entrenamiento: Contiene al menos 20 datos e idealmente representa el 80 - 85% del total de los datos disponibles.
- (ii) Set de prueba: Contiene al menos 10 datos e idealmente representa el 15 - 20% del total de los datos disponibles.

La división de datos puede ser aleatoria o racional. Comparativamente, los métodos de selección racional usan algoritmos para separar datos y requieren un menor número de datos en el set de entrenamiento para explicar la misma variabilidad que la división aleatoria [55], lo que se asocia a una mayor representatividad [37]. Esto implica un mejoramiento de la calidad interna del modelo, pero no necesariamente de su predictividad [56].

#### **2.1.6. Ecuación de regresión**

El modelo (Q)SAR se genera a partir del set de entrenamiento usando métodos estadísticos, entre los cuales el uso de regresión no lineal es propuesto para sistemas ambientales de matriz compleja, como el suelo [57].

Pese a que algunos modelos (Q)SAR no lineales existentes en suelos incluyen propiedades del suelo y grado de ionización de herbicidas [39, 58], su falta de linealidad dificulta la interpretación mecanística del proceso de adsorción, lo que limita el cumplimiento de los criterios OCDE. No obstante, los modelos (Q)SAR lineales facilitan dicha interpretación a partir de sus coeficientes, siendo la Regresión Lineal Múltiple (RLM) el método más usado para la creación de dichos modelos.

#### **2.1.7. Validaciones interna y externa**

La validación interna permite evaluar la calidad estadística del set de entrenamiento (bondad del ajuste) y la independencia de la ecuación generada respecto al *endpoint* y los descriptores (robustez). Esto último puede ser evaluado mediante [59]:

- (i) Aleatorización en Y: Corresponde al ajuste comparativo del modelo original con nuevos modelos generados a partir de valores de *endpoint* redistribuidos



aleatoriamente, manteniendo la matriz de descriptores invariable. De esta manera, se puede estimar si el modelo original representa una posible relación mecanística o una correlación azarosa.

- (ii) Validación cruzada (*Leave-One-Out*): Consiste en eliminar un valor de *endpoint* y generar nuevamente el modelo con la misma matriz de descriptores. Posteriormente, con el modelo generado se predice el valor eliminado y se repite el proceso hasta abarcar a todos los valores en el set de entrenamiento. Si la correlación obtenida para todos los casos no varía respecto de la original, se descarta la presencia de puntos de influencia.

La validación externa evalúa la correlación entre valores de *endpoint* predichos y observados en el set de prueba, para determinar la predictividad del modelo (Q)SAR.

En función del éxito en ambas validaciones se rechazará o mantendrá el modelo propuesto [37].

La mayor deficiencia de la validación estadística de modelos (Q)SAR para la predicción de coeficientes de adsorción de contaminantes en suelos radica en el no uso de todos los criterios estadísticos existentes a la fecha, tales como: (i) realizar un set de prueba, (ii) relacionar de manera confiable número de datos y número de descriptores ( $p \geq 5$ ), o (iii) informar con claridad los parámetros estadísticos considerados. Estos aspectos son observados en la Tabla A2, y se tendrán en cuenta en la realización de modelos (Q)SAR para la predicción de la adsorción de herbicidas en SDCV.

### 2.1.8. Dominio de aplicabilidad

Se entiende por dominio de aplicabilidad al espacio construido por los descriptores del modelo y los valores de *endpoint* predichos en el set de prueba. Este se emplea para estimar la incerteza en la predictividad del modelo basado en la similitud estructural de los compuestos usados en ambos sets [36, 42, 60, 61]. Con ello, se busca explicar: (i) tipo de molécula que es predicho, (ii) relación entre moléculas y descriptores usados, y (iii) aplicabilidad del modelo. Sin embargo, dicha aplicabilidad suele ser limitada debido a la no consideración del impacto del suelo en el proceso de adsorción.

El dominio de aplicabilidad de modelos (Q)SAR para predecir la adsorción de contaminantes en suelos se determina de manera estadística, analizando (i) puntos de influencia y valores atípicos que afecten la interpretación del modelo, y (ii) grupos funcionales y tipos de moléculas en el set de datos, clasificando muchas veces la robustez y predictividad del modelo en función de las clases de moléculas para una aplicación selectiva.

Sin embargo, el dominio de aplicabilidad es altamente dependiente de: (i) *endpoint*, (ii) tratamiento y calidad de los datos, (iii) descriptores iniciales y seleccionados, (iv) generación del modelo y (v) validaciones interna y externa del modelo.

Por lo tanto, las deficiencias expuestas anteriormente para las primeras fases de creación del modelo (p.ej. validación del set de datos y selección de un *endpoint* representativo de los sistemas contaminante-suelo considerados) podrían afectar la validez de los datos usados, implicando un dominio de aplicabilidad sin sentido

aplicado real (p.ej. modelos que predicen valores de  $K_{oc}$  promedio y pretenden tener un uso con propósitos regulatorios en suelos de diferente tipo).

Otros aspectos importantes a considerar en el dominio de aplicabilidad del modelo son la metodología experimental empleada para la obtención de los valores de *endpoint* y la aplicabilidad del modelo generado en condiciones reales, para lo cual es necesario incluir la variabilidad del suelo (no promediar coeficientes de adsorción).

Respecto a la metodología, Walker y Jurado-Exposito (1998) compararon  $K_d$  de herbicidas obtenidos por *sistemas batch* y por *técnica de centrifugación*, concluyendo que la metodología *batch* sobreestima el valor de  $K_d$  [62].

En un estudio similar, Kah y Brown (2007) encontraron que el valor de  $K_d$  de herbicidas obtenido mediante *sistemas batch* fue superior al obtenido por *técnica de centrifugación* solo si los compuestos se encontraban altamente adsorbidos, ocurriendo lo contrario para compuestos de baja adsorción [63].

Para asegurar la aplicabilidad de modelos (Q)SAR en condiciones reales, estos deben ser puestos a prueba en condiciones reales en que la metodología empleada para determinar el *endpoint* podría no ser representativa, y el proceso de adsorción podría ser acompañado de procesos tales como degradación, transporte, biodegradación, entre otros. Esto no implicaría necesariamente un modelo sin sentido físico ni químico, sino que acotaría su aplicación con propósitos regulatorios.

Por lo tanto, es necesario comparar la metodología de obtención de los valores de *endpoint* con las condiciones reales para establecer su correspondencia y explicar

sobreestimaciones o subestimaciones, así como caracterizar adecuadamente los escenarios de validez y aplicabilidad de los modelos (Q)SAR generados.

### **2.1.9. Interpretación mecanística**

La interpretación mecanística del proceso de adsorción de herbicidas en suelos, realizada a partir de cada modelo (Q)SAR, permite evaluar el sentido físico y químico del modelo generado y se relaciona directamente con el dominio de aplicabilidad.

Para la interpretación de los descriptores, se analiza la magnitud, signo y error de sus coeficientes asociados. No obstante, los modelos (Q)SAR existentes para la adsorción de contaminantes en suelos no son interpretados con base en el *endpoint* usado ni las condiciones bajo las cuales fueron generados (p.ej. promediando suelos).

Las inconsistencias en el sentido físico y/o químico, así como existencia de varianza no explicada por el modelo podrían deberse a variables confundentes que no hayan sido eliminadas en el tratamiento de los datos [20], lo que tampoco es considerado en los modelos (Q)SAR revisados.

## **2.2. Propuestas para facilitar la validación de modelos (Q)SAR con propósitos regulatorios**

### **2.2.1. Transferibilidad y criterios OCDE como condición necesaria para la validación de modelos (Q)SAR**

La transferibilidad es una propiedad de los modelos (Q)SAR con fines regulatorios que pretende facilitar su uso a diferentes investigadores y organizaciones mediante la

incorporación de una reproducibilidad empírica, estadística y teórica [54, 64], basada en los siguientes criterios:

- (i) Los descriptores usados deben ser reproducibles.
- (ii) La definición del modelo (métodos estadísticos, algoritmos) debe ser clara.
- (iii) La bondad del ajuste y robustez deben ser cuantificadas.
- (iv) La reproducibilidad de las predicciones debe ser cuantificada.
- (v) El modelo debe contar con documentación adecuada para validar su desarrollo y analizar su aplicación.

En general, los modelos (Q)SAR revisados consideran todos estos criterios. No obstante, su aplicación posee deficiencias, tales como:

- (i) La reproducibilidad de los descriptores está condicionada al uso de software de pago, con lo que la reproducibilidad de los modelos (Q)SAR requiere comprar los mismos programas usados por los autores.
- (ii) La cuantificación de bondad del ajuste, robustez y predictividad incluye parámetros estadísticos sensibles a los grados de libertad, de modo que no es posible comparar modelos (Q)SAR con diferente número de datos en los sets de entrenamiento y de prueba o diferente número de descriptores.
- (iii) La documentación empleada corresponde a modelos (Q)SAR previos, pero no se contextualiza cada sistema contaminante-suelo (p.ej. evaluación de la representatividad del *endpoint*), impidiendo con ello analizar la aplicación del modelo.

Por otra parte, la OCDE define la validación como el proceso en el cual son establecidas la confianza y relevancia del estudio y modelo (Q)SAR para cierto propósito particular [13], donde:

- (i) **Confianza:** Reproducibilidad y repetibilidad del modelo, incluyendo los métodos usados para la generación de sus datos.
- (ii) **Relevancia:** Significancia y utilidad del modelo en un sentido predictivo e interpretativo, según los objetivos perseguidos con su creación.

De este modo, la confianza se relaciona con la transferibilidad y requiere informar el origen del set de datos, mientras que la relevancia exige una validación empírica del modelo confeccionado, por lo que se propone considerar la transferibilidad y los criterios OCDE mencionados al momento de crear y validar modelos (Q)SAR.

### **2.2.2. Contextualización y validación empírica mediante la aplicación del ciclo PDCA en modelos (Q)SAR**

Una estrategia útil para disminuir la incerteza de los modelos (Q)SAR generados para predecir adsorción en SDCV consiste en la aplicación del ciclo de mejora continua, o ciclo PDCA (Plan-Do-Check-Act), ampliamente usado en sistemas de gestión de calidad [65].

La aplicación del ciclo PDCA en modelos (Q)SAR contemplaría las siguientes fases:

- (i) **Planificar (Plan):** Recolección de datos e información para la generación de un set de datos de calidad y descriptores en función de los objetivos del modelo (Q)SAR.

- (ii) Realizar (Do): Contextualización del set de datos y descriptores al sistema herbicida-suelo, creando finalmente el modelo.
- (iii) Verificar (Check): Validación interna y externa del modelo, validación empírica (predictividad en condiciones reales) y evaluación de las condiciones ambientales que afectan su Dominio de Aplicabilidad.
- (iv) Actuar (Act): Analizar y corregir posibles problemas en base al funcionamiento del modelo. Inclusión de nuevos datos. Recomenzo del ciclo con un modelo mejorado.

La principal dificultad para la generación de modelos (Q)SAR aplicando el ciclo PDCA es la necesidad de trabajo multidisciplinario en equipo con una componente científica y otra pública/privada, con la participación del país mediante Instrumentos de Gestión Ambiental.

Un ejemplo de lo anterior sería la aplicación de estudios de línea base en Estudios de Impacto Ambiental que apoyen la fase de planificación de los modelos (Q)SAR, facilitando su creación e incorporando en la misma un objetivo ambiental de aplicación local. De este modo, el ciclo PDCA disminuiría el reduccionismo intrínseco de los modelos (Q)SAR y favorecería la cooperación, comprensión y aprendizaje continuo [65].

Finalmente, los modelos (Q)SAR generados bajo estas consideraciones tienen las siguientes ventajas:

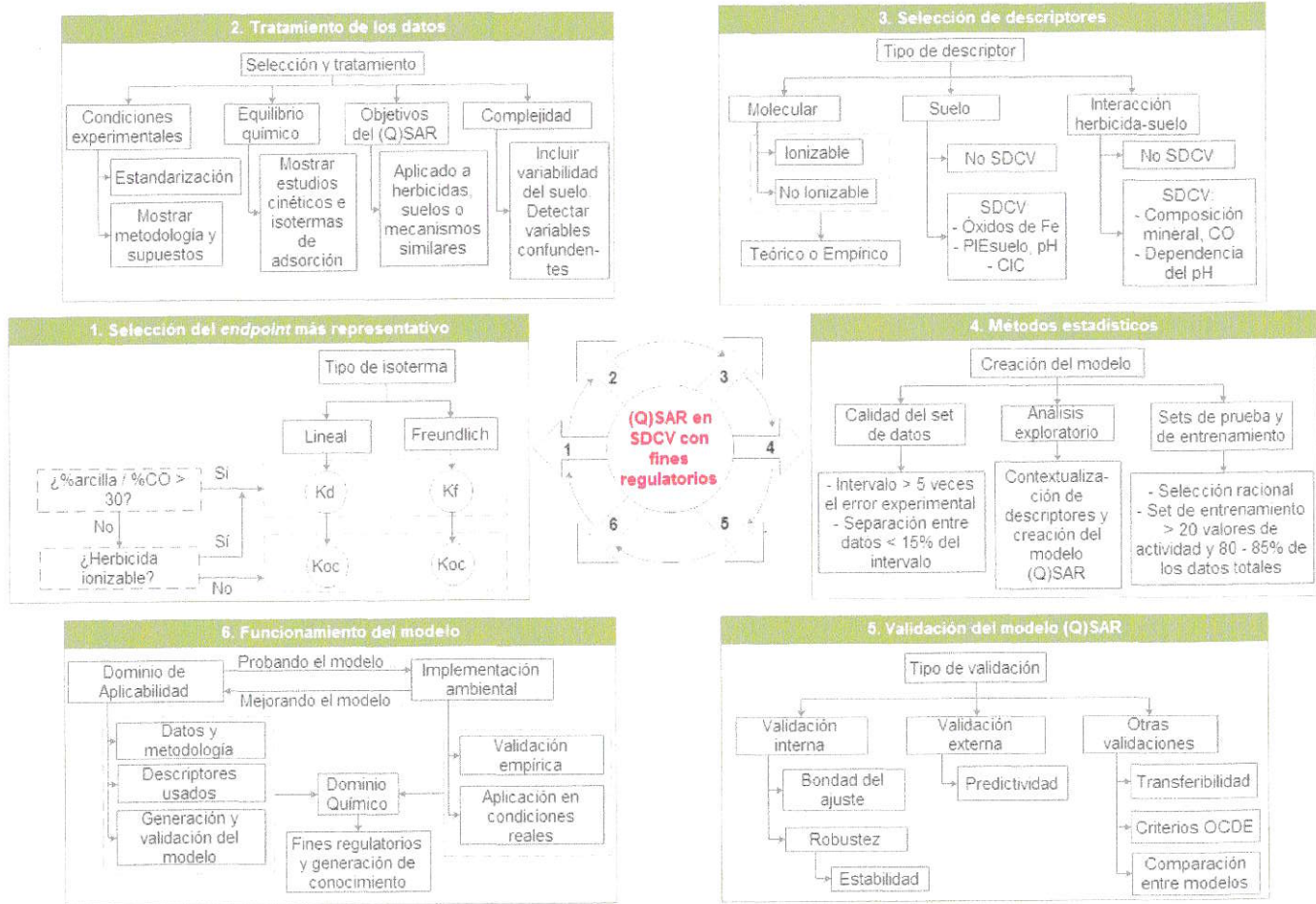
- (i) Se puede obtener información tanto de las predicciones acertadas como de las fallidas, redefiniendo la aplicabilidad e interpretación del modelo, así como sus limitaciones para una mayor certeza ambiental.

- (ii) Permite apoyar a los principios precautorio y preventivo al facilitar la detección de falencias, limitaciones y aplicaciones de los modelos (Q)SAR generados con propósitos regulatorios.
- (iii) Se disminuye el reduccionismo intrínseco de los modelos (Q)SAR y se favorece su contextualización con cada reinicio del ciclo PDCA, obligando que cada nuevo modelo presente una base teórico/conceptual y sea transferible para continuar el ciclo.

### **2.2.3. Esquema de creación de modelos (Q)SAR con propósitos regulatorios**

Con la finalidad de englobar todos los aspectos previamente mencionados y aplicarlos a la generación de modelos (Q)SAR para predecir la adsorción de contaminantes en suelos, incluyendo SDCV, se muestra en la Figura 3 el esquema para la realización de modelos (Q)SAR enfocado en SDCV, y en la Figura 4 el esquema para dar una correcta aplicación ambiental a dichos modelos.





**Figura 3.** Esquema para realizar modelos (Q)SAR para predecir adsorción de herbicidas en SDCV.

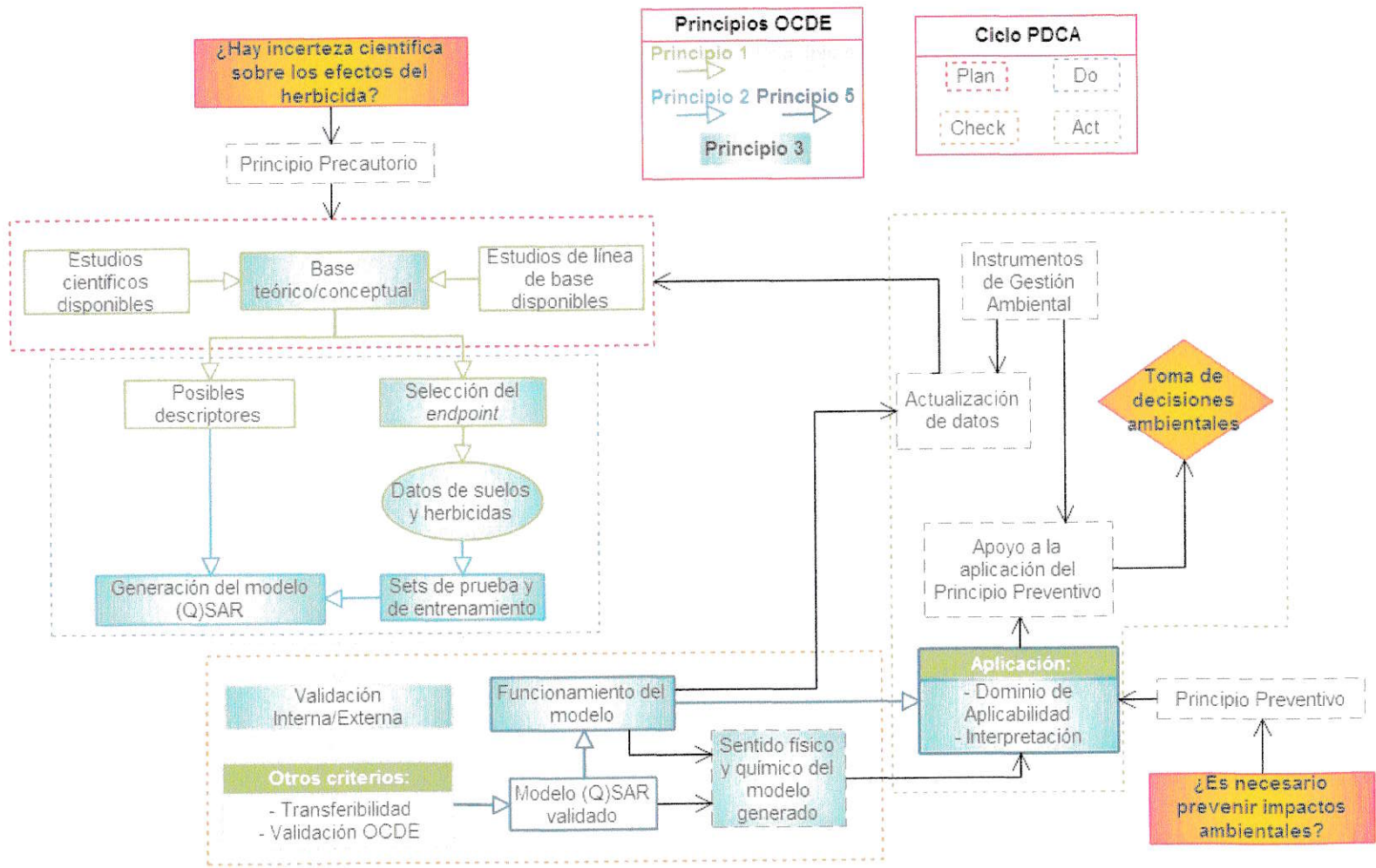


Figura 4. Ciclo de mejora continua propuesto para modelos (Q)SAR para predecir adsorción de herbicidas en SDCV.

### **3. HIPÓTESIS Y OBJETIVOS**

#### **3.1. Propuesta de investigación**

Con base en los antecedentes teóricos previamente expuestos y teniendo presente las deficiencias de los modelos (Q)SAR realizados a la fecha y la inexistencia de modelos aplicados a SDCV, el propósito de este trabajo es realizar un modelo (Q)SAR para predecir valores de *endpoint* de herbicidas ionizables y no ionizables en SDCV chilenos mediante la contextualización de los sistemas herbicida-SDCV, con la finalidad de incluir una interpretación mecanística en su formulación.

La contextualización del modelo y la metodología para su generación, serán realizadas a partir del Estado del Arte y el análisis del ajuste de los datos aplicados a modelos (Q)SAR definidos para SCP, presentando una serie de aspectos a evaluar para superar las deficiencias de los modelos (Q)SAR y asegurar su calidad con propósitos regulatorios.

#### **3.2. Hipótesis**

La adsorción de herbicidas ionizables y no ionizables en suelos derivados de cenizas volcánicas es predicha e interpretada por medio de modelos (Q)SAR lineales generados a partir de una contextualización teórico/conceptual, considerando

descriptores moleculares de herbicidas y propiedades fisicoquímicas de estos suelos, e incluyendo variabilidad inter-suelo y un análisis de alcances y limitaciones de cada modelo orientado a su aplicabilidad con propósitos regulatorios a corto y largo plazo.

### **3.3. Objetivo general**

Predecir e interpretar la dinámica de 12 herbicidas ionizables y no ionizables en 10 suelos derivados de cenizas volcánicas utilizando modelos (Q)SAR con propósitos regulatorios para el coeficiente de adsorción más representativo, englobando descriptores moleculares, del suelo y de la interacción herbicida-suelo.

### **3.4. Objetivos específicos**

- (i) Determinar mediante revisión bibliográfica las principales consideraciones para la realización de modelos (Q)SAR en SDCV.
- (ii) Generar un set de datos de calidad considerando estudios cinéticos y de adsorción, y criterios estadísticos para hallar al *endpoint* más representativo.
- (iii) Evaluar la capacidad predictiva de modelos (Q)SAR para SCP usando valores de *endpoint* de herbicidas en SDCV y realizar un análisis bibliográfico en función de los herbicidas y propiedades de los SDCV usados, con el fin de evaluar el impacto de los SDCV y contextualizar los sistemas herbicida-SDCV.
- (iv) Establecer posibles descriptores asociados a la interacción herbicida-SDCV, con la idea de mejorar la correspondencia entre descriptores y *endpoint*.
- (v) Generar modelos (Q)SAR lineales para predecir el *endpoint* de herbicidas en SDCV, utilizando descriptores moleculares del herbicida, del suelo y de la interacción herbicida-SDCV.

- (vi) Asegurar la transferibilidad y el cumplimiento de criterios OCDE de los modelos (Q)SAR generados, para facilitar su aplicabilidad con fines regulatorios y complementar a herramientas de gestión ambiental.
- (vii) Interpretar los modelos (Q)SAR generados a través del análisis de sus descriptores y coeficientes y su relación con los sistemas herbicida-SDCV.

## 4. MATERIALES Y MÉTODOS

### 4.1. Herbicidas y suelos

Se utilizaron 12 herbicidas Sigma-Aldrich, grado analítico (98% pureza), a saber: Ácido 2,4-diclorofenoxiacético (2,4-D), Atrazina (AT), Clorsulfurón (CS), Diurón (DI), Glifosato (GFS), Metabentiazurón (MBT), Metolacloro (MC), Metsulfurón-metil (MSM), Nicosulfurón (NS), Picloram (PC), Simazina (SM) y Trifluralín (TF).

Además, se usaron 10 SDCV representativos de la zona centro-sur del país, recolectados en sitios vírgenes (sin fertilizar) hasta una profundidad de 15 cm: Collipulli (COLL), Diguillín (DIG), Freire (FRE), Frutillar (FRU), Metrengo (MET), Nueva Braunau (NBR), Osorno (OSN), Ralún (RAL), Santa Bárbara (STB) y Temuco (TCO).

Cada suelo fue tamizado a un tamaño de partícula menor a 2 mm, secado a 40 °C en estufa durante 24 hrs (secado al aire) y almacenado a temperatura ambiente en envases de plástico hasta su utilización.

En el Anexo C se presentan las características de estos herbicidas (Tabla C1) y suelos (Tabla C2). La composición mineralógica de estos últimos se describe en la Tabla C3.

## 4.2. Generación de modelos (Q)SAR en SDCV

### 4.2.1. Selección de un *endpoint* representativo

Se realizó la selección del *endpoint* más representativo de manera empírica, a partir de datos generados por el Laboratorio de Físicoquímica & Analítica (PACHEM, Facultad de Ciencias Básicas, Universidad Metropolitana de Ciencias de la Educación) (Anexo D), debido a la complejidad de los mecanismos de adsorción involucrados en SDCV. Lo anterior se consiguió mediante el siguiente procedimiento:

- (i) Si la adsorción es lineal en todos los casos, utilizar  $\log K_d$ .
- (ii) Si la adsorción no es lineal en al menos un valor, utilizar  $\log K_f$  junto con el parámetro  $n$  del modelo de *Freundlich*.
- (iii) Si la relación %arcilla/%CO es menor a 30 en todos los suelos [66] y además los herbicidas ionizables se encuentran en su forma neutra, normalizar a  $\log K_{oc}$  o  $\log K_{foc}$ , según corresponda.

Donde se consideró el coeficiente de adsorción en su forma logarítmica porque así es utilizado en literatura. De este modo, es posible reducir la amplitud del error y facilitar la comparación con otros modelos (Q)SAR.

### 4.2.2. Predicción de valores de *endpoint* empleando modelos (Q)SAR para SCP

Se aplicaron los modelos (Q)SAR para predecir la adsorción de contaminantes orgánicos en SCP creados por Bintein y Devillers (1994), Sabljic y col. (1995) y Wen y col. (2012) [30, 39, 60], con la finalidad de predecir los valores de *endpoint* usados en este estudio.

Se analizó el ajuste obtenido en función del tipo de herbicida (ácido, base, no ionizable) y las propiedades de SDCV que podrían influir en el mismo, para determinar el impacto de la carga variable y la aplicabilidad de dichos modelos en el contexto nacional.

El análisis se realizó clasificando los herbicidas en no ionizables, ácidos, básicos y aquellos que no cumplieron los requisitos estadísticos para validar el *endpoint*.

#### **4.2.3. Cálculo de descriptores para generar modelos (Q)SAR en SDCV**

La obtención, generación y propuesta de descriptores moleculares y descriptores asociados a SDCV es presentada en el Anexo E.

La mayoría de descriptores moleculares calculados son teóricos, mientras que los descriptores propuestos para SDCV provienen de sus propiedades fisicoquímicas.

#### **4.2.4. Análisis exploratorio de la matriz de descriptores**

Se utilizó el software *Data Pre Treatment GUI* [67] para eliminar descriptores constantes y altamente correlacionados (Anexo E).

Este proceso fue realizado para todos los descriptores de SDCV, para todos los descriptores moleculares, y para todos los descriptores en general, con la finalidad de determinar alguna correlación.

Posteriormente, se aplicó un ACP usando el programa *R* 3.2.3 y el paquete *FactoMineR*, para analizar relaciones entre descriptores.



#### **4.2.5. Descriptores de la interacción herbicida-SDCV**

Se generaron descriptores de la interacción herbicida-SDCV considerando antecedentes teórico/conceptuales y el ACP realizado previamente (sección 4.2.4) para unificar propiedades de herbicidas y SDCV que estuvieran relacionadas entre sí por algún mecanismo de adsorción.

Se usaron descriptores basados en propiedades fisicoquímicas de herbicidas y SDCV con una fácil determinación y comprensión, que deban ser calculados constantemente para asegurar la representatividad de los valores de *endpoint*.

En este caso, la interpretación mecanística de los descriptores fue incluida en su deducción, con lo cual el uso de estos no requirió curación de la matriz de descriptores ni selección de una combinación óptima de descriptores.

#### **4.2.6. Generación de bases de datos para el análisis de los valores de *endpoint* de herbicidas en SDCV**

Se generaron 22 bases de datos con el objetivo de analizar la variabilidad del *endpoint* en función de herbicidas y SDCV:

- (i) Todos los SDCV para un solo herbicida (10 bases de datos)
- (ii) Todos los herbicidas en un solo SDCV (12 bases de datos)

Se crearon dos bases de datos para la variabilidad del *endpoint* en función de la ionización, basado en las condiciones de trabajo experimentales:

- (i) Todos los suelos con todos los herbicidas ionizados (1 base de datos)
- (ii) Todos los suelos con todos los herbicidas no ionizados (1 base de datos)

#### 4.2.7. Calidad estadística de las bases de datos

Se usaron los siguientes criterios para validar las bases de datos a usar en la generación de modelos (Q)SAR para predecir adsorción de herbicidas en SDCV [20]:

- (i) Al menos 30 valores de *endpoint* en la base de datos.
- (ii) Intervalo de valores de *endpoint*  $\geq 5$  veces el error experimental promedio.
- (iii) Diferencia entre dos valores consecutivos de *endpoint*  $\leq 10 - 15\%$  del intervalo completo de valores.

Si al menos uno de los parámetros del modelo de *Freundlich* ( $K_f$  o  $n$ ) no cumple con los criterios antes mencionados, se eliminan los dos parámetros asociados a esa matriz ambiental, para tener control sobre el comportamiento completo de la isoterma de adsorción y cumplir así con un criterio predictivo adecuado para evaluar el impacto ambiental del herbicida. De este modo, las bases de datos asociadas a ambos parámetros fueron generadas y modificadas conjuntamente.

Una vez validados los valores de *endpoint* a incluir en las bases de datos, se utilizó el software *Data Pre Treatment GUI* [67] para eliminar descriptores correlacionados (Anexo E).

#### 4.2.8. División en sets de entrenamiento y de prueba

Se utilizó el método de selección racional “*Euclidean Distance Based*” del software *Dataset Division GUI 1.2* [67] para dividir entre set de prueba y de entrenamiento las bases de datos para la generación modelos (Q)SAR (Anexo E).

#### **4.2.9. Selección de las mejores combinaciones de descriptores**

Se determinó el número máximo de descriptores capaces de ajustar los datos de manera estadísticamente significativa, con base en:

- (i) Calidad de la regresión: Al menos 5 valores de *endpoint* en el set de entrenamiento por cada descriptor a considerar en el modelo.
- (ii) Capacidad de cálculo del ordenador: Máximo 5 descriptores para el modelo.

Se modelaron todas las combinaciones posibles mediante el software *Best Subset Selection Modified v2.1* [67] (Anexo E).

En caso de seleccionarse descriptores moleculares teóricos, tales como KOWWIN y WSKOW, se validaron según su correspondencia con valores empíricos.

#### **4.2.10. Generación de la ecuación de regresión lineal**

Los modelos lineales fueron creados usando el software *MLR Plus Validation 1.3* [67], considerando:

- (i) Las mejores combinaciones de descriptores encontradas previamente.
- (ii) Los descriptores provenientes de la interacción herbicida-SDCV.

### **4.3. Validación de modelos (Q)SAR**

Los parámetros estadísticos fueron cuantificados con *Microsoft Office Excel 2007* y el software *MLR Plus Validation 1.3* [67]. La calidad de los modelos generados se evaluó mediante el cumplimiento de criterios de validación interna y externa, presentados en el Anexo E, Tablas E1 (bondad del ajuste), E2 (robustez), E3 (predictividad) y E4 (comparación entre modelos).

#### **4.4. Transferibilidad y criterios OCDE**

Para asegurar la transferibilidad del modelo (Q)SAR, se evaluaron los criterios existentes incorporando varias propuestas, a saber:

A la reproducibilidad de los descriptores usados, se agregó como condición su fácil determinación e interpretación, además de su contextualización respecto del sistema herbicida-SDCV a predecir.

A la clara definición del modelo (métodos estadísticos, algoritmos), se incorporó la contextualización de los sistemas herbicida-SDCV como una condición necesaria, apoyada en revisión bibliográfica para el correcto desarrollo y aplicación del modelo.

A la cuantificación de la bondad del ajuste y robustez, se agregó el aseguramiento de la reproducibilidad y fácil determinación de parámetros estadísticos, dando a conocer toda la información relacionada en caso de que otros investigadores quieran comparar modelos con diferentes grados de libertad, número de datos, entre otros.

A la cuantificación de la reproducibilidad de las predicciones, se incorporó la evaluación de la predicción de valores de *endpoint* previamente descartados, con la finalidad de dar una interpretación a los valores predichos de manera confiable y también a los valores problemáticos, para tener más pistas sobre las limitaciones del modelo y su sentido físico y químico.

A la documentación adecuada para validar el desarrollo y analizar la aplicación del modelo, se agregó como requisito asegurar la accesibilidad mediante software de uso libre, generando así un modelo cuya reproducibilidad no esté limitada a la componente económica.

Con lo anterior, se intentó verificar la calidad tanto en el desarrollo del modelo como en la generación y tratamiento de datos empíricos, aplicabilidad y predictividad, facilitando su comparación con otros modelos.

Respecto los criterios OCDE, la confianza se aseguró mediante la descripción de la metodología usada para la obtención de datos de calidad y la generación del modelo, el descarte de variables confundentes y la transferibilidad, mientras que la relevancia se hizo dependiente de la generación de líneas de base por parte de las autoridades con competencia ambiental, para la actualización y validación continua del modelo en condiciones reales.

#### **4.5. Dominio de Aplicabilidad e interpretación mecanística**

##### **4.5.1. Detección de valores atípicos y puntos de influencia**

Se realizó la detección de valores atípicos mediante el software *MLR Plus Validation 1.3* [67], usando el método “*Standardization Approach*”.

Fueron considerados puntos de influencia aquellos datos cuya distancia respecto a otros datos fuese  $\geq 15\%$  del intervalo completo de valores predichos.

##### **4.5.2. Interpretación y aplicabilidad del modelo en condiciones ambientales**

Se evaluaron los escenarios de validez del modelo según tipo de herbicidas predichos y sus errores asociados, considerando sus propiedades y la variabilidad entre moléculas y entre SDCV, así como posibles limitaciones en los descriptores.

Se evaluaron las limitaciones del modelo respecto a descriptores, validez temporal y propósitos regulatorios, considerando para ello la presencia de valores atípicos en el análisis, con la idea de obtener pistas mecánísticas asociadas a dichas limitaciones.

Se consideraron además las limitaciones metodológicas, basadas en la correspondencia entre condiciones ambientales y supuestos y condiciones de trabajo.

Se generaron criterios de aplicabilidad del modelo considerando las condiciones ambientales favorables para la aplicabilidad del modelo con propósitos regulatorios y el potencial apoyo a los principios preventivo y precautorio.

## 5. RESULTADOS Y DISCUSIÓN

### 5.1. Selección de un *endpoint* representativo

La mayoría de las isothermas de adsorción de los 12 herbicidas en 10 SDCV son no lineales, por lo cual  $K_d$  es descartado como *endpoint*.

El suelo Ultisol COLL es el único cuya relación %arcilla/%CO es mayor a 30, indicando un efecto significativo de la fracción mineral sobre el proceso de adsorción, asociado particularmente a Caolinita [11, 25] (Tabla C3). Sin embargo, debido al pKa de los herbicidas ácidos (Tabla C1) y las condiciones de pH durante la realización de las isothermas (aproximadamente neutro), estos herbicidas se encontrarán ionizados, descartando así la normalización a  $K_{foc}$  (Figura 3).

En vista de lo anterior, se escogen como *endpoint* representativos  $\log K_f$  y  $n$ .

Por último, fueron eliminados los valores asociados a: (i) SM en suelos COLL, MET, FRE, STB y RAL, (ii) GFS en MET, FRE y NBR, y (iii) DI en COLL, debido al incumplimiento de criterios de calidad estadística (Anexo D) para el ajuste del modelo de *Freundlich* en dichos casos. No obstante, estos valores fueron incorporados a las bases de datos para el análisis de tendencias, con el propósito de

dar una explicación mecanística a estas tendencias, pero no fueron usados en la creación de los modelos (Q)SAR.

## 5.2. Ajuste de los valores de *endpoint* de herbicidas en SDCV empleando modelos (Q)SAR para SCP

### 5.2.1. Modelo de Bintein y Devillers

El modelo de Bintein y Devillers (1994) (MBD) modelo fue propuesto para predecir la adsorción de 53 contaminantes orgánicos tanto neutros como ionizables, a partir de isothermas de *Freundlich* con  $n = 1$ , es decir, adsorción lineal ( $K_f = K_d$ ), en diferentes suelos y sedimentos con  $\%CO > 0,1$  [30].

De un total de 229 valores de  $K_d$ , los autores generaron la ecuación:

$$\log K_d = 0,93 * \log(K_{ow}) + 1,09 * \log(f_{CO}) + 0,32 * CFa - 0,55 * CFB' + 0,25 \quad (1)$$

Este modelo presenta  $\log K_d$  como *endpoint*, en vez de  $\log K_f$ . Relacionando ambos valores mediante las ecuaciones B1 y B3 (Anexo B):

$$K_f = K_d * C_{eq}^{1-n} \quad (2)$$

Luego:

$$\log K_f = \log K_d + (1 - n) * \log C_{eq} \quad (3)$$

Por lo tanto, al usar  $\log K_f$  como *endpoint*, el modelo es dependiente de la concentración en solución y del parámetro  $n$ . Sin embargo, suponiendo que todos los herbicidas de interés se encuentran a una concentración  $C_{eq} = 1$ , o bien, que son descritos por una isoterma de adsorción lineal ( $n = 1$ ):



$$\log K_f = \log K_d \quad (4)$$

Reemplazando en el modelo:

$$\log K_f = 0,93 * \log(K_{ow}) + 1,09 * \log(f_{CO}) + 0,32 * CFa - 0,55 * CFb' + 0,25 \quad (5)$$

Para analizar el sentido físico de esta ecuación, se reescribe:

$$\log K_f = C1 * \log(K_{ow}) + C2 * \log(f_{CO}) + C3 * \log(\Phi_{neu}) + I \quad (6)$$

Donde:

$$\Phi_{neu} = \frac{1}{1 + 10^{A(pH-pKa-B)}} \quad (7)$$

Con  $A = 1$  y  $B = 0$  para ácidos,  $A = -1$  y  $B = 2$  para bases, y  $A = 0$  y  $B = 0$  para compuestos no ionizables. De este modo,  $\log(\Phi_{neu}) = CFa$  para moléculas ácidas,  $\log(\Phi_{neu}) = CFb'$  para moléculas básicas, y  $\log(\Phi_{neu}) = 0$  para moléculas neutras.

Unificando los descriptores del modelo y aplicando antilogaritmo:

$$\log K_f = \log(K_{ow}^{C1} * f_{CO}^{C2} * \Phi_{neu}^{C3}) + I \quad (8)$$

$$K_f = 10^{\log(K_{ow}^{C1} * f_{CO}^{C2} * \Phi_{neu}^{C3}) + I} \quad (9)$$

En caso que  $I \ll \log(K_{ow}^{C1} * f_{CO}^{C2} * \Phi_{neu}^{C3})$ , y considerando que  $C1 \sim C2 \sim 1$ :

$$K_f \sim K_{ow} * f_{CO} * \Phi_{neu}^{C3} \quad (10)$$

Es decir, el modelo refleja la adsorción en la MO corregida por el pH, mientras que el intercepto  $I$  reflejaría la adsorción en la componente no orgánica del suelo, cuya magnitud sería independiente de la naturaleza del compuesto. De este modo, el modelo de Bintein y Devillers consideraría de forma explícita el grado de ionización de los herbicidas, pero no la variabilidad del suelo en términos de composición

(asume adsorción variable solo en la componente orgánica) ni de carga superficial (no contempla la presencia de carga variable, o efecto del  $PIE_{\text{suelo}}$ ).

Luego, reemplazando el valor de C3:

$$K_f = K_{ow} * f_{CO} * \phi_{neu}^{0,32} \quad (\text{ácidos}) \quad (11)$$

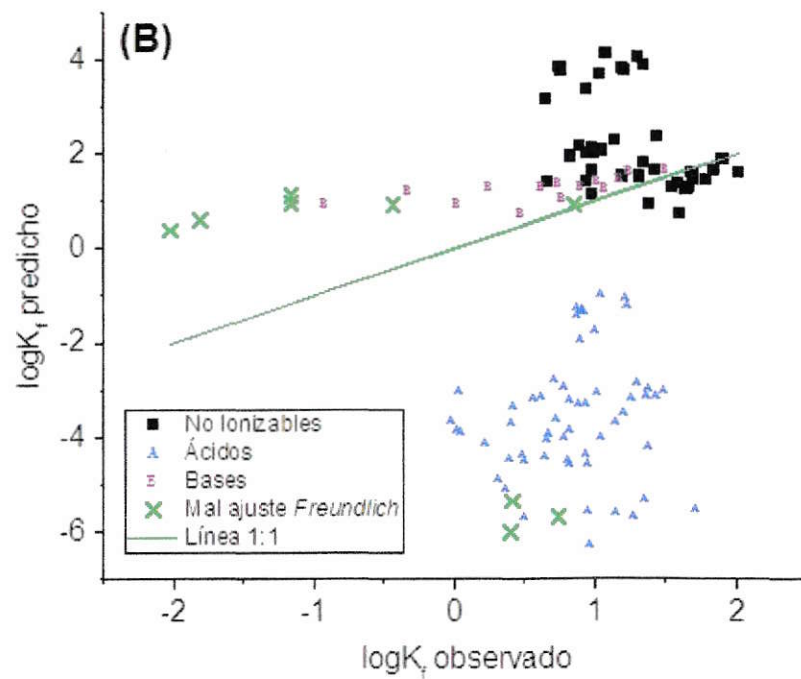
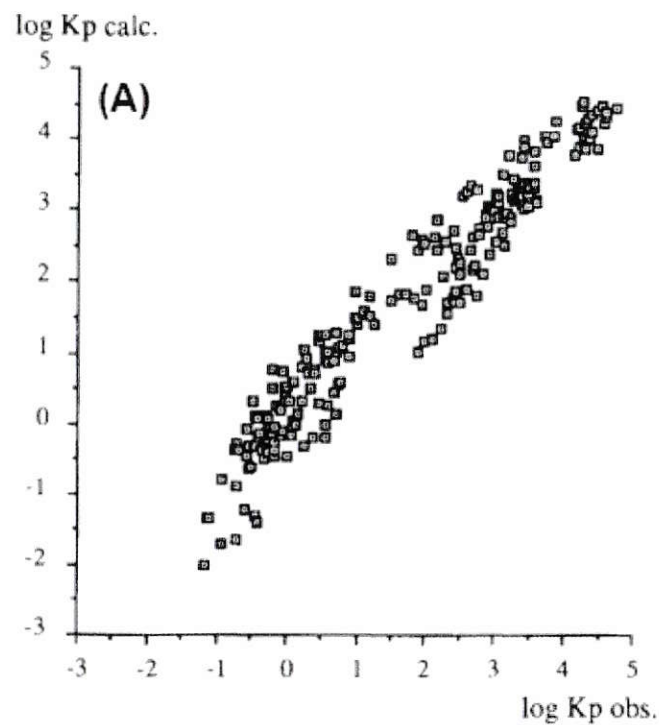
$$K_f = \frac{K_{ow} * f_{CO}}{\phi_{neu}^{0,55}} \quad (\text{bases}) \quad (12)$$

Con lo que la adsorción de compuestos ácidos disminuye a medida que aumenta su grado de ionización, lo que puede ser explicado por la mayor cantidad de aniones, favoreciéndose la repulsión electrostática del herbicida con la superficie negativa de minerales y MO del suelo, mientras que la adsorción de compuestos básicos aumenta con el grado de ionización, lo que se explicaría por el aumento de cationes en solución y con ello de la atracción electrostática con la superficie del suelo. Esto no es explícito en la definición del modelo, sino que surge del ajuste de los datos.

El ajuste al MBD de los valores de *endpoint* de los 12 herbicidas en 10 SDCV considerados en el presente trabajo se muestra en la Figura 5.

Las desviaciones observadas al aplicar el MBD en los valores previamente mencionados se exponen a continuación por tipo de herbicida:

- (i) Herbicidas no ionizables: Interacción con componentes del suelo diferentes de la MO (p. ej. arcilla) cuya variabilidad no es contemplada por el MBD.
- (ii) Herbicidas ácidos: El MBD no refleja la interacción con óxidos de Fe, incluso en presencia de carga superficial total negativa, subestimando la adsorción de la especie aniónica.



**Figura 5.** Ajuste comparativo del modelo de Bintein y Devillers (A) [30] y los valores de  $\log K_f$  del presente estudio (B).

- (iii) Herbicidas básicos: A menor adsorción se observa una mayor sobreestimación de los valores de actividad, lo que podría asociarse a la mayor significancia de la repulsión electrostática con los óxidos de Fe cuando otros procesos de adsorción, por ejemplo en MO, pierden relevancia.
- (iv) Herbicidas no ajustados al modelo *Freundlich*: Siguen las tendencias esperadas según su naturaleza, con una adsorción sobreestimada para SM (base) y subestimada para GFS (ácido). Por otro lado, DI ajusta bien al MBD, lo que puede asociarse a su carácter no ionizable, aumentando la significancia de la MO como adsorbente en el proceso de adsorción.

Adicionalmente, se debe tener presente que el MBD corrige el pH del suelo en CFb' en base a la minimización de los residuales en su set de entrenamiento, y por lo tanto se relaciona con propiedades específicas del conjunto de datos usado en su confección.

Con base en lo anterior, los parámetros del MBD son dependientes de su set de entrenamiento, con lo que la ecuación no es en esencia robusta. A esto se agrega que la aplicabilidad del modelo está restringida a isotermas lineales, o valores de  $C_{eq} = 1$  en caso de isoterma *Freundlich*, con lo cual su aporte al contexto regulatorio es limitado.

### 5.2.2. Modelo de Sabljic y col.

El modelo de Sabljic y col. (1995) (MS) fue propuesto para predecir la adsorción de 216 contaminantes no ionizables de uso agrícola, a partir de su  $K_{oc}$  promedio [60].

$$\log K_{oc} = 1,09 + 0,47 * \log K_{ow} \quad (13)$$

El MS considera solo un descriptor, con lo cual se espera un ajuste pobre pues podría no contemplar la variedad de mecanismos de adsorción e interacciones existentes.

Este modelo es aplicado a  $\log K_{oc}$ , por lo que debe ser adaptado a  $\log K_f$  para aplicarlo a los valores de *endpoint* de este trabajo. A partir de las ecuaciones B1, B2 y E4:

$$K_{oc} = \frac{K_d}{f_{CO}} \quad (14)$$

Aplicando logaritmo y reordenando:

$$\log K_d = \log K_{oc} + \log(f_{CO}) \quad (15)$$

Es posible relacionar  $K_d$  y  $K_f$  suponiendo que todos los herbicidas de interés se encuentran a una concentración  $C_{eq} = 1$  o poseen un valor de  $n = 1$  (ecuación 4).

Considerando esto en la ecuación 15:

$$\log K_f = \log K_{oc} + \log(f_{CO}) \quad (16)$$

Finalmente, el modelo queda expresado como:

$$\log K_f = 1,09 + 0,47 * \log K_{ow} + \log(f_{CO}) \quad (17)$$

Donde el coeficiente de  $\log(f_{CO})$  debe necesariamente ser 1.

Los resultados obtenidos del ajuste muestran tendencias similares al MBD, pero con desviaciones más pronunciadas (Figura 6).

Esto puede ser asociado a los supuestos más restrictivos del MS, según el cual la adsorción ocurre principalmente en la MO, pero independiente del grado de ionización. Esto tiene sentido, pues los datos usados en la confección del modelo

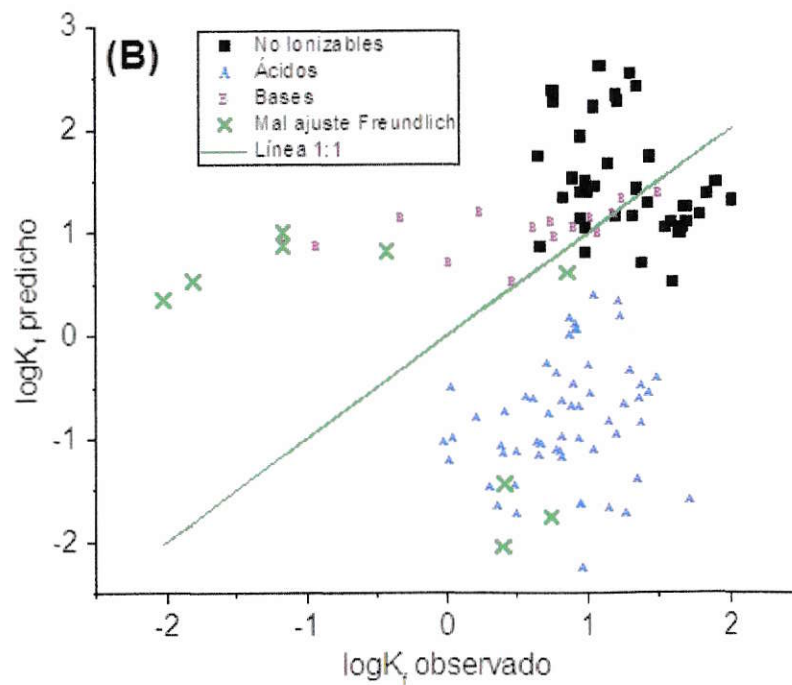
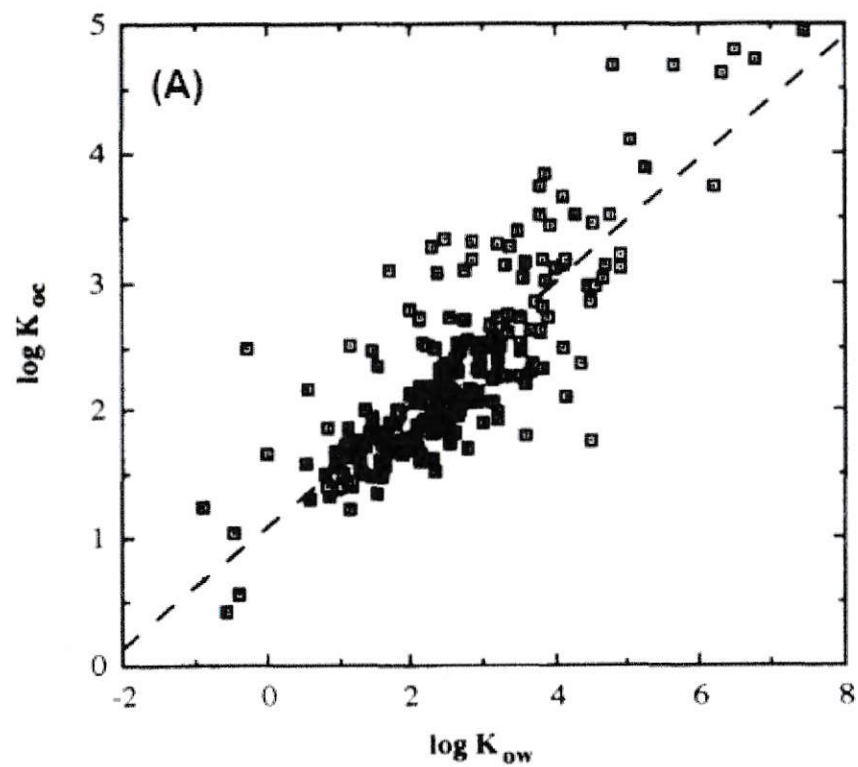


Figura 6. Ajuste comparativo del modelo de Sabljic y col. en función de  $K_{ow}$  (A) [60] y los valores de  $\log K_f$  del presente estudio (B).

corresponden a compuestos no ionizables. Sin embargo, restringe la aplicabilidad del modelo, que se supone surge con propósitos regulatorios.

En su estudio, los autores indican que la aplicabilidad del modelo desarrollado es descrita por primera vez de manera precisa y uniforme. De hecho, sugieren aplicar el modelo para una serie de agroquímicos, nombrando como única limitación que su valor de  $\log K_{ow}$  se encuentre entre -1 y 8. Sin embargo, no mencionan: (i) la procedencia de los datos usados y la metodología empleada para su obtención, (ii) el contenido de CO en los suelos, ni (iii) el cumplimiento de linealidad de la isoterma de adsorción (valor de n).

Además, cada valor de  $K_{oc}$  empleado corresponde al promedio de valores obtenidos para un mismo compuesto en diferentes suelos, con lo cual el modelo carece de aplicabilidad en condiciones reales y podría sesgar la interpretación mecanística del proceso de adsorción.

### 5.2.3. Modelo de Wen y col.

El modelo de Wen y col. (2012) (MW) fue creado con la finalidad de expandir el intervalo de aplicabilidad del descriptor  $\log K_{ow}$  en la predicción de  $\log K_{oc}$  de contaminantes orgánicos en suelos, cuya relación lineal se restringe a valores de  $\log K_{ow}$  entre 0,5 a 7,5 [39]. Para ello, los autores consideraron 594 compuestos obtenidos de literatura y generaron el siguiente modelo no lineal:

$$\log K_{oc} = 1,15 + 0,334 * \log K_{ow} + 0,0396(\log K_{ow})^2 + 0,00290(\log K_{ow})^3 - 0,000868 * e^{\log K_{ow}} \quad (18)$$

Esta ecuación pretende englobar de manera implícita los posibles cambios mecanísticos en el proceso de adsorción, asociados a interacciones hidrofílicas ( $\log K_{ow} < 0,5$ ) o altamente hidrofóbicas ( $\log K_{ow} > 7,5$ ), suponiendo que pueden ser predichas por cambios en  $\log K_{ow}$  sin necesidad de incorporar al suelo.

Respecto a la aplicabilidad de la ecuación 18, los autores exponen una serie de factores que afectan la adsorción de contaminantes en suelos, con la idea de explicar valores atípicos y puntos de influencia, siendo los más importantes: mecanismo de adsorción (polar respecto a apolar), solubilidad, ionización, degradación, volatilización, relación solución/suelo en estudios de adsorción y error experimental. De este modo, los autores asocian las desviaciones a la complejidad del proceso de adsorción y a la variabilidad metodológica dentro de su base de datos.

Adicionalmente, se tienen otros factores no mencionados por los autores que podrían dar cuenta de las desviaciones entre valores observados y predichos, a saber: (i) desconocimiento respecto del contenido de CO y tendencias de las isothermas de adsorción (lineal, no lineal), y (ii) los autores utilizan valores de  $\log K_{ow}$  considerando  $\text{pH} = 7$ , pero informan que los valores de  $\log K_{oc}$  provienen de suelos con valores de  $\text{pH}$  entre 3 y 8, con lo cual las desviaciones producto de la ionización de los herbicidas podrían ser significativas.

Para aplicar el modelo a los valores de *endpoint* del presente trabajo, se convierte  $\log K_{oc}$  a  $\log K_f$  usando la ecuación 16:

$$\log K_f = 1,15 + 0,334 * \log K_{ow} + 0,0396(\log K_{ow})^2 + 0,00290(\log K_{ow})^3 - 0,000868 * e^{\log K_{ow}} + \log(f_{CO}) \quad (19)$$



Al aplicar el MW a los valores de *endpoint* del presente trabajo se observa la misma tendencia obtenida en los MBD y MS (Figura 7). Sin embargo, la adsorción de herbicidas ácidos se aproxima más al valor esperado (línea 1:1) y su variabilidad disminuye.

Las desviaciones respecto a la línea 1:1 podrían incluir una componente mecánica, ionización y diferencias metodológicas entre los valores de *endpoint* del presente estudio y los valores usados en la confección del MW.

Para herbicidas no ionizables se observa un grupo aislado de valores (círculo rojo), correspondientes a TF. Su mayor desviación respecto a la línea 1:1 podría ser explicada por la mayor presión de vapor de TF en comparación al resto de herbicidas no ionizables (9,5 mPa a 20°C respecto a 0,00115 - 1,7 mPa a 20°C [68]), incidiendo en una mayor probabilidad de volatilización y una menor adsorción.

### **5.3. Análisis exploratorio de los descriptores propuestos**

Se descartaron 15 de un total de 186 descriptores moleculares y 0 de un total de 14 descriptores de SDCV tras eliminar descriptores constantes (varianza  $\leq 0,0001$ ) y altamente correlacionados ( $r \geq 0,9999$ ). Posteriormente, se realizó el análisis exploratorio a las matrices de descriptores para los 12 herbicidas y 10 SDCV aplicando ACP.

Aproximadamente el 80% de la variabilidad en la matriz de descriptores de SDCV es explicada por los tres primeros componentes principales (CP) del análisis exploratorio. Estos incluyen:

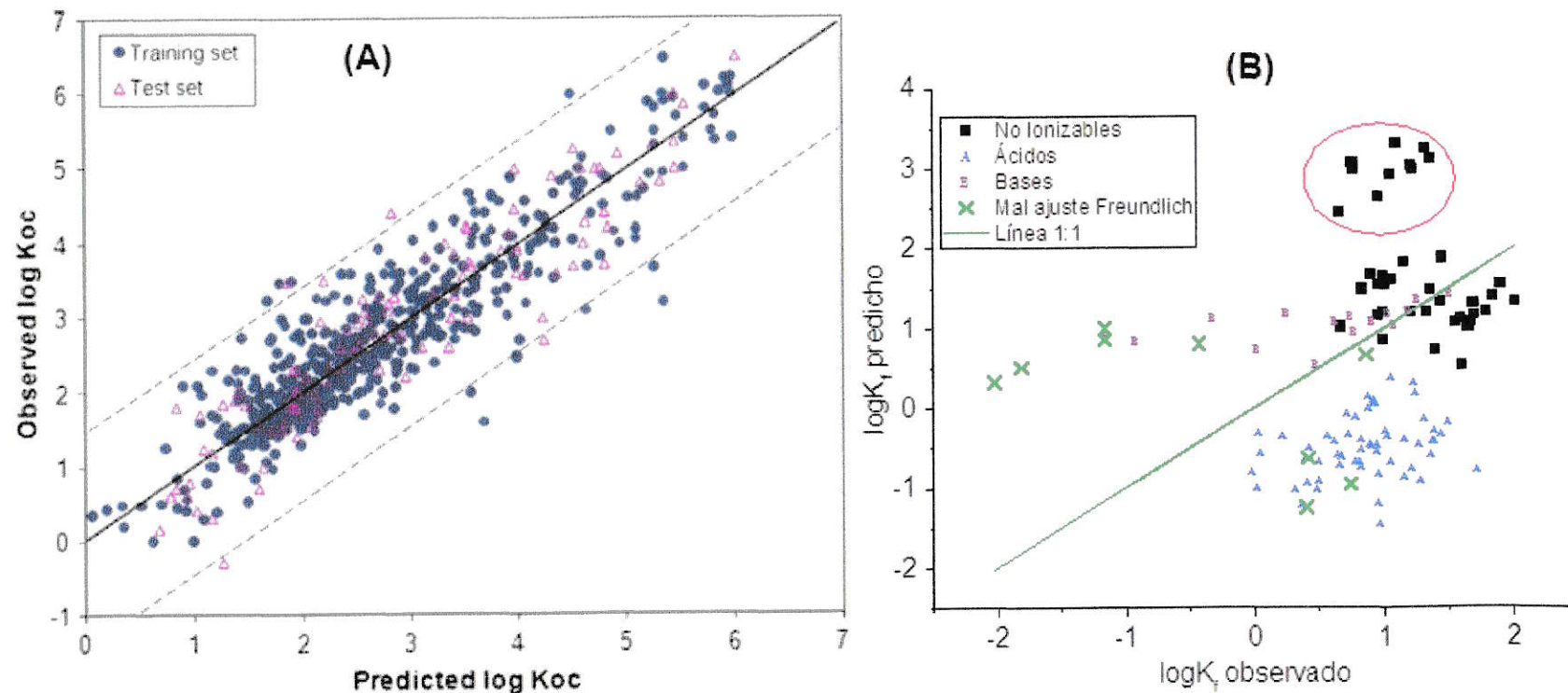


Figura 7. Ajuste comparativo del modelo de Wen y col. (A) [39] y los valores de  $\log K_f$  del presente estudio (B).

- (i) CP1 (35,0%): MO y su relación con los minerales, así como efecto de la carga variable y presencia de óxidos de Fe, lo que es reflejado en los descriptores  $\log f_{\text{CO}}$ ,  $R_{\text{Ar-CO}}$ ,  $\text{CF}_+$ ,  $\text{CF}_-$  y  $\log f_{\text{Fe DCB}}$ .
- (ii) CP2 (24,1%): CIC y fracción mineral, reflejado en los descriptores CIC y  $\log f_{\text{arcilla}}$ .
- (iii) CP3 (21,3%): Óxidos de Fe, reflejado en el descriptor  $\log f_{\text{Fe OX}}$ .

Lo anterior indica una posible relación entre la composición de los SDCV y su carga variable, lo que supone unificar ambas características en nuevos descriptores para estos suelos con la finalidad de reflejar la dependencia conjunta de la adsorción con la ambas características.

Además, los CP más relevantes incluyen descriptores para MO, arcilla y óxidos de Fe, lo que podría implicar que todos los componentes de la superficie de SDCV participan significativamente del proceso de adsorción.

Por otro lado, el 90% de la variabilidad en la matriz de descriptores moleculares es explicada por los cinco primeros CP, donde CP1 (51,3%) presenta la mayor parte de los descriptores topológicos, electrónicos, geométricos y constitucionales, así como algunos descriptores provenientes de *EPI SUITE*, de lo que se deduce una fuerte redundancia de información.

Además, se encontró una relación entre CP2 (17,9%) y los descriptores  $\log K_{\text{ow}}$ ,  $\log S_w$ ,  $\log S_p$  y  $\text{CFa}$ . De esta manera, los descriptores  $\log S_p$  y  $\text{CFa}$  no serían capaces de explicar la variabilidad asociada a CP1, resultando complementarios al resto de descriptores moleculares.

#### 5.4. Generación de descriptores para la interacción herbicida-SDCV

Los descriptores basados en la interacción herbicida-SDCV y generados a partir de una base teórico/conceptual son desarrollados en el Anexo F.

Aplicando los siguientes supuestos:

- $\sigma_{-}^{MO}$  y  $\sigma_{+}^{MO}$  no varían significativamente entre suelos, se incorporan estos términos en  $\alpha_{-}$  y  $\alpha_{+}$ , respectivamente, simplificando las expresiones de  $K_{f(CO)}$ .
- $\sigma_{+}^{Fe}$ ,  $\sigma_{0}^{Fe}$  y  $\sigma^{min}$  se calculan empleando las ecuaciones E6 y E7, considerando  $PIE_{suelo}$  para  $\sigma^{min}$  y  $PIE_{Fe}$  para  $\sigma_{+}^{Fe}$  y  $\sigma_{0}^{Fe}$ .
- $\phi_{neu} = 1$  para herbicidas no ionizables y bases (100% en su forma neutra), y  $\phi_{-} = 1$  para herbicidas ácidos (100% en su forma aniónica), debido a las condiciones experimentales de este trabajo (pH 7,0).
- $K_{ow}^{neu}$  y  $K_{ow}^{ion(-)}$  se consideran equivalentes a  $K_{ow}$  empíricos calculados a pH 7 y presentados en la Tabla C1, pues a dicho pH los ácidos se encuentran totalmente disociados y las bases completamente neutras.
- $Fe_{DCB}$  no participa del proceso de adsorción, por ser poco relevante frente a otras formas de Fe (Tabla F1). De este modo el valor de  $f_{Fe}$  en las ecuaciones F20, F21 y F22 es calculado según:

$$f_{Fe} = f_{FeOX} + f_{FePYRO} \quad (20)$$

Los descriptores para los herbicidas y suelos de este trabajo quedan definidos de la siguiente forma:

$$K_{f(\text{CO})} = \begin{cases} K_{\text{ow}} * f_{\text{CO}} & \text{Para herbicidas no ionizables y bases} \\ K_{\text{ow}} * f_{\text{CO}} * \left( \frac{1 - \sigma_{-}^{\text{min}}}{\text{CIC}} \right) & \text{Para herbicidas ácidos} \end{cases} \quad (21)$$

$$K_{f(\text{min})} = \begin{cases} \left( \frac{1 - |1 - 2\sigma_{+}^{\text{min}}|}{K_{\text{ow}}} \right) \left( \frac{m_{\text{suelo}}}{\rho_{\text{particula}}} \right) \left( \frac{f_{\text{arena}}}{\bar{r}_{\text{arena}}} + \frac{f_{\text{limo}}}{\bar{r}_{\text{limo}}} + \frac{f_{\text{arcilla}}}{\bar{r}_{\text{arcilla}}} \right) & \text{Neutros, bases} \\ \left( \frac{(\sigma_{+}^{\text{min}})^2}{K_{\text{ow}} * \text{CIC}} \right) \left( \frac{m_{\text{suelo}}}{\rho_{\text{particula}}} \right) \left( \frac{f_{\text{arena}}}{\bar{r}_{\text{arena}}} + \frac{f_{\text{limo}}}{\bar{r}_{\text{limo}}} + \frac{f_{\text{arcilla}}}{\bar{r}_{\text{arcilla}}} \right) & \text{Herbicidas ácidos} \end{cases} \quad (22)$$

$$K_{f(\text{Fe}^0)} = \frac{(f_{\text{FeOX}} + f_{\text{FePYRO}}) * (1 - \sigma_{+}^{\text{Fe}})}{K_{\text{ow}}} \quad \text{Para todos los casos} \quad (23)$$

$$K_{f(\text{Fe}^{+})} = \frac{(f_{\text{FeOX}} + f_{\text{FePYRO}}) * \sigma_{+}^{\text{Fe}} * (1 - f_{\text{CO}} - f_{\text{arcilla}})}{K_{\text{ow}}} \quad \text{Para todos los casos} \quad (24)$$

Donde los coeficientes correspondientes a herbicidas no ionizables y bases son  $a = \alpha_{\text{neu}}$ ,  $b = 3\beta_{\text{neu}}$ ,  $c = \beta_{\text{neu}}^{\text{Fe}^0}$  y  $d = \beta_{\text{neu}}^{\text{Fe}^{+}}$ , mientras que para herbicidas ácidos son  $a = \alpha_{-}$ ,  $b = 3\beta_{-}$ ,  $c = \beta_{-}^{\text{Fe}^0}$  y  $d = \beta_{-}^{\text{Fe}^{+}}$ .

De modo que el modelo (Q)SAR para la adsorción generado a partir de la contextualización es:

$$K_f = a * K_{f(\text{CO})} + b * K_{f(\text{min})} + c * K_{f(\text{Fe}^0)} + d * K_{f(\text{Fe}^{+})} \quad (25)$$

Para calcular cada descriptor, se supuso que:

- $\bar{r}_{\text{arena}} = 5,125 * 10^{-4} \text{ m}$  ,  $\bar{r}_{\text{limo}} = 1,3 * 10^{-5} \text{ m}$  y  $\bar{r}_{\text{arcilla}} = 5 * 10^{-7} \text{ m}$ , correspondientes al radio promedio del intervalo asociado a cada tipo de partícula según la clasificación del Departamento de Agricultura de Estados Unidos (USDA) [69]: diámetro entre 0,05 y 2,0 mm para arena, entre 0,002 y 0,05 mm para limo y menor a 0,002 mm para arcilla.
- $\rho_{\text{particula}}$  es función de la abundancia de alofán, considerando un valor de 1,9 g mL<sup>-1</sup> para alofán y un valor promedio de 2,6 g mL<sup>-1</sup> para el resto de

minerales (Base de datos de Mineralogía [70]). Como los suelos COLL y MET no tienen alofán, su densidad se considera  $2,6 \text{ g mL}^{-1}$ . El resto de suelos contiene alofán en abundancia ( $> 50\%$ ), estimándose su densidad en  $2,075 \text{ g mL}^{-1}$  a partir de un contenido de 75% de alofán (valor medio entre 50 y 100).

- $\text{PIE}_{\text{Fe}} = 6,3$ , correspondiente al PIE del alofán en solución  $\text{CaCl}_2$  [71]. Si bien los suelos COLL y MET no tienen alofán, se consideró el mismo valor de PIE para representar sus óxidos de Fe.

A diferencia de la metodología convencional, este modelo (Q)SAR es creado con una base física y química específica y, por lo tanto, no requiere de selección de la mejor combinación de descriptores.

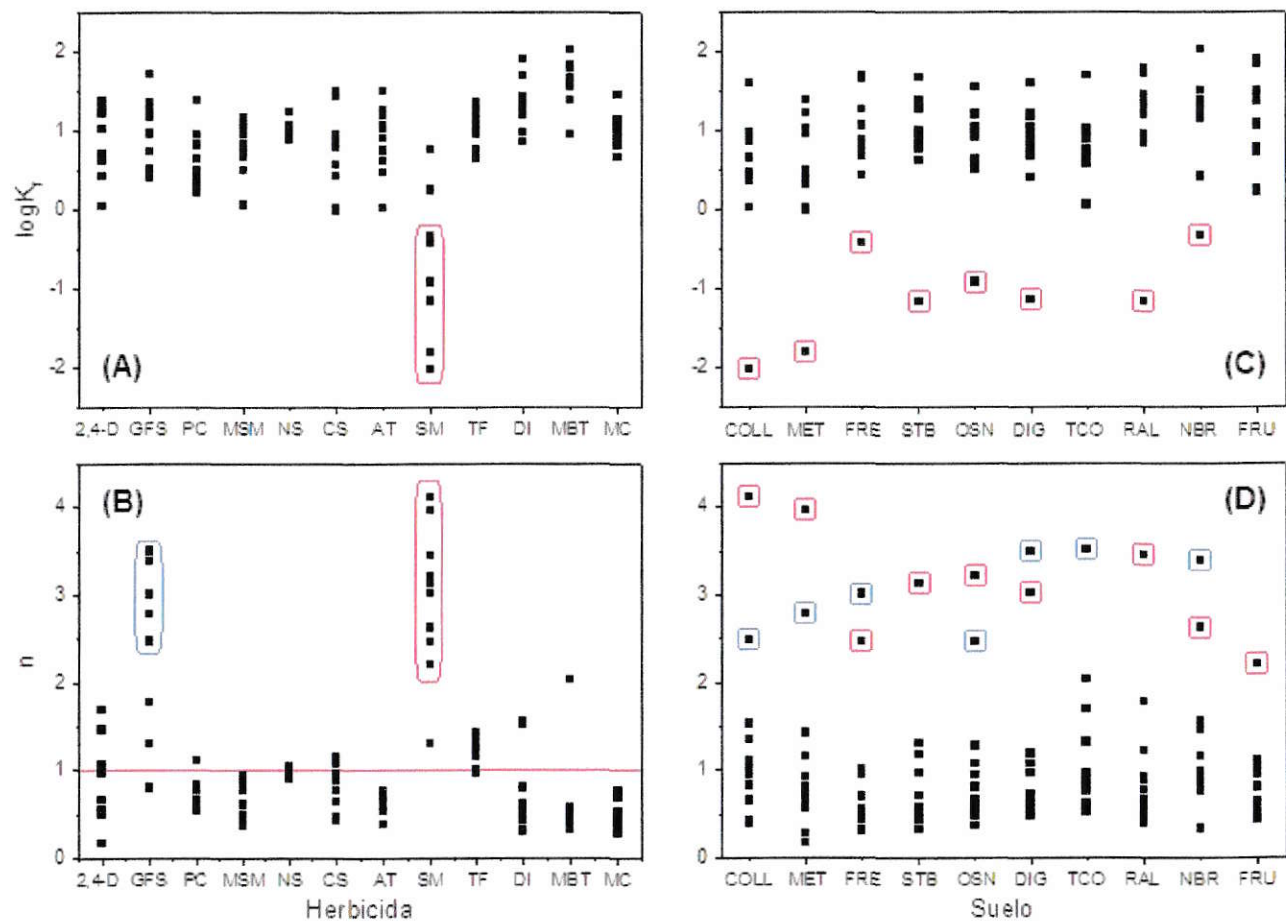
### **5.5. Análisis de los valores de *endpoint* de herbicidas en SDCV**

Los valores de *endpoint* por herbicida y SDCV se muestran en la Figura 8, mientras que los valores de *endpoint* en función del tipo de herbicida (ionizables, no ionizables) se presentan en la Figura 9.

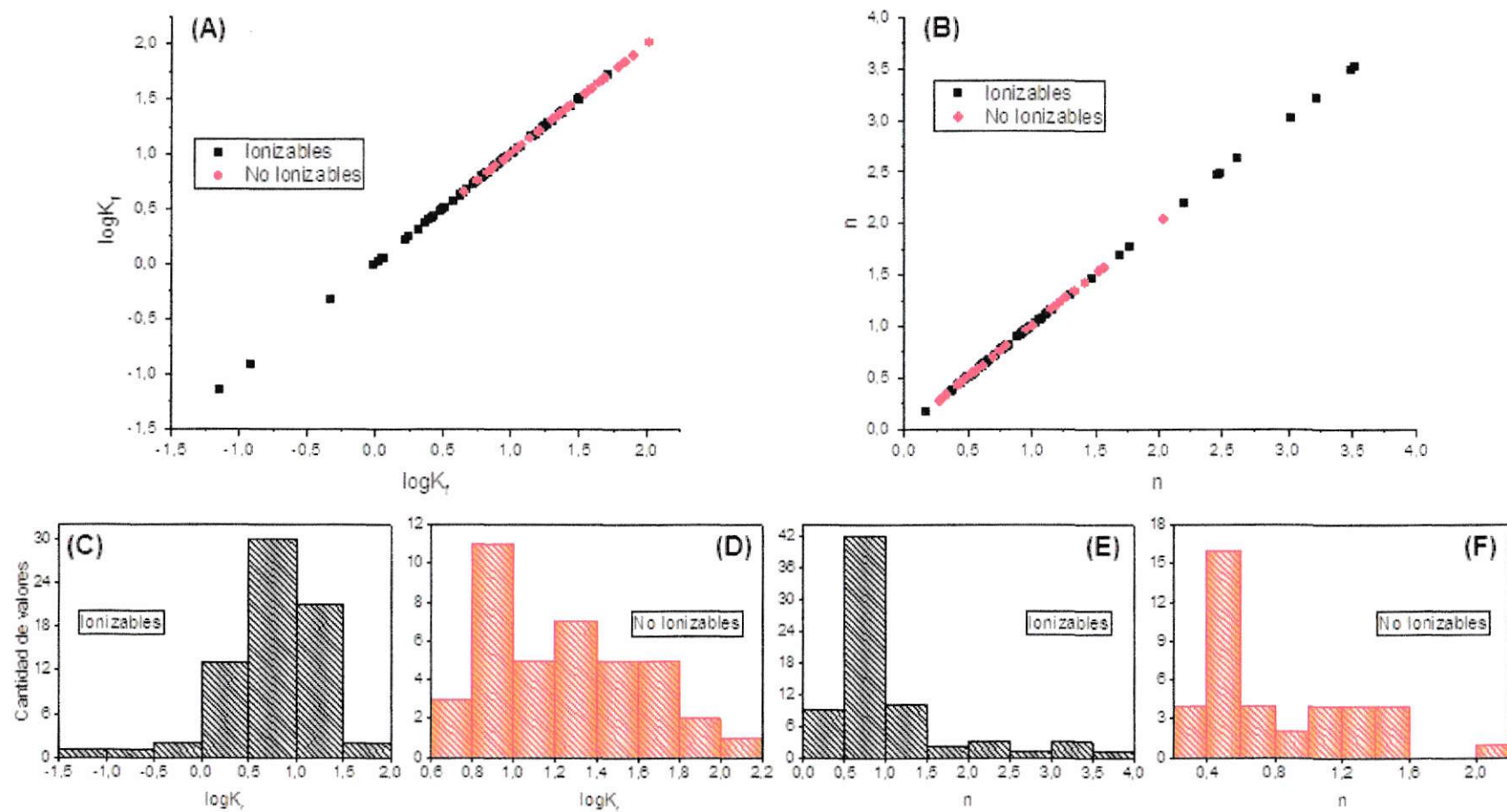
Los valores de *endpoint* son analizadas a continuación, a partir de ambas figuras.

#### **5.5.1. Variabilidad de los valores de *endpoint* en función del herbicida**

La variabilidad de los valores de *endpoint* en función de cada herbicida es presentada en las Figuras 8A y 8B. Su análisis por herbicida es realizado de manera general sin la creación de modelos (Q)SAR, debido al incumplimiento de dos requisitos estadísticos (sección 4.2.8 y Anexo E), a saber: (i) se tiene menos de 30 valores de *endpoint* por herbicida, y (ii) la variación de  $\log K_f$  y  $n$  presenta diferencias entre valores superiores al 15% del valor del intervalo de *endpoint*.



**Figura 8.** Valores de *endpoint* ( $\log K_f$  y  $n$ ) en función de cada herbicida (A, B) y SDCV (C, D).



**Figura 9.** Valores de  $K_f$  y  $n$  de todos los herbicidas en SDCV en función del tipo de herbicida.



En general, los valores de  $\log K_f$  de herbicidas no ionizables poseen mayor magnitud respecto a herbicidas ionizables (Figura 8A), así como un intervalo de valores más acotado (Figura 9A) y una distribución más amplia (Figuras 9C y 9D).

Las tendencias observadas para la magnitud de la adsorción de herbicidas no ionizables pueden asociarse a su mayor afinidad con la MO y menor repulsión electrostática [11, 25], mientras que las tendencias observadas para herbicidas ionizables pueden relacionarse a la mayor variedad de mecanismos de adsorción involucrados en el proceso de adsorción de estos herbicidas en SDCV (Tabla F1).

Por otra parte, el intervalo de valores de  $n$  de herbicidas no ionizables es similar al de herbicidas ionizables, exceptuando SM y GFS (Figuras 8B y 9B). Además, MSM y AT poseen únicamente valores de  $n < 1$ , y SM posee valores de  $n > 1$ , mientras que el resto de herbicidas presentan principalmente valores tanto de  $n < 1$  como de  $n > 1$  (Figuras 8B, 9B, 9E y 9F), implicando un cambio de tendencia en la isoterma de adsorción que estaría necesariamente relacionado con las propiedades de cada SDCV y su interacción con los herbicidas, lo que impide el uso de  $K_d$  como *endpoint*.

Respecto al impacto de SDCV en cada herbicida, una propiedad determinante en el valor de  $n$  es la porosidad y el tamaño de poros, de modo que una alta presencia de macroporos permite la formación de multicapas ( $n > 1$ ), mientras que una alta presencia de microporos impide su formación ( $n < 1$ ), siempre que las propiedades moleculares de cada herbicida le permitan unirse a sus congéneres [72, 73].

Las condiciones de agitación en la metodología *batch* bajo la cual se realizaron los estudios de adsorción de este trabajo impiden asociar de manera confiable los valores de  $n$  con la porosidad de SDCV, que podría ser dependiente de condiciones locales

como la oclusión de poros, o incluso presentar tamaño de poros dependiente del tiempo.

Para ejemplificar la incerteza entre la relación entre porosidad y valores de  $n$  mencionada anteriormente, se presentan dos estudios de adsorción de herbicidas en SDCV empleando metodología *batch*:

- (i) Cáceres-Jensen y col. (2010) señalan que el herbicida ionizable MSM presenta dos mecanismos cinéticos de adsorción en SDCV: transferencia de masa y difusión intrapartícula, siendo solo este último mecanismo el que controló la adsorción de MSM en Ultisoles [11]. De esta manera, el menor tamaño de poros asociado a la difusión intrapartícula podría explicar los valores de  $n < 1$  para MSM en COLL y MET (Ultisoles).
- (ii) Cáceres-Jensen y col. (2013) indican que la cinética de adsorción del herbicida no ionizable DI es gobernada por mecanismo de transferencia de masa tanto en Ultisoles como en Andisoles [74]. Por lo tanto, se esperaría un valor de  $n > 1$  en la mayor parte de los SDCV del presente estudio, pues la transferencia de masa ocurre en la superficie de SDCV a nivel de macroporos. Sin embargo, en la mayoría de SDCV es observada la tendencia contraria (Figura 8B).

Respecto a la variabilidad observada para  $\log K_f$  y  $n$  en la Figura 8, el herbicida SM presenta la mayor variabilidad en los valores de  $\log K_f$  (Figura 8A) y los herbicidas SM y GFS poseen la mayor variabilidad en los valores de  $n$  (Figura 8B). Por otro lado, el herbicida con menor variabilidad en  $\log K_f$  y  $n$  es NS (Figuras 8A y 8B).

Lo anterior implica que la adsorción de herbicidas ionizables es afectada en mayor o menor medida por SDCV con una variabilidad superior a la observada en herbicidas no ionizables, debido a la mayor variedad de mecanismos de adsorción. En este sentido, herbicidas no ionizables cuyo mecanismo de adsorción se relaciona principalmente con MO (menor variabilidad) presentan intervalos de  $\log K_f$  similares entre sí.

La mayor parte de los valores de  $n$  asociados a los herbicidas SM y GFS son mucho mayores a 1 en comparación al resto de herbicidas (Figura 8B, ver línea roja y rectángulos rojo y azul). Esto implica que el mecanismo de adsorción de SM y GFS ocurre mediante: (i) adsorción del herbicida sobre sus congéneres ya adsorbidos (formación de multicapa), y/o (ii) adsorción consecutiva en varios componentes del suelo mediante desbloqueo de sitios de adsorción internos, producto de la adsorción del herbicida [75].

Por otra parte, la adsorción de GFS ocurre a través del grupo amino o por formación de complejos mono y bidentados entre el grupo fosfonato y la superficie interna de los minerales [75].

Los mecanismos particulares de SM y GFS antes mencionados podrían explicar la baja calidad estadística del modelo de *Freundlich* (sección 5.1).

### **5.5.2. Variabilidad de los valores de *endpoint* en función de SDCV**

La variabilidad de los valores de *endpoint* en función de cada SDCV es presentada en las Figuras 8C y 8D. Su análisis es realizado de manera general sin la creación de

modelos (Q)SAR, debido al incumplimiento de los mismos dos requisitos estadísticos mencionados en la sección 5.5.1.

En términos generales, la variabilidad de los valores de  $\log K_f$  asociados al herbicida SM es extendida a todos los SDCV, incrementando el intervalo y variabilidad de valores de  $\log K_f$  (Figura 8C, recuadros rojos). Lo mismo ocurre con la variabilidad de los valores de  $n$  asociados a SM y GFS (Figura 8D, recuadros rojos y azules).

Luego, al descartar estos valores del análisis de tendencias, se observa que la variabilidad de los valores de  $\log K_f$  y  $n$  es similar en todos los SDCV. Lo anterior significa que la variabilidad asociada a los herbicidas es independiente del tipo de SDCV y, por lo tanto, podría estar relacionada a propiedades más generales de SDCV, como la carga variable o la cantidad de óxidos de Fe.

### **5.5.3. Variabilidad de los valores de *endpoint* de herbicidas no ionizables en SDCV bajo la aplicación de $\log K_f$ y $\log K_{oc}$**

Se analizó la reducción de la variabilidad de los valores de *endpoint* de los herbicidas no ionizables en SDCV comparando  $K_f$ ,  $K_{oc}$ ,  $\log K_f$  y  $\log K_{oc}$ , utilizando los valores de *endpoint* validados del herbicida ionizable GFS (sección 5.1) como contraste para determinar el sentido físico de las tendencias observadas.

La variabilidad de los valores de *endpoint* mencionados previamente es presentada en la Tabla 1.

Se observa una reducción de variabilidad al convertir: (i)  $K_f$  a  $\log K_f$ , (ii)  $K_f$  a  $K_{oc}$ , (iii)  $K_{oc}$  a  $\log K_{oc}$ , y (iv)  $\log K_f$  a  $\log K_{oc}$ , para los herbicidas no ionizables y GFS.

**Tabla 1.** Efecto del cambio de escala y la normalización al contenido de CO en la variabilidad de los valores de *endpoint*.

Valor de actividad	Análisis estadístico	Herbicidas no ionizables				Contraste
		TF	DI	MBT	MC	GFS
K <sub>f</sub>	Promedio	12,27	31,17	47,30	11,05	18,65
	Desviación estándar	6,22	23,64	26,32	6,42	16,21
	Variación	<b>50,69%</b>	<b>75,86%</b>	<b>55,65%</b>	<b>58,15%</b>	<b>86,97%</b>
log K <sub>f</sub>	Promedio	1,03	1,38	1,60	0,99	1,14
	Desviación estándar	0,24	0,34	0,29	0,21	0,38
	% variación	<b>23,63%</b>	<b>24,83%</b>	<b>17,97%</b>	<b>20,71%</b>	<b>33,19%</b>
K <sub>oc</sub>	Promedio	237,56	536,38	970,01	210,41	346,19
	Desviación estándar	103,46	276,70	644,58	92,53	257,72
	Variación	<b>43,55%</b>	<b>51,59%</b>	<b>66,45%</b>	<b>43,98%</b>	<b>74,44%</b>
log K <sub>oc</sub>	Promedio	2,33	2,68	2,90	2,29	2,42
	Desviación estándar	0,23	0,21	0,32	0,17	0,37
	Variación	<b>9,82%</b>	<b>7,81%</b>	<b>10,96%</b>	<b>7,42%</b>	<b>15,44%</b>

Las tendencias observadas guardan relación con la afinidad de los herbicidas no ionizables por la MO, y la reducción de la variabilidad de los valores de *endpoint* fue mayor para herbicidas no ionizables que GFS. Sin embargo, GFS se encuentra como anión en las condiciones experimentales del presente trabajo (solución a pH neutro), por lo que existe reducción de la variabilidad de los valores de *endpoint* incluso en presencia de varios adsorbentes en SDCV.

Con base en lo anterior, la reducción de variabilidad de los valores de *endpoint* en  $\log K_{oc}$  se puede interpretar como efecto del cambio de escala en conjunto con la relevancia de la MO en el proceso de adsorción de herbicidas en general. Lo anterior convierte a  $\log K_{oc}$  en un parámetro útil en un sentido estadístico, pero de sentido físico incierto debido a la subestimación del impacto de otros adsorbentes en SDCV.

#### **5.6. Tratamiento estadístico de las bases de datos empleadas en la generación de modelos (Q)SAR para adsorción de herbicidas en SDCV**

Los doce herbicidas fueron agrupados en no ionizables, ácidos y bases. Las bases de datos para la generación de modelos (Q)SAR para predecir la adsorción de herbicidas en SDCV fueron validadas mediante análisis de su calidad estadística y curación de la matriz de descriptores (sección 4.2.7).

Se eliminaron los siguientes herbicidas:

- (i) Herbicidas no ionizables: MBT en suelo TCO por diferencia  $> 15\%$  con el valor de  $n$  que le precede en magnitud.

- (ii) Herbicidas ácidos: GFS en OSN por diferencia  $> 15\%$  para el valor de  $n$  que le precede en magnitud. Se eliminan también los valores de GFS en COLL, DIG y TCO, por ser posteriores al valor de *endpoint* eliminado.

No fue posible crear un modelo (Q)SAR para predecir la adsorción de los herbicidas básicos, debido a la existencia de solo 15 valores de *endpoint* asociados, insuficientes para asegurar la calidad estadística del modelo (sección 4.2.8 y Anexo E).

Con base en lo anterior, se generan dos bases de datos orientadas a la creación de modelos (Q)SAR para predecir la adsorción de herbicidas no ionizables, por un lado, y herbicidas ácidos, por otra parte, en SDCV.

A fin de poner a prueba los modelos, se evalúa posteriormente su predictividad en:

- (i) Valores descartados por pobre ajuste del modelo *Freundlich* (sección 5.1).
- (ii) Valores descartados al validar las bases de datos (en esta sección).
- (iii) Valores de actividad de herbicidas básicos.

En la curación de la matriz de descriptores, se observa una mayor reducción de descriptores moleculares en comparación con SDCV para ambas bases de datos (Tabla 2), lo que se corresponde con la reducción de dimensionalidad sugerida por el ACP (sección 5.3).

Además, existe una menor reducción del número de descriptores moleculares para herbicidas ionizables, lo que sugiere una relación más compleja entre sus valores de  $\log K_f$  y  $n$  y los descriptores, que podría ser atribuida a la complejidad de sus mecanismos de adsorción, sujeta a la interacción con MO, arcillas y óxidos de Fe en SDCV (Tabla F1).

**Tabla 2.** Análisis de la reducción de descriptores en la generación de modelos (Q)SAR para la predicción de la adsorción de herbicidas en SDCV.

Tipo de descriptor	Análisis de reducción	Herbicidas no ionizables		Herbicidas ácidos	
		log K <sub>f</sub>	n	log K <sub>f</sub>	n
Moleculares	Nº inicial	184	184	185	185
	Nº final	6	4	39	41
	Reducción	<b>96,7%</b>	<b>97,8%</b>	<b>78,9%</b>	<b>77,8%</b>
SDCV	Nº inicial	14	14	14	14
	Nº final	12	12	12	12
	Reducción	<b>14,3%</b>	<b>14,3%</b>	<b>14,3%</b>	<b>14,3%</b>



## **5.7. División de los valores de *endpoint* entre set de entrenamiento y set de prueba**

Se dividieron los valores de *endpoint* en cada base de datos en sets de entrenamiento y de prueba, utilizando como *endpoint*: (i)  $\log K_f$  y  $n$  cuando el modelo fue generado por selección de la mejor combinación de descriptores, y (ii)  $K_f$  cuando el modelo fue generado por aplicación de la ecuación propuesta a partir de la contextualización del sistema herbicida-SDCV (sección 5.4).

En todos los casos mencionados, el set de prueba incluyó al menor valor de *endpoint* del intervalo total, lo que se corrigió de la siguiente forma:

- (i) En los sets de prueba de herbicidas no ionizables, el valor fue transferido al set de entrenamiento.
- (ii) En los sets de prueba de los herbicidas ácidos, el valor fue intercambiado con el segundo menor valor de *endpoint* en los sets de entrenamiento, para así mantener el mínimo número de datos aceptable en el set de prueba (10).

## **5.8. Creación y validación de modelos (Q)SAR para predecir la adsorción de herbicidas en SDCV**

### **5.8.1. Calidad de modelos (Q)SAR para predecir adsorción de herbicidas no ionizables**

Se crearon dos modelos (Q)SAR para predecir  $\log K_f$  de herbicidas no ionizables a partir de la selección de la mejor combinación de descriptores, a saber:

$$\log K_f = (2,40 \pm 0,35) - (0,42 \pm 0,34) * \log f_{\text{arcilla}} - (0,27 \pm 0,13) * \log f_{\text{Fe}_{\text{DCB}}} - (4,09 \pm 0,55) * \text{Weta2.unity} \quad (26)$$

$$\log K_f = (2,98 \pm 0,44) - (0,59 \pm 0,33) * \log f_{\text{arcilla}} + (0,28 \pm 0,14) * \log f_{\text{Fe}_{\text{PYRO}}} - (0,19 \pm 0,13) * \log f_{\text{Fe}_{\text{DCB}}} - (4,06 \pm 0,51) * \text{Weta2.unity} \quad (27)$$

Donde  $\log f_{\text{arcilla}}$ ,  $\log f_{\text{Fe}_{\text{DCB}}}$  y  $\log f_{\text{Fe}_{\text{PYRO}}}$  corresponden a la forma fraccional logarítmica del contenido de arcilla,  $\text{Fe}_{\text{DCB}}$  y  $\text{Fe}_{\text{PYRO}}$ , respectivamente, mientras que  $\text{Weta2.unity}$  corresponde a la densidad WHIM direccional para  $m = 2$ , también denominado  $\text{WDEN}$ ,  $\text{WHIM eta 2}$ , o  $\eta_2$ , y está basado en la proyección de los átomos de una molécula en el espacio tridimensional, capturando información 3-D del tamaño molecular, forma, simetría y distribución de dichos átomos [76].

Además, se obtuvo el siguiente modelo (Q)SAR para predecir  $K_f$  de herbicidas no ionizables a partir de las ecuaciones 21, 22, 23 y 24 (sección 5.4):

$$K_f = (13,00 \pm 5,62) + (0,00018 \pm 0,00059) * K_{f(\text{CO})} + (0,23 \pm 0,06) * K_{f(\text{min})} + (3011209 \pm 1570841) * K_{f(\text{Fe0})} + (161400 \pm 245469) * K_{f(\text{Fe+})} \quad (28)$$

Las ecuaciones 26, 27 incluyen descriptores moleculares y de SDCV, incluyendo contenido de óxidos de Fe, mientras que la ecuación 28 considera la interacción herbicida-SDCV. Por otra parte, los modelos (Q)SAR para predecir adsorción de contaminantes no ionizables en SCP no incluyen descriptores de SCP (Tabla A2), debido al uso de valores de  $\log K_{\text{oc}}$  promedio a partir de diferentes suelos (Tabla A1).

Lo anterior significa que la creación de modelos (Q)SAR para predecir la adsorción de herbicidas no ionizables en SDCV requiere:

- (i) Evitar promediar valores de *endpoint* entre diferentes SDCV.
- (ii) Considerar el impacto de la carga variable, en este caso mediante descriptores relacionados con el contenido de óxidos de Fe.

Respecto a la calidad de los modelos generados (Tabla G1), las ecuaciones 26 y 27 cumplen con todos los criterios de validación interna y externa, exceptuando  $EEN < 0,1$ . Además, se descarta la existencia de error sistemático en los ambos modelos (Q)SAR ( $EA \ll EMA$ ).

Comparativamente, la ecuación 27 tiene mayor  $R_{adj}^2$ , mayor FIT y menor CIA que la ecuación 26, convirtiéndose en la mejor alternativa al momento de aplicar dichos modelos (Q)SAR con propósitos regulatorios. Adicionalmente, la ecuación 26 presenta mayor predictividad, mientras que la ecuación 27 posee mayor robustez.

Por otra parte, ambas ecuaciones presentan coeficientes con elevado error, todos asociados a descriptores de SDCV ( $\log f_{arcilla}$ ,  $\log f_{Fe\ DCB}$ ,  $\log f_{Fe\ PYRO}$ ), lo que genera incertezas importantes en la predicción de  $\log K_f$ .

La ecuación 28 presenta buena bondad del ajuste (exceptuando EEN) y es considerada producto de una correlación no aleatoria ( $cR_p^2 > 0,5$ ), pero posee mala robustez y predictividad, por lo que no puede ser utilizada con propósitos regulatorios.

La calidad estadística de la ecuación 28, basada en la contextualización de los sistemas herbicida-SDCV, es menor en comparación a sus equivalentes obtenidas de la selección de la mejor combinación de descriptores (ecuaciones 26 y 27). Sin embargo, dicha comparación no contempla la misma escala de valores. Para evitar

este problema, se determinó el valor logarítmico de los valores de  $K_f$  observados y predichos para la ecuación 28 y se generó un nuevo modelo tal que:

$$\log K_{f\text{ obs}} = A + B * \log K_{f\text{ pred (set entrenamiento)}} \quad (29)$$

Para evitar un sesgo interpretativo debido al cambio en el número de descriptores, se limita el análisis de esta nueva ecuación a los parámetros  $R^2$ ,  $cR_p^2$ ,  $Q_{LOO}^2$ ,  $Q_{F1}^2$  y  $Q_{F2}^2$ .

Además, se usaron los valores de A y B para evaluar la correspondencia entre  $\log K_{f\text{ obs}}$  y  $\log K_{f\text{ pred}}$  y la utilidad de la ecuación con propósitos regulatorios, solo en caso que los parámetros estadísticos indicaran buena calidad del ajuste. Idealmente, se espera que  $A = 0$  y  $B = 1$ , con errores relativamente bajos.

Con base en lo anterior, a partir de la ecuación 28 se obtiene:

$$\log K_f = (0,58127 \pm 0,11625) + (0,59952 \pm 0,08615) * \log K_{f\text{ pred}} \quad (30)$$

Los parámetros estadísticos asociados a la ecuación 30 son:  $R^2 = 0,6780$ ,  $cR_p^2 = 0,6670$ ,  $Q_{LOO}^2 = 0,60304$ ,  $Q_{F1}^2 = 0,39961$ ,  $Q_{F2}^2 = 0,39292$ .

Comparando las ecuaciones 28 (Tabla G1) y 30, el cambio de escala desde  $K_f$  a  $\log K_f$  mejora los parámetros  $cR_p^2$  y  $Q_{LOO}^2$  del modelo, de modo que la forma logarítmica cumple los criterios estadísticos de robustez. Sin embargo, la predictividad sigue siendo baja.

Además, los valores asociados a intercepto y pendiente de la ecuación 30 dan cuenta de desviaciones importantes respecto a los valores ideales, que podrían relacionarse con deficiencias en la base teórico/conceptual de la ecuación 28 (Anexo F).

### 5.8.2. Calidad de modelos (Q)SAR para predecir adsorción de herbicidas ácidos

Se crearon dos modelos (Q)SAR para predecir  $\log K_f$  de herbicidas ácidos a partir de la selección de la mejor combinación de descriptores, a saber:

$$\log K_f = (2,81 \pm 0,67) - (0,73 \pm 0,45) * \log f_{\text{arcilla}} - (0,18 \pm 0,04) * \text{CIC} \\ + (0,59 \pm 0,25) * \log f_{\text{Fe}_{\text{OX}}} + (0,05 \pm 0,01) * \text{nAtomLC} \quad (31)$$

$$\log K_f = (4,13 \pm 1,18) - (0,88 \pm 0,38) * \log f_{\text{CO}} - (0,16 \pm 0,04) * \text{CIC} \\ + (0,64 \pm 0,27) * \log f_{\text{Fe}_{\text{DCB}}} + (0,05 \pm 0,01) * \text{nAtomLC} \quad (32)$$

Donde  $\log f_{\text{arcilla}}$ ,  $\log f_{\text{Fe}_{\text{OX}}}$  y  $\log f_{\text{Fe}_{\text{DCB}}}$  corresponden a la forma fraccional logarítmica del contenido de arcilla,  $\text{Fe}_{\text{OX}}$  y  $\text{Fe}_{\text{DCB}}$ , respectivamente, y  $\text{nAtomLC}$  representa el número de átomos de la cadena más larga en una molécula.

Además, se obtuvo el siguiente modelo (Q)SAR para predecir  $K_f$  de herbicidas ácidos a partir de las ecuaciones 21, 22, 23 y 24 (sección 5.4):

$$K_f = (7,37 \pm 1,08) + (2427 \pm 1613) * K_{f(\text{CO})} + (3 \pm 6) * 10^{-5} * K_{f(\text{min})} - (2,20 \\ \pm 1,85) * K_{f(\text{Fe}^0)} + (1,36 \pm 0,63) * K_{f(\text{Fe}^+)} \quad (33)$$

La ventaja de la ecuación 33 radica en la rápida detección de incongruencias respecto al sentido físico. En este caso, el bajo coeficiente asociado a  $K_{f(\text{min})}$  fue asociado a la alta repulsión electrostática, mientras que la falta de sentido físico para  $K_{f(\text{Fe}^0)}$  se asoció a problemas teórico/conceptuales en la definición del modelo. Lo anterior permite eliminar ambos descriptores para la generación de una nueva ecuación:

$$K_f = (7,18 \pm 1,08) + (2510 \pm 1619) * K_{f(\text{CO})} + (0,98 \pm 0,14) * K_{f(\text{Fe}^+)} \quad (34)$$

Respecto a la calidad de los modelos generados (Tabla G2), las ecuaciones 31 y 32 cumplen con todos los criterios de validación interna y externa, exceptuando  $R^2 \geq 0,6$  y  $EEN < 0,1$ .

Comparativamente, los parámetros  $R_{adj}^2$ , FIT y CIA de las ecuaciones 31 y 32 son similares, impidiendo la selección de alguna de estas ecuación con propósitos regulatorios por sobre la otra. De igual modo, ambas ecuaciones poseen baja calidad interna y externa, expresada en valores insuficientes para los parámetros de robustez y predictividad, incluyendo probabilidad de correlación aleatoria ( $cR_p^2 < 0,5$ ).

Además, la correlación entre  $\log f_{Fe\ DCB}$  y el resto de descriptores en la ecuación 32 se encuentra en el límite de lo aceptable ( $FIV(\log f_{Fe\ DCB}) \sim 5$ ), disminuyendo su calidad estadística frente a la ecuación 31.

La ecuación 34 cumple con el criterio  $cR_p^2 \geq 0,5$  para la aleatorización en Y, de modo que puede ser considerada un modelos (Q)SAR no aleatorio. No obstante, no se logra validar su uso con propósitos regulatorios.

Con base en lo anterior, ninguno de los modelos (Q)SAR para predecir la adsorción de herbicidas ácidos en SDCV puede ser usado con propósitos regulatorios.

Lo anterior se corresponde con la mayor dificultad asociada a la creación de modelos (Q)SAR para la adsorción de herbicidas ionizables en SCP respecto a herbicidas no ionizables, observada en menores valores de  $R^2$  [30, 41, 43] y la necesidad de incorporar descriptores de SCP (Tabla A2), evitando promediar valores de *endpoint* en dichos modelos (Tabla A1).

Por último, aplicando la conversión de  $K_f$  en la ecuación 34 a su valor logarítmico mediante la ecuación 29:

$$\log K_f = -(0,1345 \pm 0,2880) + (1,0367 \pm 0,2987) * \log K_{f_{pred}} \quad (35)$$

Los parámetros estadísticos asociados a la ecuación 35 son:  $R^2 = 0,2271$ ,  $cR_p^2 = 0,2174$ ,  $Q_{LOO}^2 = 0,1757$ ,  $Q_{F1}^2 = 0,1226$ ,  $Q_{F2}^2 = 0,1224$ .

Al realizar el cambio de escala desde la ecuación 34 a la ecuación 35, la calidad estadística del modelo (Q)SAR se reduce, presentando ajuste inaceptable ( $R^2 < 0,6$ ) y correlación aleatoria ( $cR_p^2 < 0,5$ ). La robustez disminuye y la predictividad aumenta levemente sin cumplir el requisito mínimo ( $Q_{F1}^2 < 0,6$  y  $Q_{F2}^2 < 0,6$ ).

Adicionalmente, el intercepto y la pendiente de la ecuación 35 se acercan a sus valores ideales. No obstante, la baja calidad estadística antes mencionada impide asociar un sentido físico a dicha cercanía.

### 5.8.3. Calidad de modelos (Q)SAR para predecir n de herbicidas no ionizables

Se crearon dos modelos (Q)SAR para predecir n de herbicidas no ionizables en SDCV a partir de la selección de la mejor combinación de descriptores, a saber:

$$n = (0,03 \pm 0,30) - (0,22 \pm 0,17) * \log f_{arena} + (0,15 \pm 0,41) * \log f_{arcilla} + (1,25 \pm 0,22) * SC - 5 \quad (36)$$

$$n = (0,26 \pm 0,61) + (0,66 \pm 0,75) * \log f_{limo} + (0,77 \pm 0,63) * \log f_{arcilla} + (0,05 \pm 0,06) * CIC + (1,28 \pm 0,24) * SC - 5 \quad (37)$$

Donde  $\log f_{arena}$ ,  $\log f_{arcilla}$  y  $\log f_{limo}$  corresponden a la forma fraccional logarítmica del contenido de arena, arcilla y limo, respectivamente, y  $SC - 5$ ,

también denotado como  ${}^5\chi_C$ , es un tipo de Índice de Conectividad de Kier-Hall [77] conocido como Índice de Conectividad de Orden 5 o Cluster Simple de Orden 5.

En el caso de los herbicidas ácidos, toda combinación de descriptores generó modelos (Q)SAR para predecir  $n$  en SDCV con baja correlación ( $R^2 \leq 0,15$ ) y presencia de puntos de influencia ( $Q_{LOO}^2 < 0$ ). Entonces, ningún caso cumplió con los criterios de calidad estadística necesarios para asegurar confianza (sección 4.3).

Del análisis de la calidad estadística de estos modelos (Tabla G3), las ecuaciones 36 y 37 satisfacen los criterios estadísticos para la aleatorización en  $Y$  y validación externa, con lo cual se pueden considerar no aleatorias y predictivas. Sin embargo, no poseen buena robustez ni buena bondad del ajuste, lo que limita su aplicabilidad con fines regulatorios.

Comparativamente, la bondad del ajuste y robustez de ambas ecuaciones es similar, la predictividad es mayor para la ecuación 37, pero  $R_{adj}^2$  y FIT son mayores para la ecuación 36, con valores de CIA similares en ambas ecuaciones. De esta manera, ninguna de las ecuaciones es más adecuada que la otra con propósitos regulatorios.

Finalmente, la ecuación 37 posee menor estabilidad debido a valores de FIV con mayor magnitud. En concreto, se tiene que  $FIV(CIC) \sim FIV(\log f_{arcilla}) > FIV(\log f_{limo}) > otros FIV$ . Esto podría explicarse por la relación de proporcionalidad entre CIC y densidad de carga negativa del suelo [8], dado que la correlación entre CIC y  $\log f_{arcilla}$  es mayor que CIC y  $\log f_{limo}$ , y la arcilla presenta una mayor densidad de carga negativa que el limo debido a su menor tamaño de partícula.



## **5.9. Transferibilidad y criterios OCDE en la generación de modelos (Q)SAR para la predicción de $\log K_f$ y n de herbicidas en SDCV**

La transferibilidad y criterios OCDE fueron asegurados en este estudio mediante los siguientes aspectos:

Se describió el método analítico para la obtención de los datos a usar (Anexo D y sección 5.1), el tratamiento de datos (secciones 5.6 y 5.7) y la generación del modelo (secciones 5.4 y 5.8). De este modo, se obtuvo una reproducibilidad teórico-experimental.

Además, los descriptores propuestos fueron fundamentados y explicados en detalle (Anexos E y F), los métodos estadísticos fueron definidos y expresados de manera comprensible para facilitar la comparación entre modelos (secciones 4.3 y 5.8).

Los sistemas herbicida-SDCV se contextualizaron mediante: (i) comparación con modelos (Q)SAR para la adsorción de contaminantes orgánicos en SCP (sección 5.2) y (ii) definición y tratamiento de los factores que afectan a la adsorción de herbicidas en SDCV (Tabla A1 y sección 5.4); apoyando la interpretación de tendencias mediante el análisis no solo de los herbicidas predichos de manera confiable, sino que también de los valores problemáticos, con la idea de comprender las limitaciones del modelo y su sentido físico y químico.

Por último, todos los cálculos y metodologías fueron realizados con software de uso libre, facilitando la accesibilidad de la comunidad científica al momento de poner a prueba los aspectos mencionados previamente y utilizar los modelos (Q)SAR desarrollados en este estudio para comparar o apoyar sus trabajos.

## **5.10. Dominio de aplicabilidad e interpretación mecanística de los modelos (Q)SAR creados para predecir adsorción de herbicidas en SDCV**

### **5.10.1. Detección de valores atípicos y puntos de influencia**

El ajuste de los modelos (Q)SAR para la predicción de  $\log K_f$  y  $n$  de herbicidas en SDCV obtenidos en este estudio se muestra en las Figuras 10, 11 y 12.

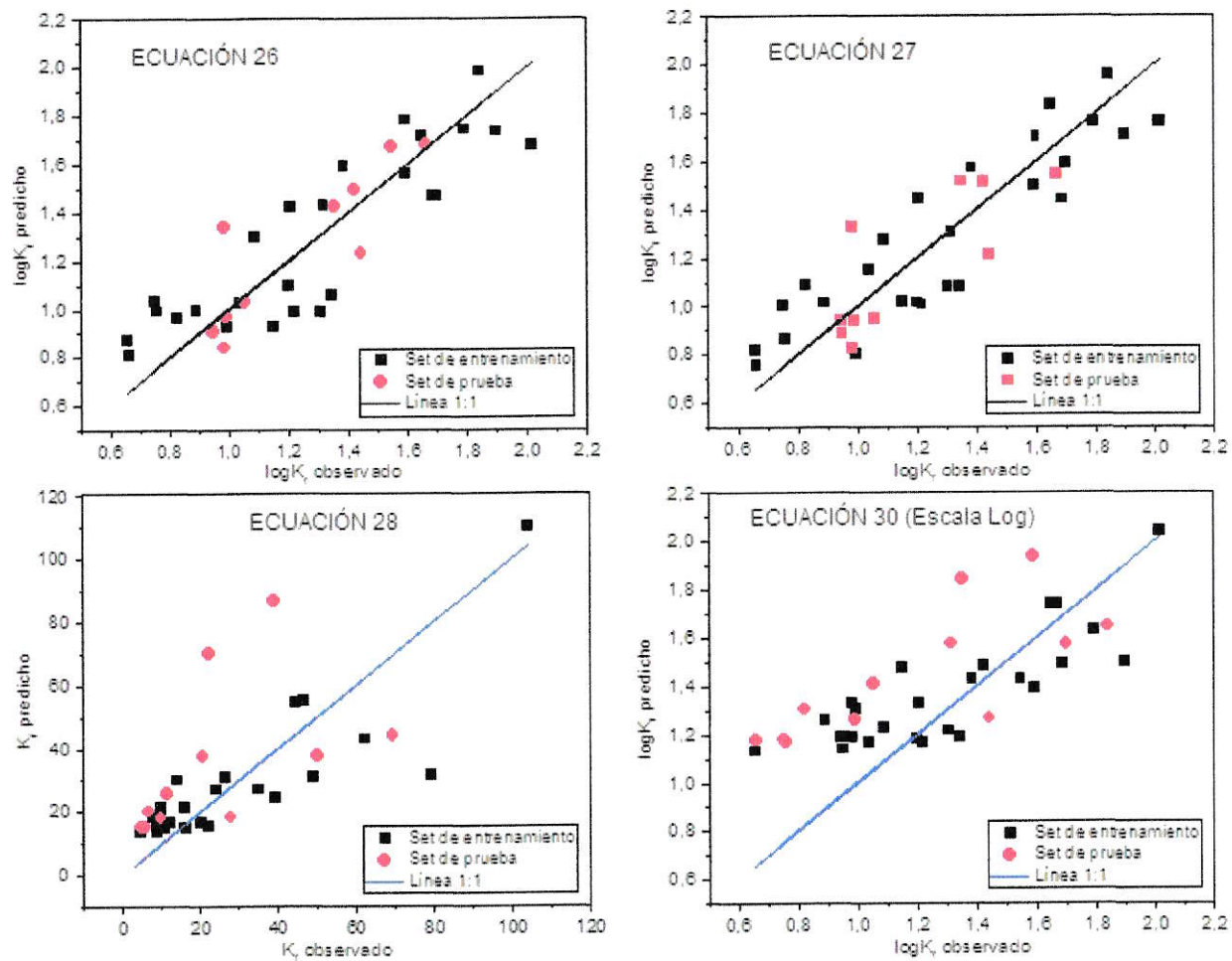
Los modelos (Q)SAR representados por las ecuaciones 28 y 34 presentaron valores atípicos en los sets de entrenamiento y de prueba. Al convertir dichas ecuaciones a escala logarítmica (ecuaciones 30 y 35), solo la ecuación 35 presentó un valor atípico en el set de entrenamiento, correspondiente a GFS en suelo RAL.

Se detectaron puntos de influencia en las ecuaciones 28, 34, 36 y 37, consideradas altamente dependientes del conjunto de datos usados para su elaboración. Sin embargo, lo anterior es corregido para las ecuaciones 28 y 34 tras el cambio a escala logarítmica (ecuaciones 30 y 35, respectivamente).

### **5.10.2. Interpretación mecanística de los modelos (Q)SAR generados para la adsorción de herbicidas no ionizables en SDCV**

La interpretación mecanística de las ecuaciones 26, 27 y 28 se realiza a través del análisis de los descriptores y coeficientes presentes en las ecuaciones mencionadas.

Los coeficientes negativos relacionados con los descriptores  $\log f_{\text{arcilla}}$  y  $\log f_{\text{Fe}_{\text{DCB}}}$  en las ecuaciones 26 y 27 indican que la adsorción de herbicidas no ionizables aumenta en SDCV que presenten un menor contenido de arcilla y  $\text{Fe}_{\text{DCB}}$ , como ocurre en la mayoría de valores de  $\log K_f$  en COLL y MET, suelos de tipo Ultisol.



**Figura 10.** Ajuste de las ecuaciones 26, 27, 28 y 30 para adsorción de herbicidas no ionizables (intervalo de aplicación  $\log K_f$ : 0,66 - 2,02).

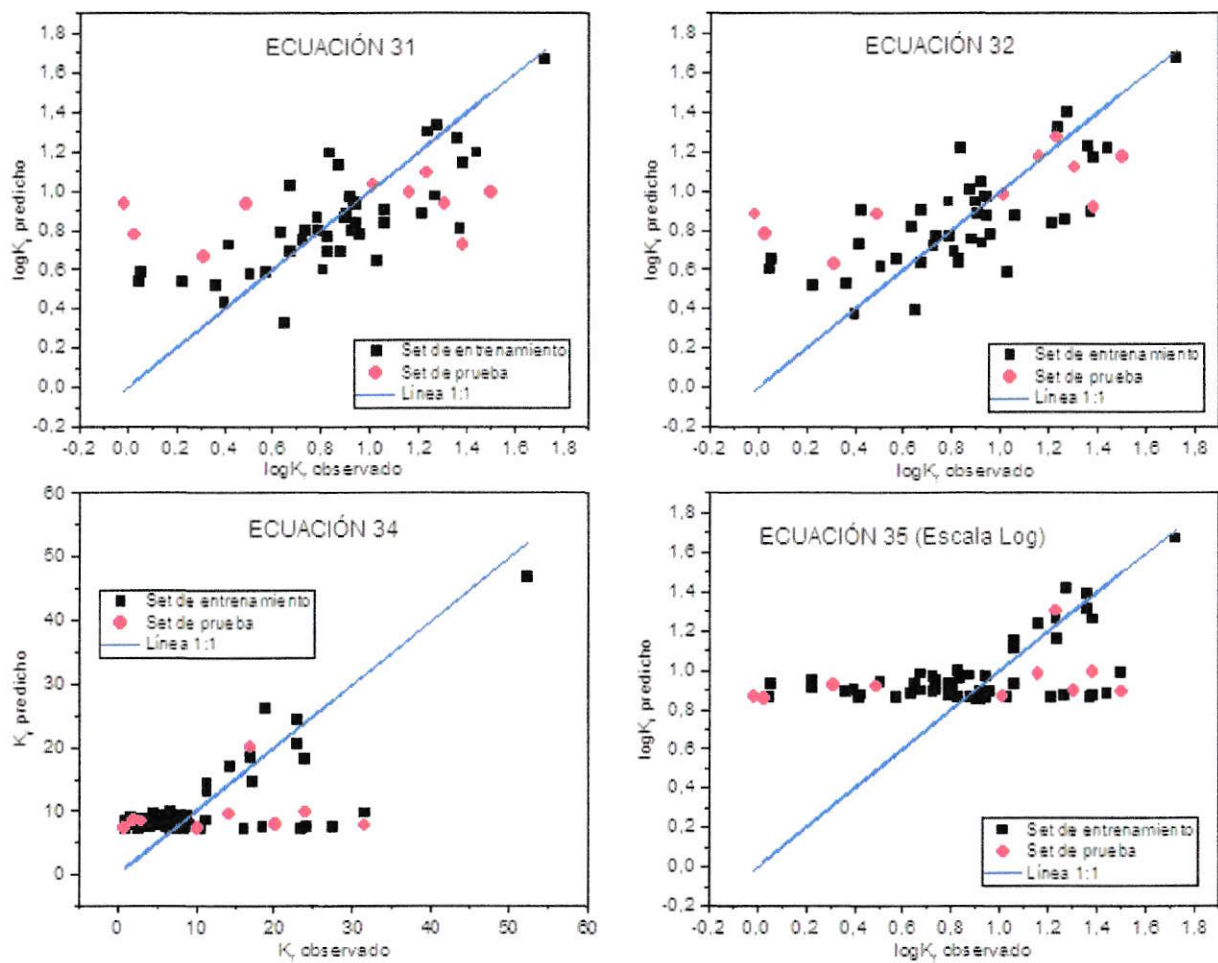
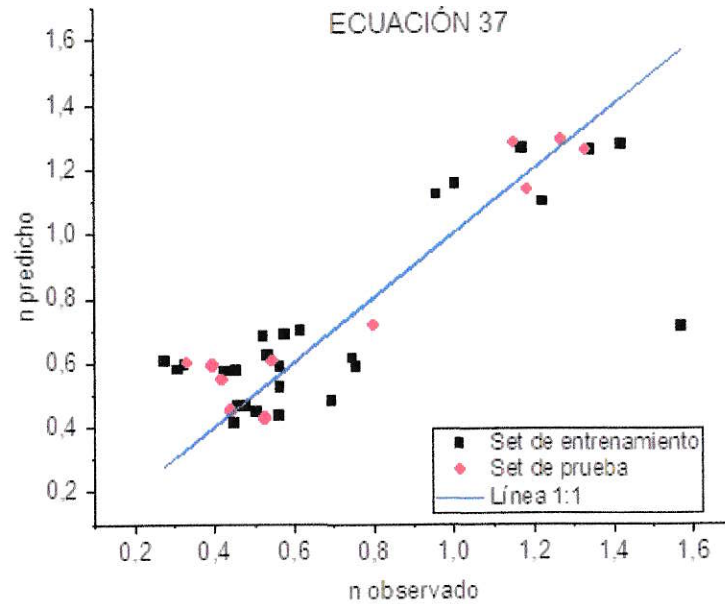
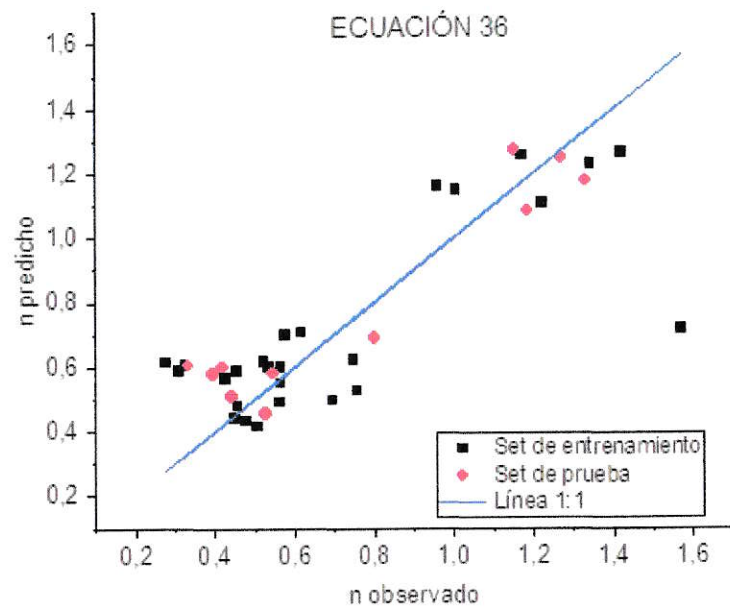


Figura 11. Ajuste de las ecuaciones 31, 32, 34 y 35 para adsorción de herbicidas ácidos (intervalo de aplicación  $\log K_f$ : 0,05 - 1,72).



**Figura 12.** Ajuste de las ecuaciones 36 y 37 para n en herbicidas no ionizables (intervalo de aplicación n: 0,28 - 1,57).

Las tendencias antes mencionadas se deberían al bloqueo de sitios de adsorción en la MO producto de la interacción con arcillas, y a la baja adsorción en óxidos de Fe libre presentes en SDCV (Tabla F1). De esta manera, la interpretación de las ecuaciones 26 y 27 se corresponde con los supuestos empleados en la generación de la ecuación 28 (sección 5.4 y Anexo F).

Matemáticamente, la adsorción predicha por las ecuaciones 26 y 27 tiende a infinito si  $\%arcilla = \%Fe_{DCB} = 0$ , lo que carece de sentido físico. Con base en la Tabla C2 es posible restringir las condiciones de aplicación de estas ecuaciones, validando su aplicabilidad para SDCV con contenidos de arcilla entre 19,4 y 45,7% y contenidos de  $Fe_{DCB}$  entre 0,6 y 7,1%.

Adicionalmente, la ecuación 27 incluye como descriptor  $\log f_{FePYRO}$ , que correlaciona positivamente con  $\log K_f$ , tendencia que podría relacionarse a la complejación de estos óxidos de Fe con la MO.

Por otra parte, ambas ecuaciones presentan el descriptor molecular  $Weta2.unity$  con coeficiente negativo. Dicho descriptor se relaciona inversamente con la densidad de átomos de una determinada molécula en el espacio 3-D [78], relacionada a su vez a la superficie de interacción del herbicida con la superficie de SDCV.

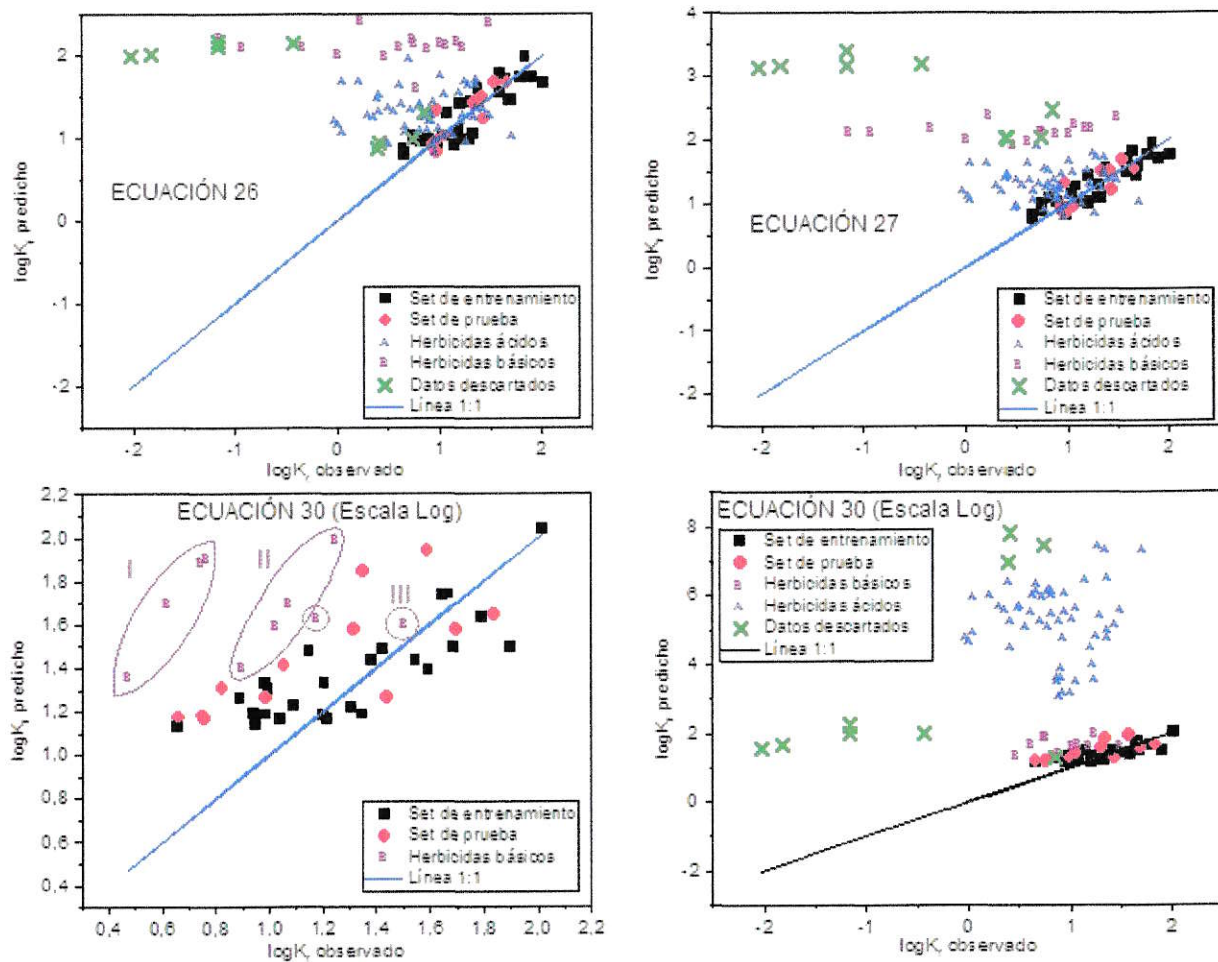
Con base en lo anterior, a mayor valor de  $Weta2.unity$  habrá menor interacción del herbicida con la superficie de SDCV, lo cual explica el coeficiente negativo asociado. Sin embargo, es difícil asociar una interpretación mecanística al descriptor  $Weta2.unity$ , puesto que no considera el tipo de interacción herbicida-SDCV.

La ecuación 28 presentó menor calidad estadística que las ecuaciones 26 y 27 (Tabla G1). Sin embargo, su interpretación mecanística es sencilla (p.ej. el intervalo de aplicación en términos de  $K_f$  siempre estará determinado por el intercepto, pues todos los coeficientes deben ser necesariamente positivos) y su calidad del ajuste (Figura 10) puede ser asociada directamente a los supuestos usados en su confección.

Adicionalmente, la ecuación 28 facilita la incorporación del ciclo PDCA (Figura 4) en la generación de modelos (Q)SAR para predecir la adsorción de herbicidas no ionizables en SDCV, permitiendo mejoras en la predictividad y aplicabilidad mediante:

- (i) Cuantificación del intercepto: Cuando la adsorción es nula en todas las componentes del suelo  $K_f = 0$ , por lo tanto se esperaría que el intercepto de la ecuación 28 sea cercano a 0. Esto no ocurre, reflejando deficiencias en el modelo (Q)SAR asociado a dicha ecuación.
- (ii) Análisis de la versión logarítmica del modelo (ecuación 30): Los parámetros A (intercepto) y B (pendiente) se alejan de sus valores ideales. Esto implica deficiencias en la ecuación propuesta y permite cuantificar su mejoría con la consideración de nuevos supuestos y/o estudios que, una vez introducidos en los descriptores y generando una nueva regresión, acerquen A y B a sus valores ideales.

Las desviaciones asociadas al ajuste de herbicidas básicos y ácidos en las ecuaciones 26, 27 y 30 generadas para predecir  $\log K_f$  de herbicidas no ionizables en SDCV son presentadas en la Figura 13.



**Figura 13.** Ajuste de las ecuaciones 26, 27 y 30 para la adsorción de herbicidas ácidos y básicos.



En todos los casos, la adsorción de bases y ácidos en SDCV es sobreestimada, sin embargo, la predicción de  $\log K_f$  de herbicidas ácidos es mayor para las ecuaciones 26 y 27. Estas ecuaciones incluyen óxidos de Fe entre sus descriptores, a diferencia de los modelos (Q)SAR para SCP, en los cuales la adsorción de estos herbicidas fue subestimada (Figuras 5, 6 y 7) por considerarse asociada únicamente a la MO.

Por otro lado, la sobreestimación de la adsorción de herbicidas ácidos en SDCV utilizando la ecuación 30 puede ser vinculada a la falta de corrección por repulsión electrostática de los aniones con las superficies de los minerales y la MO.

La adsorción de compuestos básicos en las ecuaciones 26 y 27 permanece constante, es decir, ninguno de estos modelos es sensible al cambio de herbicida ni de suelo.

Por otro lado, en la ecuación 30 en escala logarítmica se observan patrones definidos para la desviación de los herbicidas básicos (Figura 13, elipses morados).

Las bases más alejadas de la línea 1:1 (zona I) poseen un error relativo de  $(165 \pm 18)\%$ , y corresponden a suelos con menor contenido total de óxidos de hierro (2,25% en promedio), y una relación  $\frac{\text{Fe PYRO}}{\text{Fe OX}} \sim 0,3$ . Por otro lado, las bases de la zona II poseen menor error relativo ( $53 \pm 9\%$ ), mayor contenido de óxidos de hierro (3,86% en promedio) y una relación  $\frac{\text{Fe PYRO}}{\text{Fe OX}} \sim 0,7$ , mayor al caso anterior. De todos los compuestos en la zona II, el más cercano a la línea 1:1 (encerrado en un círculo) posee una relación  $\frac{\text{Fe PYRO}}{\text{Fe OX}} = 1,3$ , mayor a los casos anteriores.

Por último, la base mejor predicha por la ecuación 30, con un error relativo de 6%, presenta un contenido de óxidos de hierro bajo (1,6%), pero una relación  $\frac{\text{Fe PYRO}}{\text{Fe OX}} =$

1,7, mayor a los casos anteriores. Además, posee bajo contenido de arcilla y alto contenido de CO en comparación al resto de suelos.

De este modo, se observa una relación entre tipos de óxidos de Fe que afectaría solo a los herbicidas básicos a pesar de encontrarse en su forma neutra, con lo que no solo importaría la cantidad de óxidos de hierro, limitando con ello la aplicabilidad del modelo propuesto. Esto podría estar asociado a la afinidad de los diferentes óxidos por los herbicidas básicos, ya sea por formación de complejos, no considerada en este modelo, o por diferencias específicas en la densidad de carga superficial local.

Por analogía con las interacciones MO-mineral conocidas, otra opción para explicar las desviaciones sería una interacción Fe-mineral y Fe-MO, y la formación de sitios locales de adsorción con diferentes  $\phi_{\text{neu}}$  y  $\phi_+$  por la naturaleza heterogénea del suelo. Sin embargo, la información disponible no permite determinar si  $\phi_+ > 0$  en sitios locales, ni incorporar de manera específica las interacciones entre CO, minerales, óxidos de hierro y su impacto en sus correspondientes descriptores.

Con base en lo anterior, para predecir la adsorción de herbicidas básicos en SDCV no basta con incorporar el efecto de los óxidos de hierro, sino que deben ser consideradas interacciones específicas entre tipos de óxidos así como la relación de estos con otros componentes de los SDCV.

### **5.10.3. Interpretación mecanística de los modelos (Q)SAR generados para la adsorción de herbicidas ácidos en SDCV**

Los herbicidas ácidos empleados en la generación de las ecuaciones 31 y 32 presentan cadenas heteroatómicas (N, O, S) (Tabla C1). En ellas ocurre la ionización

de los herbicidas y pueden considerarse centros reactivos de cada molécula. Entonces, la correlación positiva entre  $\log K_f$  y el descriptor molecular nAtomLC podría deberse a la inclusión de los centros reactivos en los átomos en la cadena más larga de cada molécula.

La ecuación 31 indica que la adsorción de herbicidas ácidos aumenta en suelos con menor contenido de arcilla, menor CIC y mayor contenido de  $Fe_{OX}$ . Las interacciones electrostáticas de los herbicidas en su forma aniónica con la superficie del suelo podrían explicar la relación con el contenido de arcilla (negativa, interacciones repulsivas) y  $Fe_{OX}$  (positiva, interacciones atractivas). Por otro lado, la CIC aumenta con la carga superficial negativa y, por lo tanto, se relaciona de manera indirecta con las interacciones repulsivas entre el suelo y los aniones.

La ecuación 32 muestra una adsorción favorecida a menor contenido de CO, menor CIC y mayor contenido de  $Fe_{DCB}$ . Esto también tendría explicación en las interacciones electrostáticas, que serían mayores que la afinidad por la MO, lo que se sustenta en la baja lipofilia de estos herbicidas (Tabla C1). Por otro lado, la inclusión del contenido de  $Fe_{DCB}$  podría indicar que este tipo de óxido adquiere relevancia en la adsorción de herbicidas aniónicos y no debería ser descartado del término  $K_{f(Fe+)}$  en la ecuación 34.

Los supuestos de la ecuación 33 se relacionan con las tendencias obtenidas en las ecuaciones 31 y 32, es decir, adsorción favorecida por los óxidos de Fe, baja en la fracción orgánica y desfavorecida por la densidad de carga superficial negativa y CIC.

La forma logarítmica de la ecuación 34 (ecuación 35) posee menor validez estadística que la versión no logarítmica, e incluso puede ser considerada una correlación aleatoria ( $cR_p^2 < 0,5$ ), siendo altamente insensible a la variación de  $\log K_f$ , sobre todo para valores de  $\log K_f < 0,9$ . Luego, el análisis de los parámetros A y B carece de sentido.

Una fuente de incerteza que surge del análisis de las ecuaciones 31, 32 y 34 es la relación entre CIC y composición del suelo. La CIC es proporcional a la densidad de carga negativa del suelo [8], lo que podría explicar su correlación negativa con la adsorción. Sin embargo, la presencia de  $\log f_{CO}$  y  $\log f_{arcilla}$  podría representar información redundante en estas ecuaciones, como se observa en el análisis del parámetro estadístico FIV (Tabla G2).

Para aclarar lo anterior se utiliza la Tabla F1, donde se presenta el efecto de la CIC, contenido de arcilla y/o MO en la adsorción. Sin embargo, debido a la imposibilidad de controlar todas las variables presentes en el suelo (matriz compleja), la cuantificación del efecto de la CIC, contenido de arcilla o MO no se hace manteniendo todas las demás propiedades constantes. Luego, el impacto de cada propiedad en el proceso de adsorción es inevitablemente condicionado por las demás propiedades, aumentando su incerteza.

Para mejorar la representación de la repulsión entre aniones y sitios de carga negativa del suelo, Hyun y Lee (2004) proponen usar la capacidad de intercambio aniónica (CIA) normalizada a la CIC, es decir, CIA/CIC [79]. Según estos autores, la relación propuesta refleja el potencial encuentro entre sitios de carga positiva y aniones. Por

lo tanto, la inclusión de la CIA en la ecuación 34 podría mejorar su calidad estadística por incorporación de efectos repulsivos y oclusión de sitios en los óxidos de hierro.

Finalmente, la no consideración de mecanismos de complejación o interacciones herbicida-SDCV podría explicar las desviaciones observadas en el ajuste de la ecuación 34 (Figura 11). En este sentido, Cáceres-Jensen y col. (2010) indican que 2,4 – D y MSM podrían ser adsorbidos mediante formación de puente-Ca en la superficie de los minerales [11], interacción que no fue considerada en la formulación de la ecuación 34 y que podría existir en herbicidas con grupos -NH-.

#### **5.10.4. Interpretación mecanística de los modelos (Q)SAR generados para predecir n de herbicidas no ionizables en SDCV**

La ecuación 36 posee dos descriptores de SDCV, donde la disminución de n con el contenido de arena es la única relación clara, debido al error asociado al coeficiente de  $\log f_{\text{arcilla}}$ . Por otro lado, la ecuación 37 contempla tres descriptores de SDCV, y muestra un aumento de n con el contenido de arcilla.

De la unión de ambas ecuaciones, se tiene que a menor tamaño de partícula del suelo existiría un mayor valor de n, cuya validez está limitada a contenido de arena entre 6,2 y 47,3%, y contenido de arcilla entre 19,4 y 45,7%. Esto se puede asociar a que la pérdida de linealidad en el intervalo de concentraciones bajo el cual se realizaron las isothermas de adsorción es menos pronunciada en SDCV con mayor superficie de adsorción disponible. Por lo tanto, la aplicabilidad de estos modelos estaría limitada al cumplimiento de las condiciones experimentales, que debe ser evaluado en función de las dosis de aplicación de cada herbicida.

El cálculo del descriptor molecular SC – 5 considera el esqueleto de la molécula 2-D, en el cual se toman todas las combinaciones de 5 átomos vecinos. Cada átomo (excepto H) actúa como un vértice y se le asocia un valor equivalente al número de enlaces que forma dentro del conjunto de 5 miembros. Luego, se obtiene un valor basado en la relación entre los valores previamente generados a partir de todas las combinaciones consideradas [77].

Con base en la definición de SC – 5, se observa que este descriptor no distingue el tipo de átomo presente en la molécula. Sin embargo, es sensible al tipo de enlaces, así como al número de combinaciones de 5 átomos vecinos, por lo que puede asociarse a la complejidad de la molécula o a la distribución de sus átomos. Esto sugiere una relación inespecífica entre n y los herbicidas no ionizables.

## **5.11. Aplicabilidad de los modelos (Q)SAR generados para predecir la adsorción de herbicidas en SDCV**

### **5.11.1. Aplicabilidad en condiciones ambientales**

Las principales ventajas de los modelos (Q)SAR generados en este trabajo respecto a modelos ya existentes son su sensibilidad espacial, temporal e inter-suelo y su contextualización a SDCV con la inclusión de los óxidos de Fe como descriptores. Sin embargo, la validez de las tendencias obtenidas debe ser puesta a prueba para condiciones reales en que la metodología empleada para determinar el *endpoint* podría no ser representativa, y el proceso de adsorción podría ser acompañado de procesos tales como degradación, transporte, biodegradación, entre otros.

El uso de valores de *endpoint* determinados en *batch* para generar los modelos (Q)SAR de este trabajo podría implicar una sobreestimación de  $K_f$ , como ya fue mencionado previamente, producto de la inexistencia de pérdidas por lixiviación, volatilización, entre otros, debido al uso de sistemas cerrados. Por otro lado, la agitación constante en la metodología *batch* podría generar una subestimación del tiempo de contacto requerido para alcanzar el equilibrio químico respecto a condiciones ambientales reales.

Otro problema de la metodología en *batch* corresponde a la diferencia de la relación suelo/solución con las condiciones reales. Por ejemplo, para estimar la dosis de aplicación en suelos reales y convertirla a dosis equivalente en *batch* con la finalidad de representar mejor las condiciones reales, existen dos alternativas:

- (i) Tomar la dosis de aplicación por masa de suelo ( $g_{ia} \text{ ha}^{-1}$ ) para obtener la masa de herbicida que se requiere adicionar a la cantidad de suelo del sistema *batch* (2 g de suelo seco al aire), asegurando la misma relación inicial herbicida/suelo. Luego, la masa de herbicida es disuelta en la solución e incorporada al suelo para los estudios de cinética y adsorción.
- (ii) Tomar la dosis de aplicación por volumen de solución ( $g_{ia} \text{ L}^{-1}$ ) y usar esta concentración para la solución del sistema *batch*, manteniendo así la misma relación inicial herbicida/solución. Luego, esta concentración inicial de herbicida en solución se pone en contacto con los 2 g de suelo para los estudios de cinética y adsorción.

En ambos casos, la relación suelo/solución en condiciones reales diferirá considerablemente de la usada en *batch*. En el caso real, la relación suelo/solución

será mucho mayor a 1, mientras que en *batch* esta relación es de 0,2 (10 mL de solución por cada 2 g de suelo, es decir, suelo inundado).

Lo anterior afectará la tendencia de la isoterma, ya que en condiciones reales se tendrá una alta concentración inicial de herbicida en solución que podrá ser potencialmente adsorbida por una gran masa de suelo, pero en condiciones *batch* se tendrá un escenario opuesto.

De este modo, al generar una relación inicial herbicida/suelo equivalente, se tendrá una concentración muy baja en solución, lo que podría desplazar el comportamiento de la isoterma hacia la linealidad, mientras que generando una relación inicial herbicida/solución equivalente, habrá una mayor probabilidad de saturar el suelo, desplazando el comportamiento de la isoterma hacia *Freundlich* con  $n < 1$ .

En este sentido, la correspondencia de la metodología en *batch* y las condiciones reales dependerá también de la humedad del suelo, que en relación suelo/solución mayor a 1 podría afectar significativamente a la dilución del herbicida, y de la capacidad de retención de agua del suelo, que dará cuenta del impacto del transporte, lo que afecta al tiempo de contacto del herbicida con el suelo.

Otro proceso relevante es la degradación, que disminuye la concentración del herbicida en solución y la cantidad adsorbida, actuando como depurador de los SDCV y compitiendo con la adsorción [19]. Si es significativa, el impacto ambiental predicho por modelos (Q)SAR basados en adsorción podría ser sobreestimado, o bien, transferido a los productos de degradación.



Adicionalmente, la movilidad del herbicida afectará a la superficie de adsorción disponible y con ello la saturación del suelo y la probabilidad de encuentro entre el herbicida y la microbiota, relacionándose a la biodegradación.

Para un mismo grado de humedad, una mayor capacidad de retención de agua implicará una menor movilidad, lo que puede traducirse en menor velocidad de infiltración. Si la velocidad de infiltración es mayor a la velocidad de adsorción, aumentará la lixiviación, contaminando cuerpos de agua subterránea. Por lo tanto, la relación suelo/solución es importante para asegurar la representatividad del proceso de adsorción en la dinámica herbicida-SDCV y, con ello, la representatividad de  $\log K_f$  como *endpoint*.

Por lo tanto, los modelos (Q)SAR de adsorción de herbicidas en SDCV deben ser acotados a las propiedades de cada suelo para permitir su representatividad, y a las implicancias de otros procesos en la salud humana para su potencial aplicación con propósitos regulatorios. Luego, los valores entregados por modelos (Q)SAR usando la metodología *batch* deben ser entendidos como la máxima adsorción posible, generando una variedad de escenarios probables para la toma de decisiones racionales frente al potencial impacto ambiental del uso de herbicidas en SDCV.

### **5.11.2. Aplicabilidad en el contexto chileno**

La falta de información a nivel país suele implicar el uso de estrategias y modelos extranjeros con la idea de agilizar la toma de decisiones en materia medioambiental. Sin embargo, este trabajo deja en evidencia el gran error asociado a este tipo de actos producto de la descontextualización.

Los tres modelos (Q)SAR realizados a partir de SCP muestran desviaciones importantes al momento de predecir valores de *endpoint* obtenidos de SDCV chilenos (Figuras 5, 6 y 7).

De esta manera, la generación de modelos (Q)SAR para la adsorción de herbicidas en SDCV permite favorecer el cumplimiento del artículo 39 de la Ley 19.300 modificada por la Ley 20.417, que indica “la ley velará porque el uso del suelo se haga en forma racional, a fin de evitar su pérdida y degradación” [80].

Luego, el aseguramiento de la conservación de los recursos naturales es deber del Ministerio del Medio Ambiente y organismos asociados al aprovechamiento de los recursos naturales en un área determinada (Artículo 42), y se asume que un proyecto o actividad cumple con la legislación cuando su Declaración o Estudio de Impacto Ambiental es aprobado.

Por lo tanto, en Chile conviene incorporar modelos (Q)SAR como herramienta legal en la creación de líneas de base para los Estudios de Impacto Ambiental y para la generación de información nacional, así como en la verificación del cumplimiento de la norma por parte de algún proyecto o actividad.

La sugerencia anterior queda en manos del Ministerio del Medio Ambiente, quien debe proponer políticas, normas, planes y programas en materia de residuos y suelos contaminados (Artículo 70, letra g).

Por último, para facilitar la aplicabilidad de los modelos y acelerar la toma de decisiones racional, una posibilidad es abandonar la visión determinista de los descriptores, es decir, la suposición de que estos son conocidos con certeza, y optar

por intervalos de valores para cada descriptor. De este modo, en vez de tomar diferentes muestras de suelo, caracterizarlas y promediar los valores obtenidos, podría generarse un intervalo de valores en función de la variabilidad espacial. Así, un descriptor relacionado por ejemplo al contenido de CO no presentaría un único valor, sino un intervalo que reflejaría la variabilidad del suelo y, por lo tanto, generaría potenciales valores de  $\log K_f$  locales para dicho suelo, implicando mejores estrategias para el cultivo y uso de herbicidas en países como Chile, que no cuentan con una cultura desarrollada en materia de complementar sus procesos agroindustriales con rotación de cultivos y control integrado de plagas.

### **5.12. Discusión metodológica**

La metodología propuesta para la creación de modelos (Q)SAR en SDCV permite:

- (i) Asegurar un *endpoint* representativo del set de datos empleado: Las tendencias empíricas obtenidas para el proceso de adsorción en SDCV evidencian la inaplicabilidad de  $K_d$  (isotermas no lineales), la incerteza frente a los supuestos requeridos para el uso de  $K_{oc}$  (sección 5.5.3), y la necesidad de incluir  $n$  como *endpoint* debido a su variabilidad (Anexo B).
- (ii) Conocer el origen y metodología de determinación de los datos empleados en la creación del modelo (Q)SAR, facilitando la propuesta de descriptores, la interpretación de los modelos generados y la detección de limitaciones.
- (iii) Incorporar SDCV en la variabilidad del *endpoint*: Los modelos (Q)SAR generados en este trabajo consideran el efecto de SDCV en el proceso de adsorción y, por lo tanto, son sensibles a diferentes suelos, mejorando su aplicación con propósitos regulatorios.

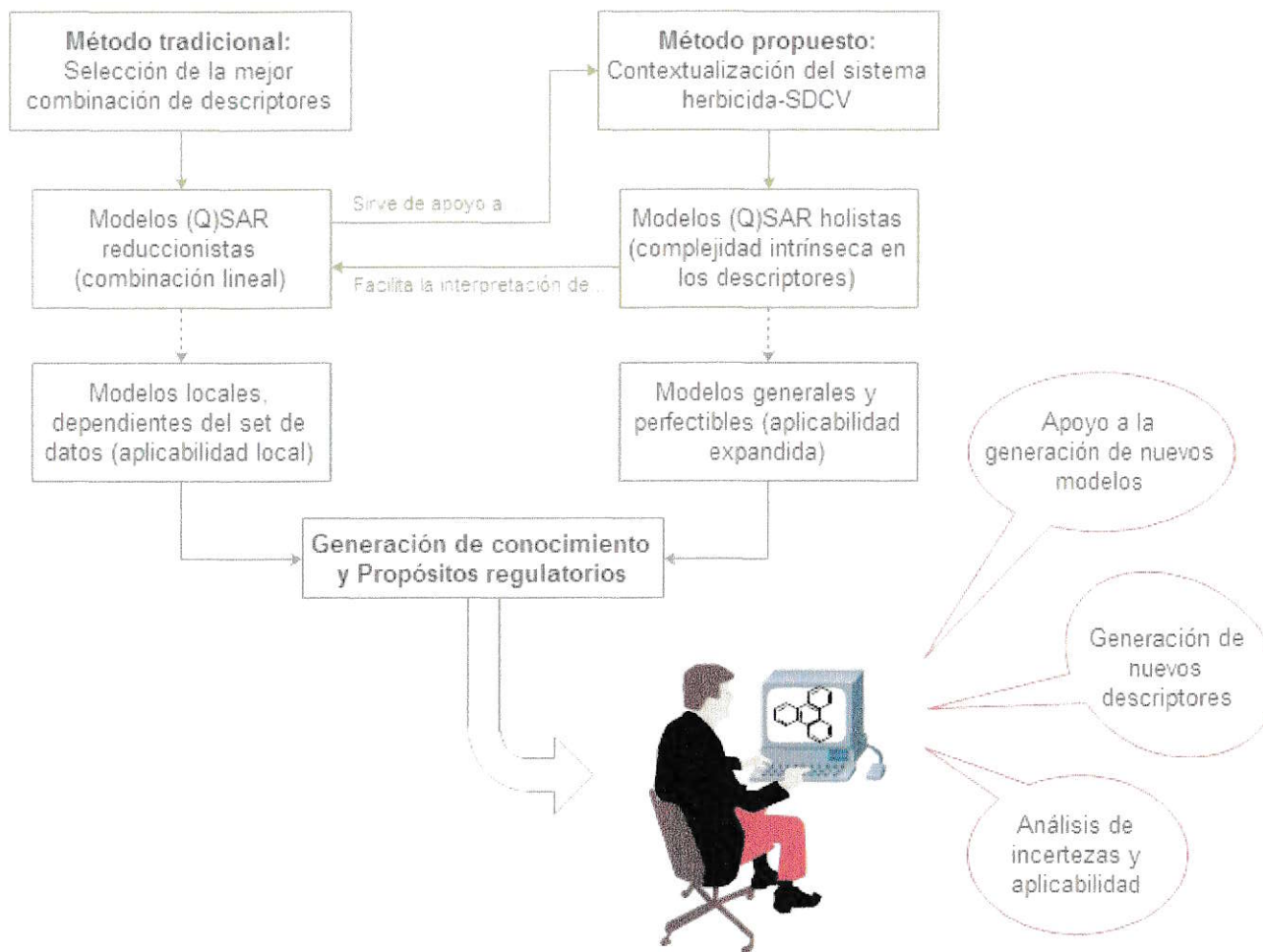
- (iv) Definir nuevos descriptores para SDCV y su interacción con los herbicidas:  
Creación de modelos (Q)SAR con descriptores cada vez más representativos si estos se condicionan a una base teórico/conceptual en constante desarrollo (p.ej. estudios científicos y líneas de base), incluyendo a todos los participantes del proceso de adsorción (SDCV, herbicidas, interacciones).
- (v) Analizar y comparar la calidad estadística de manera confiable: El uso de diferentes parámetros de validación y la inclusión de parámetros estadísticos para la comparación entre modelos, facilita el análisis entre modelos, la transferibilidad de los mismos y la aplicación del ciclo PDCA.
- (vi) Generar un dominio de aplicabilidad más complejo que la metodología tradicional: En este caso, el dominio de aplicabilidad no solo es función de los herbicidas usados, sino que también involucra SDCV y la metodología bajo la cual se cuantificaron los datos usados para crear el modelo (Q)SAR.
- (vii) Dar una interpretación mecanística más elaborada que la metodología tradicional: La inclusión de SDCV y dos tipos de modelos (Q)SAR permite ahondar en los descriptores seleccionados, sus errores, el sentido físico y químico de los modelos generados, y la correspondencia entre diferentes modelos debido a su análisis conjunto.
- (viii) Detectar aplicaciones y limitaciones de manera más directa y realista que la metodología tradicional: Se incluye la validación empírica del modelo (Q)SAR como condición necesaria a su aplicación con propósitos regulatorios, es decir, analizar su predictibilidad bajo escenarios reales en los que se consideren posibles variables no contempladas en la generación del set de datos o del modelo mismo (p. ej. lixiviación), lo que permite

dirigir el uso de los modelos (Q)SAR de manera cuidadosa, apoyando a los principios preventivo y precautorio del Derecho Ambiental.

Además de las mejoras mencionadas respecto a los modelos (Q)SAR ya existentes, la aplicación de la metodología propuesta (Figuras 3 y 4) permite complementar los modelos (Q)SAR generados, a través de su uso conjunto (Figura 14).

Por un lado, se tienen los modelos basados en la metodología tradicional (ecuaciones 26, 27, 31, 32, 36 y 37), creados a partir de la mejor combinación de descriptores, con una fuerte componente estadística y un enfoque reduccionista (supone que la relación entre descriptores es lineal). La ventaja de estos modelos (Q)SAR radica en su simpleza (p. ej. no requieren contextualización detallada), su buena correlación y la posibilidad de detectar relaciones no consideradas en estudios previos, o descartar relaciones poco significativas para un conjunto particular de datos. Sin embargo, su aplicabilidad está condicionada a los sets de entrenamiento y prueba utilizados, convirtiéndolos en modelos de uso específico.

Por otra parte, se incluyen modelos generados a partir de una base teórico/conceptual (ecuaciones 28 y 34), con una mayor contextualización a los sistemas herbicida-SDCV y un enfoque holista que privilegia el sentido físico y químico por sobre la validación estadística. Su ventaja se encuentra en la rápida detección de limitaciones y una interpretación clara, debido a su base teórico/conceptual, lo que permite la consideración de nuevas variables o reformulación de ciertos supuestos, dando utilidad a estos modelos incluso si no son estadísticamente validados. Además, el hecho de representar el proceso de adsorción en forma global implica que estos modelos (Q)SAR son de uso general e independiente del set de datos empleado.



**Figura 14.** Uso complementario de modelos (Q)SAR generados en este trabajo y su contexto en el ciclo PDCA.

El complemento a la Figura 4, correspondiente al tratamiento de cada modelo (Q)SAR individual, se consigue con la aplicabilidad conjunta de los modelos (Q)SAR generados en este trabajo, de la siguiente manera:

- (i) Cada modelo es en sí mismo un estudio científico disponible, mejorando la base teórico/conceptual de futuros modelos (Q)SAR.
- (ii) La interpretación de estos modelos y los descriptores propuestos y usados en los mismos facilitan la generación de nuevos descriptores para futuros modelos (Q)SAR, sobre todo si involucran moléculas o suelos similares.
- (iii) El análisis de todos los modelos (Q)SAR da pistas sobre las incertezas presentes y su origen, lo que permite evaluar su reducción y mejorar la aplicabilidad y representatividad de cada modelo.

De esta forma, todos los modelos (Q)SAR tienen una utilidad propia y conjunta, obteniéndose información tanto de los aciertos como de las fallas, compensando sus limitaciones e incertezas y facilitando la futura aplicación del ciclo PDCA, al dar pistas sobre diferentes aspectos mecanísticos.

## 6. CONCLUSIONES

1. El impacto de la carga variable presente en SDCV y la presencia de adsorbentes que difieren de la MO, que son relevantes para el proceso de adsorción, impiden aplicar los modelos (Q)SAR existentes para predecir adsorción de contaminantes orgánicos en SCP.
2. La representatividad del sistema herbicida-SDCV aumenta en los sets de datos cuando se consideran los valores de  $\log K_f$  de herbicidas de cada SDCV por separado, evitando la normalización al contenido de CO.
3. Se crearon modelos (Q)SAR lineales, estadísticamente confiables, para predecir la adsorción de herbicidas no ionizables en SDCV empleando  $\log K_f$  como *endpoint*, descriptores moleculares (Weta2.unity) y de SDCV ( $\log f_{\text{arcilla}}$ ,  $\log f_{\text{FeDCB}}$ ,  $\log f_{\text{FePYRO}}$ ). Estos modelos presentan una interpretación mecanística sencilla y aplicabilidad limitada en cuanto a propósitos regulatorios.
4. El uso conjunto de modelos (Q)SAR generados a partir de la selección de la mejor combinación de descriptores y modelos (Q)SAR basados en la interacción herbicida-SDCV permite considerar nuevos supuestos y/o estudios



y vislumbrar posibles descriptores para facilitar la interpretación mecanística, detectar falencias y perfeccionar modelos (Q)SAR.

5. Para la adecuada aplicación de los modelos (Q)SAR y la oportuna toma racional de decisiones, se propone: (i) usar diferentes valores de descriptores de naturaleza empírica para reflejar la variabilidad espacial de SDCV en los valores de  $\log K_f$  predichos, y (ii) medir reiteradamente los descriptores en el tiempo, dando cuenta de la variabilidad temporal del sistema herbicida-SDCV.
6. Para la aplicación de modelos (Q)SAR en Chile bajo condiciones ambientales locales, se sugiere considerar: (i) relevancia del proceso de adsorción y efecto de las condiciones reales, (ii) correspondencia entre condiciones reales y metodología empleada en la generación de datos para el modelo (Q)SAR, y (iii) participación del Ministerio del Medio Ambiente en el financiamiento, creación y puesta en práctica de modelos (Q)SAR con propósitos regulatorios.

## 7. BIBLIOGRAFÍA

- [1] M. Kah, C.D. Brown, Adsorption of ionisable pesticides in soils, in: *Reviews of Environmental Contamination and Toxicology*, Vol 188, Springer, New York, 2006, pp. 149-217.
- [2] C. Lu, Y. Wang, C. Yin, W. Guo, X. Hu, QSPR study on soil sorption coefficient for persistent organic pollutants, *Chemosphere*, 63 (2006) 1384-1391.
- [3] Servicio Agrícola y Ganadero, Informe de venta de plaguicidas de uso agrícola en Chile. Año 2008, in, SAG, Chile, 2008, pp. 118.
- [4] Servicio Agrícola y Ganadero, Informe de venta de plaguicidas de uso agrícola en Chile. Año 2012, in, SAG, Chile, 2012, pp. 113.
- [5] J.N. Fernandez-Cornejo, Richard; Osteen, Craig; Wechsler, Seth; Martin, Andrew; Vialou, Alex Pesticide Use in U.S. Agriculture: 21 Selected Crops, 1960-2008, in, U.S. Department of Agriculture, Economic Research Service, 2014, pp. 80.
- [6] A.K. Sarmah, J. Sabadie, Hydrolysis of Sulfonylurea Herbicides in Soils and Aqueous Solutions: a Review, *Journal of Agricultural and Food Chemistry*, 50 (2002) 6253-6265.
- [7] A. Delle Site, Factors Affecting Sorption of Organic Compounds in Natural Sorbent/Water Systems and Sorption Coefficients for Selected Pollutants. A Review, *Journal of Physical and Chemical Reference Data*, 30 (2001) 187-439.
- [8] T.R. Yu, *Chemistry of Variable Charge Soils*, Oxford University Press, Oxford, 1997.
- [9] S.N. Shoji, Masami; Dahlgren, Randy, *Volcanic ash soils: genesis, properties and utilization*, Elsevier Science Publishers, Elsevier, 1993.
- [10] M. Escudey, G. Galindo, J.E. Förster, M. Briceño, P. Diaz, A. Chang, Chemical forms of phosphorus of volcanic ash-derived soils in Chile, *Communications in Soil Science and Plant Analysis*, 32 (2001) 601-616.
- [11] L. Cáceres Jensen, R. Fuentes, M. Escudey, E. Fuentes, M.E. Baez, Metsulfuron-methyl Sorption/Desorption Behavior on Volcanic Ash-Derived Soils. Effect of Phosphate and pH, *Journal of Agricultural and Food Chemistry*, 58 (2010) 6864-6869.
- [12] United Nations, Rio Declaration on Environment and Development at Rio de Janeiro, UN Doc. A/CONF. 151/26 (Vol I) / 31 ILM 874, in, 1992.
- [13] OECD, *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models*, OECD Publishing, 2014.
- [14] K.N. Reddy, M.A. Locke, Prediction of soil sorption (Koc) of herbicides using semiempirical molecular-properties, *Weed Science*, 42 (1994) 453-461.

- [15] A. Sabljic, On the prediction of soil sorption coefficients of organic pollutants from molecular structure: application of molecular topology model, *Environmental Science & Technology*, 21 (1987) 358-366.
- [16] Y. Wang, J. Chen, X. Yang, F. Lyakurwa, X. Li, X. Qiao, In silico model for predicting soil organic carbon normalized sorption coefficient (K<sub>oc</sub>) of organic chemicals, *Chemosphere*, 119 (2015) 438-444.
- [17] R.R. dos Reis, S.C. Sampaio, E.B. de Melo, An alternative approach for the use of water solubility of nonionic pesticides in the modeling of the soil sorption coefficients, *Water Research*, 53 (2014) 191-199.
- [18] K. Roy, S. Kar, R.N. Das, Chapter 1 - Background of QSAR and Historical Developments, in: *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*, Academic Press, Boston, 2015, pp. 1-46.
- [19] C. Wenlin, L. Volker, S.K. Rai, C.K. William, Coupled Sorption and Degradation Kinetics and Non-First Order Behavior, in: *Non-First Order Degradation and Time-Dependent Sorption of Organic Chemicals in Soil*, American Chemical Society, 2014, pp. 5-37.
- [20] A. Tropsha, Best Practices for QSAR Model Development, Validation, and Exploitation, *Molecular Informatics*, 29 (2010) 476-488.
- [21] M.T.D. Cronin, J.S. Jaworska, J.D. Walker, M.H.I. Comber, C.D. Watts, A.P. Worth, Use of QSARs in international decision-making frameworks to predict ecological effects and environmental fate of chemical substances, *Environmental Health Perspectives*, 111 (2003) 1391-1401.
- [22] Thomson Reuters, Web of Science, in, <https://webofknowledge.com/>.
- [23] A. Sabljic, QSAR models for estimating properties of persistent organic pollutants required in evaluation of their environmental fate and risk, *Chemosphere*, 43 (2001) 363-375.
- [24] A.R. Katritzky, M. Kuanar, S. Slavov, C.D. Hall, M. Karelson, I. Kahn, D.A. Dobchev, Quantitative Correlation of Physical and Chemical Properties with Chemical Structure: Utility for Prediction, *Chemical Reviews*, 110 (2010) 5714-5789.
- [25] L. Cáceres-Jensen, J. Gan, M. Báez, R. Fuentes, M. Escudey, Adsorption of Glyphosate on Variable-Charge, Volcanic Ash-Derived Soils, *Journal of Environmental Quality*, 38 (2009) 1449-1457.
- [26] R.D. Clark, A perspective on the role of quantitative structure-activity and structure-property relationships in herbicide discovery, *Pest Manag. Sci.*, 68 (2012) 513-518.
- [27] P. Gramatica, S. Cassani, N. Chirico, QSARINS-Chem: Insubria Datasets and New QSAR/QSPR Models for Environmental Pollutants in QSARINS, *Journal of Computational Chemistry*, 35 (2014) 1036-1044.
- [28] L. Cáceres-Jensen, Modelación de la dinámica de glifosato y metil-metsulfurón en suelos derivados de cenizas volcánicas, in: *Facultad de Química y Biología., Universidad de Santiago de Chile, Santiago, 2010.*
- [29] Z. Gerstl, Estimation of organic chemical sorption by soils, *Journal of Contaminant Hydrology*, 6 (1990) 357-375.
- [30] S. Bintein, J. Devillers, QSAR for organic chemical sorption in soils and sediments, *Chemosphere*, 28 (1994) 1171-1188.

- [31] G.G. Briggs, Theoretical and experimental relationships between soil adsorption, octanol-water partition coefficients, water solubilities, bioconcentration factors, and the parachor, *Journal of Agricultural and Food Chemistry*, 29 (1981) 1050-1059.
- [32] W. Meylan, P.H. Howard, R.S. Boethling, Molecular topology/fragment contribution method for predicting soil sorption coefficients, *Environmental Science & Technology*, 26 (1992) 1560-1567.
- [33] S. Tao, X. Lu, Estimation of organic carbon normalized sorption coefficient (K<sub>oc</sub>) for soils by topological indices and polarity factors, *Chemosphere*, 39 (1999) 2019-2034.
- [34] P. Gramatica, M. Corradi, V. Consonni, Modelling and prediction of soil sorption coefficients of non-ionic organic pesticides by molecular descriptors, *Chemosphere*, 41 (2000) 763-777.
- [35] P.L. Andersson, U. Maran, D. Fara, M. Karelson, J.L.M. Hermens, General and Class Specific Models for Prediction of Soil Sorption Using Various Physicochemical Descriptors, *Journal of Chemical Information and Computer Sciences*, 42 (2002) 1450-1459.
- [36] G. Schüürmann, R.-U. Ebert, R. Kühne, Prediction of the Sorption of Organic Compounds into Soil Organic Matter from Molecular Structure, *Environmental Science & Technology*, 40 (2006) 7005-7011.
- [37] P. Gramatica, E. Giani, E. Papa, Statistical external validation and consensus modeling: A QSPR case study for K<sub>oc</sub> prediction, *Journal of Molecular Graphics and Modelling*, 25 (2007) 755-766.
- [38] P.R. Duchowicz, M. Perez Gonzalez, A. Morales Helguera, M.N. Dias Soeiro Cordeiro, E.A. Castro, Application of the replacement method as novel variable selection in QSPR. 2. Soil sorption coefficients, *Chemometrics and Intelligent Laboratory Systems*, 88 (2007) 197-203.
- [39] Y. Wen, L.M. Su, W.C. Qin, L. Fu, J. He, Y.H. Zhao, Linear and non-linear relationships between soil sorption and hydrophobicity: Model, validation and influencing factors, *Chemosphere*, 86 (2012) 634-640.
- [40] Y.H. Shao, J.N. Liu, M.X. Wang, L.L. Shi, X.J. Yao, P. Gramatica, Integrated QSPR models to predict the soil sorption coefficient for a large diverse set of compounds by using different modeling methods, *Atmospheric Environment*, 88 (2014) 212-218.
- [41] A. Franco, S. Trapp, Estimation of the soil-water partition coefficient normalized to organic carbon for ionizable organic chemicals, *Environmental Toxicology and Chemistry*, 27 (2008) 1995-2004.
- [42] J. Huuskonen, Prediction of Soil Sorption Coefficient of a Diverse Set of Organic Chemicals From Molecular Structure, *Journal of Chemical Information and Computer Sciences*, 43 (2003) 1457-1462.
- [43] M. Kah, C.D. Brown, Prediction of the Adsorption of Ionizable Pesticides in Soils, *Journal of Agricultural and Food Chemistry*, 55 (2007) 2312-2322.
- [44] S.W. Karickhoff, D.S. Brown, T.A. Scott, Sorption of hydrophobic pollutants on natural sediments, *Water Research*, 13 (1979) 241-248.
- [45] J.C. Means, S.G. Wood, J.J. Hassett, W.L. Banwart, Sorption of polynuclear aromatic hydrocarbons by sediments and soils, *Environmental Science & Technology*, 14 (1980) 1524-1528.

- [46] R.P. Schwarzenbach, J. Westall, Transport of nonpolar organic compounds from surface water to groundwater. Laboratory sorption studies, *Environmental Science & Technology*, 15 (1981) 1360-1367.
- [47] J.C. Means, S.G. Wood, J.J. Hassett, W.L. Banwart, Sorption of amino- and carboxy-substituted polynuclear aromatic hydrocarbons by sediments and soils, *Environmental Science & Technology*, 16 (1982) 93-98.
- [48] B.M. Gawlik, N. Sotiriou, E.A. Feicht, S. Schulte-Hostede, A. Kettrup, Alternatives for the determination of the soil adsorption coefficient, *K<sub>oc</sub>*, of non-ionic organic compounds — a review, *Chemosphere*, 34 (1997) 2525-2551.
- [49] Z. Gerstl, U. Mingelgrin, Sorption of organic substances by soils and sediments, *Journal of Environmental Science and Health, Part B*, 19 (1984) 297-312.
- [50] A. Sabljic, Predictions of the nature and strength of soil sorption of organic pollutants by molecular topology, *Journal of Agricultural and Food Chemistry*, 32 (1984) 243-246.
- [51] M. González, A. Helguera, I. Collado, A topological substructural molecular design to predict soil sorption coefficients for pesticides, *Mol Divers*, 10 (2006) 109-118.
- [52] G. Sposito, The Operational Definition of the Zero Point of Charge in Soils, *Soil Science Society of America Journal*, 45 (1981).
- [53] T.L. ter Laak, W.A. Gebbink, J. Tolls, Estimation of soil sorption coefficients of veterinary pharmaceuticals from soil properties, *Environmental Toxicology and Chemistry*, 25 (2006) 933-941.
- [54] A. Cherkasov, E.N. Muratov, D. Fourches, A. Varnek, I.I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y.C. Martin, R. Todeschini, V. Consonni, V.E. Kuz'min, R. Cramer, R. Benigni, C. Yang, J. Rathman, L. Terfloth, J. Gasteiger, A. Richard, A. Tropsha, QSAR Modeling: Where Have You Been? Where Are You Going To?, *Journal of Medicinal Chemistry*, 57 (2014) 4977-5010.
- [55] T. Pötter, H. Matter, Random or Rational Design? Evaluation of Diverse Compound Subsets from Chemical Structure Databases, *Journal of Medicinal Chemistry*, 41 (1998) 478-488.
- [56] T.M. Martin, P. Harten, D.M. Young, E.N. Muratov, A. Golbraikh, H. Zhu, A. Tropsha, Does Rational Selection of Training and Test Sets Improve the Outcome of QSAR Modeling?, *Journal of Chemical Information and Modeling*, 52 (2012) 2570-2578.
- [57] J.D. Walker, J. Jaworska, M.H.I. Comber, T.W. Schultz, J.C. Dearden, Guidelines for developing and using quantitative structure-activity relationships, *Environmental Toxicology and Chemistry*, 22 (2003) 1653-1665.
- [58] A. Franco, W. Fu, S. Trapp, Influence of soil pH on the sorption of ionizable chemicals: Modeling advances, *Environ Toxicol Chem*, 28 (2009) 458-464.
- [59] K. Roy, S. Kar, R.N. Das, Chapter 7 - Validation of QSAR Models, in: *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*, Academic Press, Boston, 2015, pp. 231-289.
- [60] A. Sabljic, H. Güsten, H. Verhaar, J. Hermens, QSAR modelling of soil sorption. Improvements and systematics of log *K<sub>oc</sub>* vs. log *K<sub>ow</sub>* correlations, *Chemosphere*, 31 (1995) 4489-4514.
- [61] J.R. Baker, J.R. Mihelcic, A. Sabljic, Reliable QSAR for estimating *K<sub>oc</sub>* for persistent organic pollutants: correlation with molecular connectivity indices, *Chemosphere*, 45 (2001) 213-221.

- [62] Walker, E. Jurado, Adsorption of isoproturon, diuron and metsulfuron-methyl in two soils at high soil:solution ratios, *Weed Research*, 38 (1998) 229-238.
- [63] M. Kah, C.D. Brown, Changes in pesticide adsorption with time at high soil to solution ratios, *Chemosphere*, 68 (2007) 1335-1343.
- [64] J.C. Dearden, M.T.D. Cronin, K.L.E. Kaiser, How not to develop a quantitative structure–activity or structure–property relationship (QSAR/QSPR), SAR and QSAR in *Environmental Research*, 20 (2009) 241-266.
- [65] S. Hillmer, D. Karney, In support of the assumptions at the foundation of Deming's management theory, *Journal of Quality Management*, 6 (2001) 371-400.
- [66] J. Villaverde, M. Kah, C.D. Brown, Adsorption and degradation of four acidic herbicides in soils from southern Spain, *Pest Manag. Sci.*, 64 (2008) 703-710.
- [67] K. Roy, QSAR Tools, in, Drug Theoretics and Cheminformatic Laboratory, <http://dtclab.webs.com/software-tools>.
- [68] University of Hertfordshire, PPDB: Pesticide Properties DataBase, in, <http://sitem.herts.ac.uk/aeru/ppdb/en/atoz.htm>.
- [69] United States Department of Agriculture, Soil Mechanics Level I. Module 3 - USDA Textural Soil Classification. Study Guide, in: USDA (Ed.), Soil Conservation Service, 1987.
- [70] D. Barthelmy, Mineralogy Database, in, <http://webmineral.com/>.
- [71] P.M. Huang, Environmental Impacts of Soil Component Interactions: Metals, Other Inorganics, and Microbial Activities, Volume II, CRC Press, 1995.
- [72] K.S.W. Sing, Reporting physisorption data for gas/solid systems with special reference to the determination of surface area and porosity (Recommendations 1984), *Pure Appl. Chem.*, 57 (1985) 603-619.
- [73] Á. Neira, Efecto de la granulometría sobre el proceso de adsorción de Carbaril en un suelo chileno, in: Departamento de Química, Universidad de Chile, Santiago, 2015.
- [74] L. Cáceres-Jensen, J. Rodríguez-Becerra, J. Parra-Rivero, M. Escudey, L. Barrientos, V. Castro-Castillo, Sorption kinetics of diuron on volcanic ash derived soils, *Journal of Hazardous Materials*, 261 (2013) 602-613.
- [75] G.A. Khoury, T.C. Gehris, L. Tribe, R.M. Torres Sánchez, M. dos Santos Afonso, Glyphosate adsorption on montmorillonite: An experimental and theoretical study of surface complexes, *Applied Clay Science*, 50 (2010) 167-175.
- [76] R. Todeschini, V. Consonni, R–Z, in: *Handbook of Molecular Descriptors*, Wiley-VCH Verlag GmbH, 2008, pp. 366-510.
- [77] R. Todeschini, V. Consonni, A–D, in: *Handbook of Molecular Descriptors*, Wiley-VCH Verlag GmbH, 2008, pp. 1-123.
- [78] R. Todeschini, P. Gramatica, New 3D molecular descriptors: the WHIM theory and QSAR applications, *Perspectives in Drug Discovery and Design*, 9 (1998) 355-380.
- [79] S. Hyun, L.S. Lee, Factors Controlling Sorption of Prosulfuron by Variable-Charge Soils and Model Sorbents, *Journal of Environmental Quality*, 33 (2004) 1354-1361.
- [80] Ministerio del Medio Ambiente, Ley N°19.300 sobre Bases Generales del Medio Ambiente. Ley Orgánica de la Superintendencia del Medio Ambiente, in: Ministerio del Medio Ambiente (Ed.), Chile, 2011, pp. 121.

- [81] C.T. Chiou, P.E. Porter, D.W. Schmedding, Partition equilibria of nonionic organic compounds between soil organic matter and water, *Environmental Science & Technology*, 17 (1983) 227-231.
- [82] C.T. Chiou, L.J. Peters, V.H. Freed, A Physical Concept of Soil-Water Equilibria for Nonionic Organic Compounds, *Science*, 206 (1979) 831-832.
- [83] E.G. Kenaga, CAI, Relationship Between Water Solubility, Soil Sorption, Octanol-Water Partitioning, and Concentration of Chemicals in Biota, in: J.P. Eaton, PR; Hendricks, AC (Ed.) *Aquatic Toxicology*, American Society of Testing Materials, 1980, pp. 78-115.
- [84] R. Koch, Molecular connectivity index for assessing ecotoxicological behaviour of organic compounds, *Toxicological & Environmental Chemistry*, 6 (1983) 87-96.
- [85] K. Schellenberg, C. Leuenberger, R.P. Schwarzenbach, Sorption of chlorinated phenols by natural sediments and aquifer materials, *Environmental Science & Technology*, 18 (1984) 652-657.
- [86] H.M. Seip, J. Alstad, G.E. Carlberg, K. Martinsen, R. Skaane, Measurement of mobility of organic compounds in soils, *Science of The Total Environment*, 50 (1986) 87-101.
- [87] J.J. Hassett, J.C. Means, W.L. Banwart, S.G. Wood, S. Ali, A. Khan, Sorption of Dibenzothiophene by Soils and Sediments, *Journal of Environmental Quality*, 9 (1980) 184-186.
- [88] D.A. Bahnick, W.J. Doucette, Use of molecular connectivity indices to estimate soil sorption coefficients for organic chemicals, *Chemosphere*, 17 (1988) 1703-1715.
- [89] A. Khan, J.J. Hassett, W.L. Banwart, Sorption of Acetophenone by sediments and soils, *Soil Science*, 128 (1979) 297-302.
- [90] B.T. Walton, M.S. Hendricks, T.A. Anderson, W.H. Griest, R. Merriweather, J.J. Beauchamp, C.W. Francis, Soil Sorption of Volatile and Semivolatile Organic Compounds in a Mixture, *Journal of Environmental Quality*, 21 (1992) 552-558.
- [91] P. Lagas, Sorption of chlorophenols in the soil, *Chemosphere*, 17 (1988) 205-216.
- [92] C.T. Jafvert, Sorption of organic acid compounds to sediments: Initial model development, *Environmental Toxicology and Chemistry*, 9 (1990) 1259-1268.
- [93] B. von Oepen, W. Kördel, W. Klein, Sorption of nonpolar and polar compounds to soils: Processes, measurements and experience with the applicability of the modified OECD-Guideline 106, *Chemosphere*, 22 (1991) 285-304.
- [94] W.L. Banwart, J.J. Hassett, S.G. Wood, J.C. Means, Sorption of Nitrogen-Heterocyclic compounds by soils and sediments, *Soil Science*, 133 (1982) 42-47.
- [95] C.A.M. van Gestel, W.-c. Ma, An approach to quantitative structure-activity relationships (QSARs) in earthworm toxicity studies, *Chemosphere*, 21 (1990) 1023-1033.
- [96] K.N. Reddy, M.A. Locke, Relationships between molecular properties and log P and soil sorption (K<sub>oc</sub>) of substituted phenylureas: QSAR models, *Chemosphere*, 28 (1994) 1929-1941.
- [97] R.D. Wauchope, T.M. Buttler, A.G. Hornsby, P.W.M. Augustijn-Beckers, J.P. Burt, The SCS/ARS/CES Pesticide Properties Database for Environmental Decision-Making, in: G. Ware (Ed.) *Reviews of Environmental Contamination and Toxicology*, Springer New York, 1992, pp. 1-155.
- [98] H. Lohninger, Estimation of soil partition coefficients of pesticides from their chemical structure, *Chemosphere*, 29 (1994) 1611-1626.

- [99] M.T. Saçan, I.A. Balcioglu, Prediction of the soil sorption coefficient of organic pollutants by the characteristic root index model, *Chemosphere*, 32 (1996) 1993-2001.
- [100] G.G. Briggs, A simple relationship between soil sorption of organic chemicals and their Octanol-water partitioning behavior, *Proceeding of the Seventh British Insecticide and Fungicide Conference*, 11 (1973) 475-478.
- [101] J.R. Baker, J.R. Mihelcic, D.C. Luehrs, J.P. Hickey, Evaluation of Estimation Methods for Organic Carbon Normalized Sorption Coefficients, *Water Environment Research*, 69 (1997) 136-145.
- [102] J.R. Baker, J.R. Mihelcic, E. Shea, Estimating Koc for persistent organic pollutants: limitations of correlations with Kow, *Chemosphere*, 41 (2000) 813-817.
- [103] G. Liu, J. Yu, QSAR analysis of soil sorption coefficients for polar organic chemicals: Substituted anilines and phenols, *Water Research*, 39 (2005) 2048-2055.
- [104] F. Yang, R. Qu, M. Wang, Y. Tang, M. Feng, Z. Wang, Experimental and QSPR study of sorption properties of polychlorinated diphenyl sulfides (PCDPSs) in Yangtze River plain soil, *Geoderma*, 193-194 (2013) 140-148.
- [105] B. Yaron, R. Calvet, R. Prost, Retention of Pollutants on and Within the Soil Solid Phase, in: *Soil Pollution*, Springer Berlin Heidelberg, 1996, pp. 99-148.
- [106] B. Yaron, R. Calvet, R. Prost, Reversible and Irreversible Retention — Release and Bound Residues, in: *Soil Pollution*, Springer Berlin Heidelberg, 1996, pp. 151-182.
- [107] B. Pan, B. Xing, Sorption Comparison between Pharmaceuticals and Hydrophobic Organic Chemicals in Soils and Sediments, in: J. Xu, D.L. Sparks (Eds.) *Molecular Environmental Soil Science*, Springer Netherlands, 2013, pp. 323-357.
- [108] G.G. Mauricio Escudey, Juan E. Forster, Margarita Briceño, Patricia Diaz, and Andrew Chang Chemical forms of phosphorus of volcanic ash-derived soils in Chile *Commun Soil Sci Plant Anal*, 32 (2001) 601- 616.
- [109] A.E.C. Lara, Estudio de Adsorción-Desorción y Cinética de Simazina, Atrazina, Picloram y Metolacloro en suelos derivados de cenizas volcánicas., in: *Facultad de Ciencias Naturales, Matemáticas y del Medio Ambiente. Escuela de Química., Universidad Tecnológica Metropolitana, Santiago, Chile, 2013.*
- [110] J.V. Herrera, Estudio de Adsorción-Desorción y Cinética de Glifosato en suelos derivados de cenizas volcánicas mediante detección electroquímica, in: *Departamento de Química, Facultad de Ciencias Básicas., Universidad Metropolitana de Ciencias de la Educación, Santiago, Chile, 2014.*
- [111] M.C. Huentemilla, Predicción de Transporte de Nicosulfurón en Suelos Derivados de Ceniza Volcánica utilizando STANMOD y HYDRUS-1D, in: *Departamento de Química, Facultad de Ciencias Básicas, Universidad Metropolitana de Ciencias de la Educación, Santiago, Chile, 2015.*
- [112] S. Azizian, Kinetic models of sorption: a theoretical analysis, *J. Colloid Interface Sci.*, 276 (2004) 47-52.
- [113] PubChem, PubChem Compound Database, in, <https://www.ncbi.nlm.nih.gov/pccompound>.
- [114] Web Interfaces for CDK Libraries, Web Server for Molecular Descriptor Calculation, in, <http://crdd.osdd.net:8081/webcdk/desc.jsp>.



- [115] K. Roy, I. Mitra, S. Kar, P.K. Ojha, R.N. Das, H. Kabir, Comparative Studies on Some Metrics for External Validation of QSPR Models, *Journal of Chemical Information and Modeling*, 52 (2012) 396-408.
- [116] E. Kolossov, R. Stanforth, The quality of QSAR models: problems and solutions, *SAR and QSAR in Environmental Research*, 18 (2007) 89-100.
- [117] K. Roy, S. Kar, R.N. Das, Chapter 6 - Selected Statistical Methods in QSAR, in: *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*, Academic Press, Boston, 2015, pp. 191-229.
- [118] G. Schüürmann, R.-U. Ebert, J. Chen, B. Wang, R. Kühne, External Validation and Prediction Employing the Predictive Squared Correlation Coefficient — Test Set Activity Mean vs Training Set Activity Mean, *Journal of Chemical Information and Modeling*, 48 (2008) 2140-2145.
- [119] A. Golbraikh, A. Tropsha, Beware of  $q^2$ !, *Journal of Molecular Graphics and Modelling*, 20 (2002) 269-276.
- [120] D. Werner, J.A. Garratt, G. Pigott, Sorption of 2,4-D and other phenoxy herbicides to soil, organic matter, and minerals, *Journal of Soils and Sediments*, 13 (2013) 129-139.
- [121] A.G. Ahangar, R.J. Smernik, R.S. Kookana, D.J. Chittleborough, Separating the effects of organic matter–mineral interactions and organic matter chemistry on the sorption of diuron and phenanthrene, *Chemosphere*, 72 (2008) 886-890.
- [122] R.J. Smernik, R.S. Kookana, The effects of organic matter–mineral interactions and organic matter chemistry on diuron sorption across a diverse range of soils, *Chemosphere*, 119 (2015) 99-104.
- [123] B. Xing, N. Senesi, P.M. Huang, *Biophysico-Chemical Processes of Anthropogenic Organic Compounds in Environmental Systems*, Wiley, 2011.

**ANEXO A:**

**Deficiencias en la creación de modelos (Q)SAR**

**Tabla A1.** Origen y tratamiento de los datos usados en la generación de modelos (Q)SAR.

Ref.	Origen de los datos	Características de los datos recopilados	Tratamiento de los datos
[50]	[44, 81-84]	Diferentes metodologías, adsorbentes (suelo, sedimento), y promedio de coeficientes de adsorción entre suelos	No se realiza
[15]	[31, 45-47, 50, 83, 85-87]	Diferentes metodologías ( <i>batch</i> y <i>columna</i> ), adsorbentes (suelo, sedimento), <i>endpoints</i> ( $K_d$ y $K_f$ son normalizados to $K_{oc}$ , incluso a bajo %CO). Se promedian los valores de $K_{om}$ entre suelos	No se realiza. Además, los autores asumen que $K_{om} = K_{oc}$ , es decir, $MO = CO$ en todos los adsorbentes
[32]	[15, 50, 83, 84, 88]	Ver ref. [15, 50]	Conversión de $K_{om}$ y $K_f$ a $K_{oc}$ . Se excluyen sales y suelos con bajo CO. Al promediar $K_{oc}$ , se excluyen valores muy diferentes entre sí
[30]	[44-47, 81, 82, 89-95]	Ver ref. [15, 50]	Conversión de $K_{om}$ y $K_f$ a $K_{oc}$ . Solo se usan isothermas lineales ( $n = 1$ ). Se descartan sedimentos y suelos con %CO < 0,1. No se promedian los valores de $K_d$ , con la finalidad de incorporar la variabilidad entre suelos
[96]	[31, 97]	Promedio de $K_{oc}$ entre suelos. Se asume que los compuestos ionizables se encuentran en su forma neutra	No se realiza
[98]	[15, 97]	Ver ref. [15]	Solo se usan compuestos con C, H, O, N, S, P, F, Cl y Br. Se excluyen sales, polímeros y mezclas de sustancias
[60]	[29, 32]	Ver ref. [15, 50]	Los valores de $K_{oc}$ son promediados
[99]	[31, 81]	Experimentos <i>batch</i> con condiciones similares. Algunas isothermas se asumen lineales. Se promedia $K_{om}$ entre suelos	Conversión de $K_{om}$ a $K_{oc}$
[33]	[60, 83, 100]	Ver ref. [15, 50]	Conversión de $K_{om}$ a $K_{oc}$ . Los valores de $K_{oc}$ son promediados
[34]	[60]	Ver ref. [15, 50]	Valores de $K_{oc}$ con bajo contenido de CO son excluidos
[61]	[60, 101, 102]	Ver ref. [15, 50]. En base a la ref. [101], se considera el efecto pH y la degradación, y se confirma el equilibrio químico	No se realiza
[35]	[15, 29, 60]	Ver ref. [15, 50]	No se realiza

**Tabla A1** (continuación).

Ref.	Origen de los datos	Características de los datos recopilados	Tratamiento de los datos
[42]	[60, 98] y bases de datos	Ver ref. [15, 50]	No se realiza
[103]	[60]	Ver ref. [15, 50]	No se realiza
[36]	[42]	Ver ref. [15, 50]	Los valores de $K_{oc}$ son promediados
[2]	[61]	Ver ref. [15, 50]	No se realiza
[51]	[34]	Ver ref. [15, 50]	No se realiza
[37]	[33, 42, 60]	Ver ref. [15, 50]	Los valores de $K_{oc}$ son promediados
[43]	[43]	Valores determinados por los mismos autores	Datos cuantificados en <i>batch</i> . Se incluye la variabilidad del suelo, sin embargo, se asume isoterma lineal ( $K_d$ )
[38]	[34, 60]	Ver ref. [15, 50]	No se realiza
[39]	[36, 37]	Ver ref. [15, 50]	No se realiza
[17]	[34, 60]	Ver ref. [15, 50]	No se realiza

**Tabla A2.** Modelos (Q)SAR para la adsorción de contaminantes orgánicos en suelos.

Ref.	Compuesto	Endpoint	n <sub>e</sub>	n <sub>p</sub>	Descriptores	Validación estadística	ρ	Intervalo de aplicación
<i>Modelos no clase-específicos para compuestos no ionizables</i>								
[15]	Apolar	log K <sub>om</sub>	185	-	<sup>1</sup> χ	EE, F <sup>1,70</sup> , R <sup>2</sup>	72	n.i.
[32]	Polar y apolar	log K <sub>oc</sub>	189	205	<sup>1</sup> χ, P <sub>f</sub>	R <sup>2</sup> , DE, EM, F	125	0.0 - 6.5 *
[60]	Polar y apolar	log K <sub>oc</sub>	81	-	<sup>1</sup> χ	R <sup>2</sup> , EE <sub>e</sub> , F <sup>1,79</sup> , Q <sup>2</sup> <sub>LOO</sub>	81	1.0 - 6.5
[60]	Polar y apolar	log K <sub>oc</sub>	390	-	log K <sub>ow</sub>	R <sup>2</sup> , EE <sub>e</sub> , F <sup>1,388</sup> , Q <sup>2</sup> <sub>LOO</sub>	390	0.0 - 6.5
[33]	Polar y apolar	log K <sub>oc</sub>	400	143	<sup>1</sup> χ <sup>v</sup> , <sup>2</sup> χ, <sup>6</sup> χ <sub>ch</sub> , P <sub>f</sub>	R <sup>2</sup> , EMA	133.3	0.0 - 6.4
[34]	Polar y apolar	log K <sub>oc</sub>	141	20	MW, nNO, nHA, CIC, MAXDP, Ts	R <sup>2</sup> , EE, Q <sup>2</sup> <sub>LOO</sub> , Q <sup>2</sup> <sub>LMO</sub> , DE <sub>e</sub> , DE <sub>p</sub> , F	23.5	1.0 - 5.5 *
[35]	Polar y apolar	log K <sub>oc</sub>	68	274	log K <sub>ow</sub> , R, P, SASA, SAV	Q <sup>2</sup> <sub>CV</sub> , DE <sub>e</sub> , DE <sub>p</sub> , ACP	13.6	n.i.
[42]	Polar y apolar	log K <sub>oc</sub>	403	165	log K <sub>ow</sub> , ROT, NAR, MW, I <sub>acid</sub>	R <sup>2</sup> , EE, F, Q <sup>2</sup> <sub>LOO</sub> , EE <sub>LOO</sub> , R <sup>2</sup> <sub>p</sub> , EE <sub>p</sub>	80.6	0.0 - 6.5
[36]	Polar y apolar	log K <sub>oc</sub>	457	114	MW, ε, E - state, P <sub>f</sub>	R <sup>2</sup> , EE, Q <sup>2</sup> <sub>LMO</sub> , Q <sup>2</sup> <sub>YRAND</sub> , Q <sup>2</sup> <sub>p</sub>	152.3	n.i.
[51]	Polar y apolar	log K <sub>oc</sub>	143	20	μ <sup>Dip</sup> <sub>10</sub> , μ <sup>Dip</sup> <sub>15</sub> , μ <sup>Dist</sup> <sub>4</sub> , μ <sup>H</sup> <sub>1</sub> , μ <sup>H</sup> <sub>5</sub> , μ <sup>P</sup> <sub>7</sub>	R <sup>2</sup> , EE, F, Q <sup>2</sup> <sub>LOO</sub> , R <sup>2</sup> <sub>prueba</sub> , EE <sub>p</sub>	23.8	0.1 - 5.1 *
[37]	Polar y apolar	log K <sub>oc</sub>	93	550	VED1, nHA, MAXDP, CIC0	R <sup>2</sup> , EE, Q <sup>2</sup> <sub>LOO</sub> , Q <sup>2</sup> <sub>BOOT</sub> , Q <sup>2</sup> <sub>YRAND</sub> , Q <sup>2</sup> <sub>p</sub> , EE <sub>p</sub> , DE <sub>e</sub> , DE <sub>p</sub> , ACP	23.25	0.0 - 6.0 *
[38]	Polar y apolar	log K <sub>oc</sub>	143	-	nCONN, ATS2p, O - 058, nP, Ds, Vm	R, EE, F, R <sub>LOO</sub> , EE <sub>LOO</sub> , AIC, FIT	23.8	n.i.
[39]	Polar y apolar	log K <sub>oc</sub>	594	107	log P, (log P) <sup>2</sup> , (log P) <sup>3</sup> , e <sup>log P</sup>	R <sup>2</sup> , EE, F, EM <sub>e</sub> , EMA <sub>e</sub> , DE <sub>e</sub> , EM <sub>p</sub> , EMA <sub>p</sub> , DE <sub>p</sub>	148.5	0.0 - 6.2 *
[17]	Polar y apolar	log K <sub>oc</sub>	143	20	log S <sub>p</sub> , Mv, VAR, MAXDP	R <sup>2</sup> , EE, EMA, F, Q <sup>2</sup> <sub>LOO</sub> , EE <sub>LOO</sub> , SEC <sub>LOO</sub> , Q <sup>2</sup> <sub>LMO</sub> , Q <sup>2</sup> <sub>YRAND</sub> , Q <sup>2</sup> <sub>p</sub> , EE <sub>p</sub> , SEC <sub>p</sub>	37.8	0.42 - 5.31
[40]	Polar y apolar	log K <sub>oc</sub>	644	320	MLOGP, X4v, DP15, AAC	R <sup>2</sup> , Q <sup>2</sup> <sub>LOO</sub> , Q <sup>2</sup> <sub>p</sub> , DE <sub>e</sub> , DE <sub>p</sub> , Q <sup>2</sup> <sub>YRAND</sub>	161	-1.0 - 6.5 *
[16]	Polar y apolar	log K <sub>oc</sub>	618	206	MLOGP2, α, O - 058, ATSC8v, nN, nROH, P - 117, SpMaxA_G/D, Mor16u	R <sup>2</sup> <sub>adj</sub> , DE <sub>e</sub> , Q <sup>2</sup> <sub>LOO</sub> , Q <sup>2</sup> <sub>p</sub> , DE <sub>p</sub> , FIV	68.7	0.00 - 6.96

**Tabla A2** (continuación).

Ref.	Compuesto	Endpoint	$n_e$	$n_p$	Descriptores	Validación estadística	$\rho$	Intervalo de aplicación
<i>Modelos clase-específicos para compuestos no ionizables</i>								
[96]	Fenilureas sustituidas	$\log K_{oc}$	44	-	$VDW_V, (VDW_V)^2, \mu, \mu^2, e_{LUMO}$	$R^2, F, CCM$	8.8	0.8 - 3.48
[99]	PCBs, fenoles y bencenos clorados	$\log K_{oc}$	36	-	CRI	$R^2, EE, F$	36	1.72 - 6.98
[34]	Carbamatos	$\log K_{oc}$	29	-	$nO, nX, nNO, \xi^C$	$R^2, EE, Q_{LOO}^2, Q_{LMO}^2, F^{4,24}$	7.25	0.42 - 3.35
[61]	COPs	$\log K_{oc}$	18	14	${}^1\chi, {}^4\chi_V, {}^3\chi_C$	$R^2, Q_{VC}^2, EE, F^{3,14}$	6	4.31 - 6.66
[103]	Anilinas sustituidas y fenoles	$\log K_{oc}$	42	-	$\log K_{ow}, E_{HOMO}, \alpha, \mu$	$R^2, R_{adj}^2, EE, F, FIV, Q_{CV}^2$	10.5	1.0 - 4.5*
[2]	COPs	$\log K_{oc}$	32	-	$Lu, DAI(CH_3-), DAI(\sim CH \sim), DAI(\sim C \approx), DAI(-C \approx)$ and $DAI(-O-)$	$R^2, EE, F, R_{LMO}, R_{LMO}^2, EE_{LMO}$	5.3	4.14 - 6.66
[104]	PCDPSs	$\log K_{oc}$	20	-	$E_{LUMO}, \alpha, S^{\theta}, ZPE, N_{3(5)}, N_o, N_m$	$R^2, EE, Q_{LOO}^2$	2.9	3.49 - 6.54
<i>Modelos no clase-específicos para compuestos ionizables</i>								
[43]	Ácidos	$\log K_d$	54	-	$\log D, OC, GATS7v$	$R^2$	18	-1.7 - 0.5*
[41]	Bases	$\log K_{oc}$	43	22	$\log P_n, \phi_n, \phi_{ion}$	$R^2, DE$	14.3	0.5 - 5.5*
<i>Modelos no clase-específicos para compuestos ionizables y no ionizables</i>								
[30]	Iónicos y no iónicos	$\log K_d$	229	500	$\log K_{ow}, \log f_{oc}, CFa, CFb'$	$EE, F$	57.25	-1.5 - 4.5

$n_e$ : Número de datos en el set de entrenamiento,  $n_p$ : Número de datos en el set de prueba,  $\rho$ : Relación entre número de observaciones y número de descriptores, **DE**: Desviación estándar. Subíndices **LOO** y **LMO** son tipos de validación cruzada (VC).  $Q_p^2$  y  $R_p^2$  se aplican al set de prueba. **PCBs**: Bifenilos policlorados, **COPs**: Compuestos orgánicos persistentes, **PCDPSs**: Sulfuros Bifenílicos policlorados. \*: Intervalo aproximado a partir del gráfico informado. Para información específica sobre cada descriptor, revisar la referencia respectiva.

**Tabla A3.** Problemas recurrentes y aspectos a evaluar en la realización de modelos (Q)SAR para adsorción de contaminantes en suelos.

<b>Deficiencias</b>	<b>Repercusiones</b>	<b>Aspectos a evaluar</b>
<i>Selección, tratamiento, representatividad y calidad de los datos</i>		
Variabilidad del suelo (espacial, temporal e inter-suelo)	Pérdida de información para la interpretación y predicción del <i>endpoint</i> . Inaplicabilidad del modelo con propósitos regulatorios	Considerar la variabilidad espacial, temporal e inter-suelo y descriptores que reflejen dicha variabilidad, que deban ser calculados constantemente
Contextualización del sistema herbicida-suelo	Incerteza en la representatividad del <i>endpoint</i> usado, metodología utilizada, descriptores propuestos y aplicabilidad real del modelo	Contextualización de los sistemas herbicida-suelo a modelar, basado en estudios científicos y/o líneas de base
Diferentes metodologías de obtención, sin un tratamiento posterior. Uso del mismo set de datos en diferentes estudios	Alta probabilidad de pérdida de sentido físico y químico, producto de variabilidad no explicada asociada a las diferentes metodologías	Tratamiento de los datos, considerando misma metodología. Recopilación de datos basada en objetivos (ejemplo: similitud con los sistemas herbicida-suelo que se pretende modelar)
<i>Selección de descriptores y generación del modelo</i>		
Exclusividad de descriptores moleculares	Sesgo en la generación de modelos, mayor probabilidad de baja calidad estadística, inaplicabilidad de estos modelos en suelos a diferentes condiciones locales	Incluir descriptores para los suelos y su interacción. Basar la selección de descriptores en la contextualización de los sistemas a modelar como complemento al criterio estadístico
Error asociado a coeficientes y claridad en la definición de descriptores moleculares	Incerteza frente al sentido físico y químico. Incerteza en la calidad estadística de los modelos generados	Incluir el error de todos los descriptores empíricos. En caso de descriptores teóricos asociados a empíricos, asegurar que sean representativos de los mismos
Análisis de colinealidad, significado, alcance y limitaciones de los descriptores que componen al modelo	Incerteza de la calidad estadística de los modelos, dificultades para definir su interpretación y aplicabilidad con propósitos regulatorios	Evaluar la colinealidad mediante FIV y dar una explicación a los descriptores seleccionados para abordar el Dominio de Aplicabilidad y la interpretación
<i>Aplicación de métodos estadísticos de validación</i>		
Ajuste de datos, número de descriptores usados	Baja calidad estadística de los modelos generados, dificultad al momento de comparar modelos	Usar como criterio el parámetro $\rho$ (relación datos/descriptores $\geq 5$ ). Usar un máximo de 5 descriptores en los modelos
Uso de set de prueba y cantidad de datos asociados al mismo	Incerteza en el poder predictivo de los modelos generados	Siempre considerar la creación de un set de prueba, en base al criterio de Tropsha
Comparación o validación estadística	Malas prácticas, interpretación errada, impedimento de la toma de decisiones racional frente a problemas ambientales	Aplicar métricas variadas. Si los modelos a comparar presentan diferente número de descriptores o datos, usar parámetros que den cuenta de eso (por ejemplo $R_{adj}^2$ )

## ANEXO B:

### Base teórico/conceptual para la generación y validación de datos y *endpoint*

Previo a la selección de un *endpoint* representativo, la calidad de los sets de datos está condicionada a los siguientes aspectos metodológicos:

- (i) Determinar los valores de *endpoint* bajo condiciones experimentales uniformes, describiendo la metodología para asegurar su reproducibilidad.
- (ii) Confirmar la existencia de equilibrio químico mediante estudios cinéticos y definir el intervalo de concentraciones bajo el cual es aplicable el *endpoint*, con base en las isothermas de adsorción.
- (iii) Analizar posibles variables confundentes, tales como: variación del pH y presencia de degradación durante los estudios de cinética y adsorción, suelo contaminado previo al estudio experimental; que afecten a la determinación de los valores de *endpoint* o la interpretación del modelo.

Muchos modelos (Q)SAR para predecir adsorción en suelos presentan buena calidad estadística aun sin realizar el tratamiento metodológico mencionado (Anexo A), dejando en evidencia que la validación estadística es una condición necesaria pero no suficiente para asegurar la calidad y confianza de dichos modelos.



Una vez generados y validados los datos, se escoge un *endpoint* a partir de las concentraciones de contaminante en el suelo ( $q_{eq}$ ) y la fase acuosa ( $C_{eq}$ ) [40, 103]:

$$q_{eq} = K_d * C_{eq} \quad (B1)$$

$$q_{eq} = K_f * C_{eq}^n \quad (B2)$$

$$K_{oc} = \{K_d \text{ o } K_f\} * \left(\frac{100}{\%CO}\right) \quad (B3)$$

Donde  $K_d$ ,  $K_f$  y  $K_{oc}$  son los tres *endpoint* usados para predecir adsorción en suelos, y  $n$  es un parámetro empírico para el ajuste del modelo de *Freundlich*.

Debido a su naturaleza empírica, todas las variantes del coeficiente de reparto tienen como ventaja la inclusión implícita de las propiedades locales de cada suelo y las interacciones contaminante-suelo asociadas, debido a su sensibilidad a las condiciones ambientales, como la erosión, y propiedades físicas y químicas de los suelos y contaminantes, como CE, CO, CIC,  $PIE_{suelo}$ , pH, reversibilidad, histéresis, hidratación del suelo, entre otras; así como a la presencia de varios compuestos y su interacción, por ejemplo competencia entre adsorbatos [105-107].

Lo anterior confiere al *endpoint* una interpretación compleja, convirtiéndolo en un parámetro útil para analizar tendencias y posibles escenarios de interés ambiental, pero no necesariamente para analizar mecanismos de adsorción. Esto último aumenta la necesidad de contextualizar el *endpoint* seleccionado al tipo de sistema contaminante-suelo y al interés perseguido con la creación del modelo (Q)SAR.

Por ejemplo, la representatividad de  $K_d$  está limitada a isotermas de adsorción lineales, mientras que  $K_f$  anula esta restricción mediante la inclusión del parámetro  $n$ .

La pérdida de linealidad suele producirse a altos valores de  $C_{eq}$  y se relaciona con altas concentraciones iniciales ( $C_0$ ), relacionadas con la dosis de aplicación. Por lo tanto,  $K_f$  es mejor *endpoint* que  $K_d$  si el set de datos incluye herbicidas no selectivos, eliminando la dependencia entre  $K_d$  y  $C_0$  cuando  $n \neq 1$  [29, 30]. Sin embargo, el uso de  $K_f$  como *endpoint* exige necesariamente que los valores de  $n$  sean similares, para facilitar la interpretación mecanística del proceso de adsorción global.

Basado en lo expuesto, la predicción del impacto ambiental a diferentes dosis de aplicación utilizando  $K_f$  requiere reducir la variabilidad de  $n$ , o incorporarlo como *endpoint* para la toma de decisiones ambientales mediante modelos (Q)SAR [29, 30].

Por otra parte, la normalización de  $K_d$  o  $K_f$  en Chile presenta dos problemas:

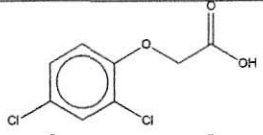
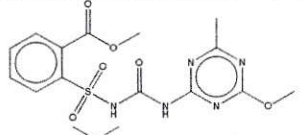
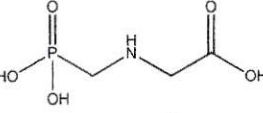
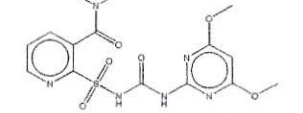
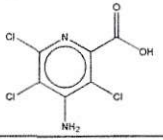
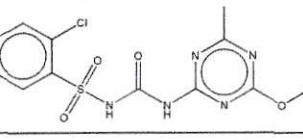
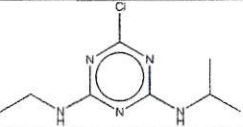
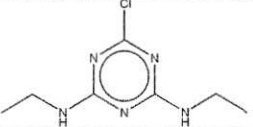
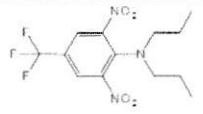
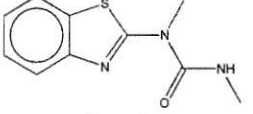
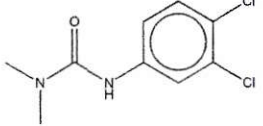
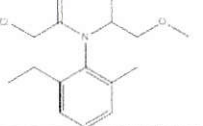
- (i) Para herbicidas polares e ionizables, no existe una correlación clara entre lipofilia y adsorción [93], con lo cual  $CO$  pierde representatividad y  $K_{oc}$  podría depender del contenido de arcilla, pH, área superficial, CIC y naturaleza o tipos de materia orgánica (MO) presentes en el suelo [35].
- (ii) En SDCV, la relevancia de la interacción herbicida-óxidos de Fe en el proceso de adsorción de herbicidas en SDCV aumenta la incerteza frente a la normalización del coeficiente de adsorción, sobre todo para la adsorción de herbicidas ionizables.

Lo anterior sugiere evitar cualquier tipo de normalización en caso que exista adsorción significativa en varios componentes del suelo [1] o los suelos presenten un bajo contenido de  $CO$ , donde la normalización a  $K_{oc}$  pierde sentido físico.

## **ANEXO C:**

### **Propiedades y caracterización de herbicidas y suelos empleados en este trabajo**

**Tabla C1.** Herbicidas usados en el presente estudio.

Nombre	Molécula	pKa	K <sub>ow</sub>	S <sub>w</sub>	Nombre	Molécula	pKa	K <sub>ow</sub>	S <sub>w</sub>
<i>Ácidos</i>									
2,4-D		3,4	1,51*10 <sup>-1</sup>	1,10*10 <sup>-1</sup>	Metsulfurón metilo (MSM)		3,75	1,35*10 <sup>-2</sup>	7,32*10 <sup>-3</sup>
Glifosato (GFS)		2,34	6,31*10 <sup>-4</sup>	6,21*10 <sup>-2</sup>	Nicosulfurón (NS)		4,78	4,07	1,83*10 <sup>-2</sup>
Picloram (PC)		2,3	1,20*10 <sup>-2</sup>	2,32*10 <sup>-3</sup>	Clorsulfurón (CS)		3,4	1,02*10 <sup>-1</sup>	3,49*10 <sup>-2</sup>
<i>Básicos</i>									
Atrazina (AT)		1,7	5,01*10 <sup>2</sup>	1,62*10 <sup>-4</sup>	Simazina (SM)		1,62	2,00*10 <sup>2</sup>	2,48*10 <sup>-5</sup>
<i>No ionizables</i>									
Trifluralín (TF)		-	1,86*10 <sup>5</sup>	6,59*10 <sup>-7</sup>	Metabentiazurón (MBT)		-	4,37*10 <sup>2</sup>	2,71*10 <sup>-4</sup>
Diurón (DI)		-	7,41*10 <sup>2</sup>	1,53*10 <sup>-4</sup>	Metolacloro (MC)		-	2,51*10 <sup>3</sup>	1,87*10 <sup>-3</sup>

Valores de pKa (25°C), K<sub>ow</sub> (pH 7, 20°C) y S<sub>w</sub> (mol L<sup>-1</sup>, pH 7, 20°C, calculado a partir de S<sub>w</sub> en mg L<sup>-1</sup> y M en g mol<sup>-1</sup>) de la Base de Datos de Propiedades de Pesticidas (PPDB), de la Universidad de Hertfordshire [68]

**Tabla C2.** Propiedades fisicoquímicas principales de SDCV usados en el presente estudio\*.

Nombre	Clasificación	Ubicación	CO (%)	Arena (%)	Limo (%)	Arcilla (%)	Fe <sub>PYRO</sub> (%)	Fe <sub>OX</sub> (%)	Fe <sub>DCB</sub> (%)	PIE <sub>suelo</sub>	CIC (cmol kg <sup>-1</sup> )	pH agua (2:5)	Humedad (%)
<i>Ultisol</i>													
Collipulli (COLL)	<i>Fine, Mesic, Xeric, Paleumult</i>	36°58'S; 72°09'W	1,5	13,7	40,7	45,7	0,7	0,9	6,2	2	8,7	5,2	12,7
Metrengo (MET)	<i>Fine, Mesic, Paleumult</i>	38°34'S; 72°22'W	2,3	8	56,7	35,3	0,8	1,8	7,1	2,5	9,3	4,7	9,5
<i>Andisol</i>													
Freire (FRE)	<i>Medial, Mesic, Xeric, Placandept</i>	38°57'S; 72°36'W	4,5	21,3	54,2	24,5	2,2	2,5	4,3	3,1	9,9	4,4	10,1
Santa Bárbara (STB)	<i>Ashy, Medial, Mesic, Typic, Dystrandept</i>	36°50'S; 71°55'W	5,1	7,2	66,5	26,3	0,3	1,9	5,3	3,8	10,3	5,7	14,3
Osorno (OSN)	<i>Medial, Mesic, Typic, Dystrandept</i>	40°32'S; 73°05'W	5,1	10,1	50,9	39,1	1,4	2,0	3,0	2,1	9,8	4,6	8,7
Diguillín (DIG)	<i>Medial, Thermic, Typic, Dystrandept</i>	36°53'S; 72°10'W	5,8	35,5	45,1	19,4	0,4	1,9	3,5	2,6	11,8	6,2	21,3
Temuco (TCO)	<i>Medial, Mesic, Entic, Dystrandept</i>	38°56'S; 72°36'W	6,4	16,1	58,2	25,7	0,7	2,2	3,9	2,9	12,1	5,4	9,9
Ralún (RAL)	<i>Mesic, Umbric, Vitrandept</i>	41°32'S; 73°05'W	6,9	47,3	38,5	12,9	1,8	1,4	1,4	3,3	7,1	4,4	45,1
Nueva Braunau (NBR)	<i>Ashy, Mesic, Hydric, Dystrandept</i>	41°19'S; 73°06'W	9,5	6,2	66,2	27,6	1,8	3,3	5,1	3,3	10,3	4,1	46,5
Frutillar (FRU)	<i>Medial, Isomesic, Typic, Placandept</i>	41°06'S; 73°07'W	11	16,3	63,9	19,7	1,0	0,6	0,6	2,9	9,5	4,1	39,6

\* Datos provenientes del Informe final del Proyecto FONDECYT INICIACIÓN N° 11110421. Todos los suelos fueron caracterizados bajo la misma metodología.

**Fe<sub>PYRO</sub>**: Óxidos de Fe complejados con la MO, **Fe<sub>OX</sub>**: Óxidos de Fe amorfo, **Fe<sub>DCB</sub>**: Óxidos de Fe libre.

**Tabla C3.** Composición mineralógica de los SDCV [108].

Mineral	COLL	MET	FRE	STB	OSN	DIG	TCO	RAL	NBR	FRU
Alofán			+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++
$\alpha$ -Cristobalita	+		+		+		++	+	+	+
Clorita-Al				+				++		
Feldespato					+		+		+	
Ferrihidrita			+			+	+		+	
Gibbsita			++	+			++		++	
Goetita		+								
Haloisita	+	+++++		++	+++	++				+
Caolinita	+++++									
Montmorillonita								+		
Organo-alofánico			++	+	++	+	++		++	+
Plagioclasa					+	++		++		+
Cuarzo		+	+							
Vermiculita	+			++	+	+				++

+++++ Dominante (> 50 %), ++++ Abundante (20 – 50 %), +++ Común (5 – 20 %), ++ Presente (1 – 5 %), + Fracción traza (< 1 %).

## ANEXO D:

### Validación de los valores empíricos empleados en el set de datos

Los parámetros de calidad analítica para el método de detección de cada herbicida junto con los datos de estudios cinéticos y de adsorción fueron obtenidos del Laboratorio de Fisicoquímica & Analítica (PACHEM, Facultad de Ciencias Básicas, Universidad Metropolitana de Ciencias de la Educación), usando la misma metodología en todos los casos [11, 109-111].

#### Parámetros de calidad analítica y validación del método

La cuantificación de los 12 herbicidas se realizó por Cromatografía Líquida de Alta Resolución con Detector de Arreglo de Diodos (HPLC-DAD, *SHIMADZU*), conformado por los siguientes módulos: desgasificador (*serie DGU-20A<sub>5</sub>*), bomba de gradiente cuaternaria (*serie LC-20AT*), automuestreador (*serie SIL-20AC*), detector DAD (*serie SPD-M20A*) y horno (*serie CTO-20AC*).

Las condiciones de trabajo fueron: temperatura de 35°C, flujo de 1,1 mL min<sup>-1</sup>, volumen de inyección de 20 µL, columna *MultoHigh 100 RP C<sub>18</sub> 5µ* (150 x 4,6 mm ID) y el software *LabSolution* para la adquisición y procesamiento de datos.

La descripción detallada del procedimiento analítico incluye: (i) Fase móvil compuesta por ACN:H<sub>2</sub>O en diferentes proporciones y en condiciones ácidas (pH entre 2,3 y 2,8) para la cuantificación de herbicidas ácidos, (ii) Detección a longitudes de onda entre 212 y 240 nm, (iii) tiempos de retención entre 2,8 y 5,6 min, (iv) Límite de detección entre 0,0022 y 0,0194 µg mL<sup>-1</sup> y (v) Límite de cuantificación entre 0,075 y 0,0648 µg mL<sup>-1</sup>.

### **Estudio cinético**

La concentración de herbicida adsorbida en cada SDCV se calculó a partir de las concentraciones inicial ( $C_0$ , obtenida de la preparación de la muestra) y final ( $C_t$ , detectadas por HPLC-DAD), el volumen de solución ( $V$ ) y la masa de suelo seco ( $m_s$ ):

$$q_t = \frac{V * (C_0 - C_t)}{m_s} \quad (D1)$$

La ecuación D1 supone que la concentración adsorbida equivale a la disminución de la concentración de herbicida en solución, ( $C_0 - C_t$ ), la cual es normalizada por  $V/m_s$  debido al cambio de matriz (de solución a suelo). Sin embargo, lo anterior implica que la concentración no disminuye por procesos laterales como degradación, ni variaciones externas tales como contaminación difusa o dependencia del pH. Por lo tanto, con la finalidad de reducir variables confundentes durante la metodología y posibles sesgos en la interpretación de los valores obtenidos, se analizó:

- (i) pH de cada muestra, en caso de dependencia del pH en el grado de ionización del herbicida y la carga superficial del suelo durante el estudio cinético.



- (ii) Presencia de metabolitos generados durante los estudios de cinética e isotermas, a partir de los cromatogramas obtenidos para cada herbicida.
- (iii) Concentración de cada herbicida en suelos sin contaminar para descartar contaminación difusa que pudiese afectar las mediciones.

### **Tiempo de equilibrio**

Para la determinación del tiempo de equilibrio, se consideró el modelo de pseudo-segundo orden, que ajusta correctamente a bajas concentraciones iniciales de adsorbato [74, 112]. Este modelo está definido por la ecuación:

$$\frac{dq_t}{dt} = k_2(q_{m\acute{a}x} - q_t)^2 \quad (D2)$$

Donde  $k_2$  es la constante de velocidad de pseudo-segundo orden y  $q_{m\acute{a}x}$  la concentración máxima adsorbida por el suelo. Ambos parámetros fueron proporcionados por el Laboratorio de Físicoquímica & Analítica [11, 109-111], a partir de la integración (suponiendo que  $q_t = 0$  en  $t = 0$ ) y posterior linealización de la ecuación D2:

$$\frac{t}{q_t} = \frac{1}{(k_2 * q_{m\acute{a}x})^2} + \frac{1}{q_{m\acute{a}x}} t \quad (D3)$$

El ajuste de la ecuación D3 fue validado estadísticamente solo si (i)  $R^2 \geq 0,96$ , (ii) error relativo  $\leq 20\%$  para todos los parámetros y (iii) valor de  $p < 0,05$ .

Se estableció el equilibrio cuando la diferencia entre  $q_t$  y  $q_{m\acute{a}x}$  fue menor a 5%.

### **Estudio de adsorción**

La ecuación utilizada para la determinación de la concentración adsorbida fue:

$$q_{eq} = \frac{V * (C_0 - C_{eq})}{m_s} \quad (D4)$$

Con  $q_{eq}$  y  $C_{eq}$  equivalentes a  $q_t$  y  $C_t$  del estudio cinético (ecuación D1), respectivamente.

El uso de los datos fue validado para tiempos de equilibrio menores a 24 h (tiempo de agitación del estudio de adsorción).

Por último, con la finalidad de reducir variables confundentes se realizó el mismo análisis del estudio cinético.

### **Modelos de adsorción**

Se utilizaron los coeficientes de adsorción provenientes de los modelos lineal y *Freundlich*, calculados por el Laboratorio de Fisicoquímica & Analítica para todas las isothermas de adsorción mediante el programa *SigmaPlot 12.0*, usando la ecuación “Standard Curves, Linear Curve” ( $f = y_0 + a * x$ ) para probar linealidad ( $K_d$ ) y “Power, 2 Parameter” ( $f = a * x^b$ ) para casos no lineales ( $K_f$ ) [11, 109-111].

Para la validación de los modelos, se consideraron (i)  $R^2 \geq 0,96$ , (ii) error relativo  $\leq 20\%$  para todos los parámetros y (iii) valor de  $p < 0,05$ .

Adicionalmente, se utilizaron los coeficientes de adsorción normalizados al contenido de CO solo de los valores previamente validados.

### **Validación de la existencia de equilibrio químico**

Los datos cinéticos asociados a los 12 herbicidas en los 10 SDCV empleados en el presente trabajo fueron estadísticamente validados. Su correlación con el modelo de *pseudo-segundo orden* podría sugerir (i) similitud mecanística para el proceso

cinético global de adsorción de los herbicidas en SDCV, con quimisorción dependiente del número de sitios de adsorción disponibles [74], y/o (ii) relación con las condiciones experimentales, particularmente la baja relación suelo/solución (2,0 g de suelo seco al aire, 10 mL de solución) o la baja concentración inicial de herbicida en solución [74].

El tiempo de equilibrio fue inferior al tiempo de contacto empleado en los estudios de adsorción (24 h), lo que permite utilizar los valores de coeficientes de adsorción de estos herbicidas en SDCV para la generación de modelos (Q)SAR.

Respecto al posible impacto de variables confundentes en la determinación de los valores de *endpoint* para la generación e interpretación del modelo (Q)SAR: (i) no se encontraron diferencias significativas de pH durante el proceso de adsorción, (ii) no se detectaron productos de degradación en los cromatogramas para ninguno de los doce herbicidas, y (iii) la concentración de herbicidas en las muestras de suelo sin contaminar se encontró bajo el límite de detección, descartando contaminación difusa. Por lo tanto, se descarta la presencia de variables que afecten a la determinación de los valores de *endpoint* y a la interpretación del modelo.

## ANEXO E:

### Descripción de la metodología y software usados para la creación de modelos

#### (Q)SAR para predecir coeficientes de adsorción de herbicidas en SDCV

#### Cálculo y generación de descriptores moleculares

Se utilizó la página web *PubChem* [113] para conocer la estructura molecular 3D de los herbicidas a usar, usadas como datos de entrada (.sdf) en el “Servidor Web para el Cálculo de Descriptores Moleculares” [114], obteniendo 174 descriptores teóricos:

- (i) Topológicos: Carbon Connect Order Zero, Carbon Connect Order One, Chi Chain Indices, Chi Cluster Indices, Chi path Indices, Chi path Cluster Indices, Eccentric Connectivity Index, Kappa Shape Indices, MDE, Autocorrelation Polarizability, Autocorrelation Charge, Autocorrelation Mass, Petitjean Number, TPSA, Vertex Adj Magnitude, Weighted Path, Weinner Number, Zagreb Index, WHIM, BCUT, Atomic valence connectivity index (order 0), Atomic valence connectivity index (order1), Fragment Complexity.
- (ii) Electrónicos: Ionizational Potential, Charged Partial Surface Area, Bpol.
- (iii) Geométricos: Gravitational Index, Length Over Breadth, Moment Of Inertia, Petitjean Shape Index.

- (iv) Constitucionales: XlogP, Rotatable Bonds Count, Molecular Weight, Longest Aliphatic Chain, Lipinski Rule of Five, Largest Pi System, Largest Chain, Atom Count, Bond Count, Aromatic Bond Count, Hydrogen Bond Acceptors, Hydrogen Bond Donator, Aromatic Atom Count, Atomic Polarizations.

A lo anterior, se sumaron descriptores calculados con el programa *EPI SUITE 4.11* (EPA): MW (Masa molecular), KOWWIN ( $K_{ow}$  teórico), BoilPoint (Punto de ebullición), MeltPoint (Punto de fusión), VapPres (Presión de vapor), WSKOW ( $S_w$  teórico) y  $\log K_{oa}$  (Logaritmo del coeficiente de partición octanol/aire).

Por último, se incluyó la forma logarítmica de la solubilidad en agua ( $\log S_w$ ) y la lipofilia ( $\log K_{ow}$ ), obtenidos de la Tabla C1, junto con los siguientes descriptores:

- (i)  $\log S_p$ , propuesto por Do Reis y col. (2014) como corrección al uso individual de  $S_w$  y  $K_{ow}$  en compuestos no iónicos [17]:

$$\log S_p = \log S_w - \log(K_{ow} + 1) \quad (E1)$$

- (ii) El grado de ionización de los herbicidas se incorporó mediante los descriptores CFa y CFb', propuestos por Bintein y Devillers [30]:

$$CFa = \log\left(\frac{1}{1 + 10^{pH-pKa}}\right) \quad (E2)$$

$$CFb' = \log\left(\frac{1}{1 + 10^{pKa-(pH-2)}}\right) \quad (E3)$$

Estos descriptores corresponden al logaritmo de la fracción de herbicida que permanece neutra al pH de la solución, donde CFa se aplica a compuestos ácidos y CFb' a compuestos básicos, presentando un valor nulo en los demás casos.

## Cálculo y generación de descriptores de SDCV

Se utilizaron propiedades fisicoquímicas, como contenido de CO, arena, limo, arcilla,  $Fe_{PYRO}$ ,  $Fe_{OX}$  y  $Fe_{DCB}$  (Tabla C2), para calcular descriptores para SDCV, las cuales fueron expresadas en su forma fraccional logarítmica. Ejemplificando con el contenido de CO, se obtiene el siguiente descriptor:

$$\log f_{CO} = \log \left( \frac{\%CO}{100} \right) \quad (E4)$$

Adicionalmente, se consideró como descriptor la relación entre contenido de arcilla y CO ( $R_{Arc-CO}$ ), para dar cuenta de la adsorción en la fracción mineral en caso que CO no sea representativo por si solo del proceso de adsorción [66].

$$R_{Arc-CO} = \log \left( \frac{\%arcilla}{\%CO} \right) \quad (E5)$$

Con base en las ecuaciones E2 y E3, se generó una representación de la fracción de carga total positiva ( $CF_+$ ) y negativa ( $CF_-$ ) que presenta cada SDCV:

$$CF_+ = \log \left( \frac{1}{1 + 10^{pH_{suelo} - PIE_{suelo}}} \right) \quad (E6)$$

$$CF_- = \log \left( \frac{1}{1 + 10^{PIE_{suelo} - pH_{suelo}}} \right) = 1 - CF_+ \quad (E7)$$

Usando como  $pH_{suelo}$  el pH medido en solución acuosa en relación 2:5 (Tabla C2).

## Condiciones de los software empleados

### 1. Análisis exploratorio de la matriz de descriptores (*Data Pre Treatment GUI*)

- (i) Variance cut-off = 0.0001
- (ii) Correlation Coefficient Cut-off = 0.9999

## 2. Calidad estadística de las bases de datos (*Data Pre Treatment GUI*)

- (i) Variance cut-off = 0.0001
- (ii) Correlation Coefficient Cut-off = 0.9

El proceso se realizó para cada base de datos, incorporando los valores de *endpoint* en el archivo de entrada (.xlsx), así el software considere su variabilidad y relación con los descriptores.

El uso de  $r = 0,9$  como coeficiente de corte podría generar una alta colinealidad entre los descriptores del modelo. Sin embargo, la posible existencia de efectos compensatorios al combinar descriptores altamente correlacionados con el *endpoint* y otros poco correlacionados con el mismo, implica no descartar estos últimos. Luego, la colinealidad fue descartada durante la validación interna.

## 3. División en sets de entrenamiento y de prueba (*Dataset Division GUI 1.2*)

- (i) Number of compound in each cluster: Para una alta disponibilidad de datos (>50) se usó como valor 5, mientras que para una escasa cantidad de datos (30 - 50), se utilizó como valor  $(0,1 * n^{\circ} \text{ datos})$  aproximado al número entero inferior.
- (ii) Seed number: 1

Adicionalmente, para validar la calidad de los valores que componen al set de prueba se aseguró que:

- (i) Cubrieran un amplio intervalo del set de entrenamiento.
- (ii) No afectaran a la calidad estadística del set de entrenamiento.
- (iii) No fueran valores extremos del intervalo de datos.

Para asegurar la representatividad estadística del set de prueba:

- (i) Se compuso de al menos 10 datos.
- (ii) Si el número de datos en cada matriz lo permitía, se tomó entre el 15 y 20% de los datos totales. En caso contrario (matrices pequeñas), la restricción fue  $\geq 15\%$ .

#### 4. Selección de las mejores combinaciones de descriptores (*Best Subset Selection Modified v2.1*)

- (i) Combination of 'n' descriptors = Entre 1 y 5.
- (ii) Cut-off for  $r^2$  value = 0.6
- (iii) Inter correlation between descriptors cut-off = 1

Se usaron el set de entrenamiento y de prueba como datos de entrada. Los parámetros output del modelo ( $R^2$ ,  $Q_{LOO}^2$ ,  $Q_{F1}^2$  y  $Q_{F2}^2$ ), fueron contrastados con los requisitos expuestos en las Tablas E1, E2 y E3. En caso de incumplimiento de dichos requisitos, se seleccionaron las dos combinaciones más próximas a cumplirlos.

#### **Parámetros de calidad estadística para la validación de los modelos (Q)SAR**

Para la validación de los modelos (Q)SAR, fueron utilizados los parámetros de calidad estadística mencionados en las Tablas E1, E2, E3 y E4.



**Tabla E1.** Parámetros de calidad estadística para la bondad del ajuste de modelos (Q)SAR.

Parámetro de calidad estadística	Ecuación asociada	Requisito	Interpretación
Relación entre número de observaciones y número de descriptores ( $\rho$ )	$\rho = \frac{n_e}{p}$	$\rho \geq 5$	Modelo no sobreajustado [64]
Coefficiente de determinación ( $R^2$ )	$R^2 = 1 - \frac{\sum_{i=1}^n (Y_{\text{obs}(i)[e]} - Y_{\text{pred}(i)[e]})^2}{\sum_{i=1}^n (Y_{\text{obs}(i)[e]} - \bar{Y}_{\text{obs}[e]})^2}$	$R^2 > 0,6$	Grado de ajuste aceptable [115]. No detecta errores sistemáticos [116]
Estadístico F ( $F^{\text{m,gl}}$ )	$F^{\text{m,gl}} = \frac{(n_e - p - 1) * R^2}{p * (1 - R^2)}$	$F^{\text{m,gl}} > F_{\text{crítico}}^{\text{m,gl}}$	Relación lineal estadísticamente significativa [115]
Suma de los errores cuadráticos ( $\text{SEC}_e$ )	Con $m = p$ y $gl = n_e - p - 1$ $\text{SEC}_e = \sum_{i=1}^{n_{e,p}} (Y_{\text{obs}(i)[e,p]} - Y_{\text{pred}(i)[e,p]})^2$	Minimizar $\text{SEC}_e$	Bajo error en términos de sus errores cuadráticos [115]
Error estándar ( $\text{EE}_e$ )	$\text{EE}_e = \sqrt{\frac{\text{SSE}_e}{n_{e,p} - p - 1}}$	Minimizar $\text{EE}_e$	Bajo error del modelo, considerando sus grados de libertad [115]
Error estándar normalizado al intervalo de <i>endpoint</i> (EEN)	$\text{EEN} = \frac{\text{SE}_e}{Y_{\text{máx obs}[e]} - Y_{\text{mín obs}[e]}}$	$\text{EEN} \leq 0,1$	Calidad estadística, considerando la dependencia del intervalo de valores de <i>endpoint</i> [2]
Error medio (EM) y error medio absoluto (EMA)	$\text{EM} = \sum_{i=1}^{n_e} \frac{(Y_{\text{obs}(i)[e]} - Y_{\text{pred}(i)[e]})}{n_e}$ $\text{EMA} = \sum_{i=1}^{n_e} \frac{ Y_{\text{obs}(i)[e]} - Y_{\text{pred}(i)[e]} }{n_e}$	$\text{EM} \ll \text{EMA}$	Se descarta la presencia de errores sistemáticos [64]
Factor de Inflación de la Varianza (FIV)	$\text{FIV}(x) = \frac{1}{1 - R_x^2}$	$\text{VIF}(x) \leq 5$	Correlación aceptable entre descriptores [117]

$n_e$ : Número de datos en el set de entrenamiento,  $p$ : Número de descriptores del modelo,  $Y_{\text{obs}(i)[e,p]}$  y  $Y_{\text{pred}(i)[e,p]}$ : Dato  $i$  observado y predicho en el set de entrenamiento [e] o de prueba [p], respectivamente,  $\bar{Y}_{\text{obs}[e,p]}$ : Promedio de los datos empíricos en el set de entrenamiento [e] o de prueba [p],  $Y_{\text{máx obs}[e]}$  y  $Y_{\text{mín obs}[e]}$ : Máximo y mínimo valor empírico en el set de entrenamiento, respectivamente.  $R_x^2$ : Coeficiente de correlación de cada descriptor  $x$  en función del resto de descriptores del modelo.

**Tabla E2.** Parámetros de calidad estadística para la robustez del modelo (Q)SAR.

Ecuaciones asociadas	Requisito	Interpretación
<i>Aleatorización en Y</i>		
$Q_{\text{YRAND}}^2 = 1 - \frac{\sum_{i=1}^n (Y_{\text{obs}(i)[e]} - Y_{\text{pred}(i)[e]}^{\text{rand}})^2}{\sum_{i=1}^n (Y_{\text{obs}(i)[e]} - \bar{Y}_{\text{obs}[e]})^2}$	$\bar{Q}_{\text{YRAND}}^2 = \sum_{j=1}^{50} \frac{Q_{\text{YRAND}(j)}^2}{50}$ $cR_p^2 = \sqrt{R^2 * (R^2 - \bar{Q}_{\text{YRAND}}^2)}$	$cR_p^2 \geq 0,5$ La correlación obtenida para el modelo (Q)SAR no es azarosa [115]
<i>Coficiente de determinación de la validación cruzada (Leave-One-Out)</i>		
$Q_{\text{LOO}}^2 = 1 - \frac{\sum_{i=1}^n (Y_{\text{obs}(i)[e]} - Y_{\text{pred}(i)[e]}^{\text{VC}})^2}{\sum_{i=1}^n (Y_{\text{obs}(i)[e]} - \bar{Y}_{\text{obs}[e]})^2}$	$Q_{\text{LOO}}^2 > 0,5$	El modelo (Q)SAR es robusto [115]
<i>Métrica <math>r_m^2</math></i>		
$Y_{\text{sc obs}(i)} = \frac{Y_{\text{obs}(i)[e]} - Y_{\text{min obs}[e]}}{Y_{\text{max obs}[e]} - Y_{\text{min obs}[e]}}$	$Y_{\text{sc pred}(i)} = \frac{Y_{\text{pred}(i)[e]} - Y_{\text{min obs}[e]}}{Y_{\text{max obs}[e]} - Y_{\text{min obs}[e]}}$	
$k_{\text{LOO}} = \frac{\sum_{i=1}^n Y_{\text{sc obs}(i)} * Y_{\text{sc pred}(i)}}{\sum_{i=1}^n (Y_{\text{sc pred}(i)})^2}$	$k'_{\text{LOO}} = \frac{\sum_{i=1}^n Y_{\text{sc obs}(i)} * Y_{\text{sc pred}(i)}}{\sum_{i=1}^n (Y_{\text{sc obs}(i)})^2}$	
$r_{0(\text{LOO})}^2 = 1 - \frac{\sum_{i=1}^n (Y_{\text{sc obs}(i)} - k_{\text{LOO}} * Y_{\text{sc pred}(i)})^2}{\sum_{i=1}^n (Y_{\text{sc obs}(i)} - \bar{Y}_{\text{sc obs}})^2}$	$r'_{0(\text{LOO})}^2 = 1 - \frac{\sum_{i=1}^n (Y_{\text{sc obs}(i)} - k'_{\text{LOO}} * Y_{\text{sc pred}(i)})^2}{\sum_{i=1}^n (Y_{\text{sc obs}(i)} - \bar{Y}_{\text{sc pred}})^2}$	$\bar{r}_{m(\text{LOO})}^2 > 0,5$
$r_{m(\text{LOO})}^2 = R^2 * \left(1 - \sqrt{(R^2 - r_{0(\text{LOO})}^2)}\right)$	$r'_{m(\text{LOO})}^2 = R^2 * \left(1 - \sqrt{(R^2 - r'_{0(\text{LOO})}^2)}\right)$	$\Delta r_{m(\text{LOO})}^2 < 0,2$
$\bar{r}_{m(\text{LOO})}^2 = \frac{r_{m(\text{LOO})}^2 + r'_{m(\text{LOO})}^2}{2}$	$\Delta r_{m(\text{LOO})}^2 =  r_{m(\text{LOO})}^2 - r'_{m(\text{LOO})}^2 $	El modelo (Q)SAR es robusto [115]. Esta robustez corrige la falsa robustez en $Q_{\text{LOO}}^2$ cuando el intervalo de valores de <i>endpoint</i> es amplio

$R^2$ : Coeficiente de determinación del modelo,  $Y_{\text{obs}(i)[e]}$  y  $Y_{\text{pred}(i)[e]}$ : Dato  $i$  observado y predicho en el set de entrenamiento, respectivamente,  $\bar{Y}_{\text{obs}[e]}$ : Promedio de los datos empíricos en el set de entrenamiento,  $Y_{\text{pred}(i)[e]}^{\text{rand}}$ : Dato  $i$  predicho por el modelo aleatorio.  $Y_{\text{pred}(i)[e]}^{\text{VC}}$ : Dato  $i$  predicho por la validación cruzada.  $Y_{\text{max obs}[e]}$  y  $Y_{\text{min obs}[e]}$ : Máximo y mínimo valor empírico en el set de entrenamiento, respectivamente,  $\bar{Y}_{\text{sc obs}}$  y  $\bar{Y}_{\text{sc pred}}$ : Valores promedio para  $Y_{\text{sc obs}(i)}$  y  $Y_{\text{sc pred}(i)}$ , respectivamente.

**Tabla E3.** Parámetros de calidad estadística para la predictividad del modelo (Q)SAR.

Ecuaciones asociadas		Requisito	Interpretación
<i>Coefficientes de determinación</i>			
$Q_{F1}^2 = 1 - \frac{\sum_{i=1}^n (Y_{\text{obs}(i)[p]} - Y_{\text{pred}(i)[p]})^2}{\sum_{i=1}^n (Y_{\text{obs}(i)[p]} - \bar{Y}_{\text{obs}[e]})^2}$		$Q_{F1}^2 > 0,5$	El modelo (Q)SAR es predictivo [59]
$Q_{F2}^2 = 1 - \frac{\sum_{i=1}^n (Y_{\text{obs}(i)[p]} - Y_{\text{pred}(i)[p]})^2}{\sum_{i=1}^n (Y_{\text{obs}(i)[p]} - \bar{Y}_{\text{obs}[p]})^2}$		$Q_{F2}^2 > 0,5$	El modelo (Q)SAR es predictivo [118]
<i>Métrica <math>r_m^2</math> y Criterio de Golbraikh y Tropsha</i>			
$Y_{\text{sc obs}(i)} = \frac{Y_{\text{obs}(i)[p]} - Y_{\text{mín obs}[e]}}{Y_{\text{máx obs}[e]} - Y_{\text{mín obs}[e]}}$	$Y_{\text{sc pred}(i)} = \frac{Y_{\text{pred}(i)[p]} - Y_{\text{mín obs}[e]}}{Y_{\text{máx obs}[e]} - Y_{\text{mín obs}[e]}}$	$\bar{r}_{m(\text{prueba})}^2 > 0,5$	
$k_p = \frac{\sum_{i=1}^n Y_{\text{sc obs}(i)} * Y_{\text{sc pred}(i)}}{\sum_{i=1}^n (Y_{\text{sc pred}(i)})^2}$	$k'_p = \frac{\sum_{i=1}^n Y_{\text{sc obs}(i)} * Y_{\text{sc pred}(i)}}{\sum_{i=1}^n (Y_{\text{sc obs}(i)})^2}$	$\Delta r_{m(\text{prueba})}^2 < 0,2$	
$r_{0(p)}^2 = 1 - \frac{\sum_{i=1}^n (Y_{\text{sc obs}(i)} - k_p * Y_{\text{sc pred}(i)})^2}{\sum_{i=1}^n (Y_{\text{sc obs}(i)} - \bar{Y}_{\text{sc obs}})^2}$	$r_{0(p)}'^2 = 1 - \frac{\sum_{i=1}^n (Y_{\text{sc obs}(i)} - k'_p * Y_{\text{sc pred}(i)})^2}{\sum_{i=1}^n (Y_{\text{sc obs}(i)} - \bar{Y}_{\text{sc pred}})^2}$	$r_{\text{prueba}}^2 > 0,6$	
$r_{m(\text{prueba})}^2 = R^2 * \left(1 - \sqrt{(R^2 - r_{0(p)}^2)}\right)$	$r_{m(\text{prueba})}'^2 = R^2 * \left(1 - \sqrt{(R^2 - r_{0(p)}'^2)}\right)$	$\left(\frac{r_{\text{prueba}}^2 - r_{0(p)}^2}{r_{\text{prueba}}^2}\right) < 0,1$	El modelo (Q)SAR es predictivo [115, 119]. Estos parámetros corrigen la falsa predictividad en $Q_{F2}^2$ cuando el intervalo de valores de <i>endpoint</i> es amplio
$\bar{r}_{m(\text{prueba})}^2 = \frac{r_{m(\text{prueba})}^2 + r_{m(\text{prueba})}'^2}{2}$	$\Delta r_{m(\text{prueba})}^2 =  r_{m(\text{prueba})}^2 - r_{m(\text{prueba})}'^2 $	$\left(\frac{r_{\text{prueba}}^2 - r_{0(p)}'^2}{r_{\text{prueba}}^2}\right) < 0,1$	
$r_{\text{prueba}}^2 = \left( \frac{\sum_{i=1}^n (Y_{\text{obs}(i)[p]} - \bar{Y}_{\text{obs}[p]}) (Y_{\text{pred}(i)[p]} - \bar{Y}_{\text{pred}[p]})}{\sqrt{\sum_{i=1}^n (Y_{\text{obs}(i)[p]} - \bar{Y}_{\text{exp[test]})^2 \sum_{i=1}^n (Y_{\text{pred}(i)[p]} - \bar{Y}_{\text{pred}[p]})^2}} \right)^2$		$0,85 \leq k_p \text{ o } k'_p \leq 1,15$	
		$ r_{0(\text{test})}^2 - r_{0(\text{test})}'^2  < 0,3$	

$Y_{\text{obs}(i)[p]}$  y  $Y_{\text{pred}(i)[p]}$ : Dato  $i$  observado y predicho en el set de prueba, respectivamente,  $\bar{Y}_{\text{obs}[e,p]}$ : Promedio de los datos empíricos en el set de prueba,  $\bar{Y}_{\text{pred}[p]}$ : Promedio de los datos predichos en el set de prueba,  $Y_{\text{máx obs}[e]}$  y  $Y_{\text{mín obs}[e]}$ : Máximo y mínimo valor empírico en el set de entrenamiento, respectivamente,  $\bar{Y}_{\text{sc obs}}$  y  $\bar{Y}_{\text{sc pred}}$ : Valores promedio para  $Y_{\text{sc obs}(i)}$  y  $Y_{\text{sc pred}(i)}$ , respectivamente.

**Tabla E4.** Parámetros de calidad estadística para la comparación entre modelos (Q)SAR.

<b>Parámetro</b>	<b>Ecuación asociada</b>	<b>Modelo de mayor calidad</b>
Coefficiente de determinación ajustado ( $R_{adj}^2$ )	$R_{adj}^2 = \frac{(n_e - 1) * R^2 - p}{n_e - 1 - p}$	Mayor $R_{adj}^2$ . Permite comparar modelos con distinto número de descriptores [54]
Función Kubinyi (FIT)	$FIT = \frac{R^2 * (n_e - p - 1)}{(1 - R^2) * (n_e - p^2)}$	Mayor FIT [38]
Criterio de Información de Akaike (CIA)	$CIA = SEC_e * \frac{n_e + p}{(n_e - p)^2}$	Menor CIA [38]

$R^2$ : Coeficiente de determinación del modelo,  $n_e$ : Número de datos en el set de entrenamiento,  $p$ : Número de descriptores del modelo,  $SEC_e$ : Suma de los errores cuadráticos para el set de entrenamiento.

## ANEXO F:

### Generación de descriptores para la interacción herbicida-SDCV

Se analizaron los factores más relevantes que afectan a la adsorción de herbicidas ionizables y no ionizables en SDCV para facilitar la reducción de dimensionalidad e interpretación mecanística de los futuros modelos (Q)SAR.

La información recopilada se presenta en la Tabla F1 y cumple con la fase de planificación expuesta en la Figura 4.

Se incorpora el efecto de los SDCV en  $K_d$  separando el proceso de adsorción en MO, mineral y óxidos de Fe, denotados con los subíndices CO, min y Fe, respectivamente.

$$K_d = \frac{q_{ads}}{C_{sol}} = \frac{q_{ads-CO} + q_{ads-min} + q_{ads-Fe}}{C_{sol}} = \frac{q_{ads-CO}}{C_{sol}} + \frac{q_{ads-min}}{C_{sol}} + \frac{q_{ads-Fe}}{C_{sol}} \quad (F1)$$

Por lo tanto,  $K_d$  puede ser considerada una combinación lineal de tres coeficientes de adsorción:

$$K_d = K_{d(CO)} + K_{d(min)} + K_{d(Fe)} \quad (F2)$$

Y, de manera equivalente:

$$K_f = K_{f(CO)} + K_{f(min)} + K_{f(Fe)} \quad (F3)$$

**Tabla F1.** Principales factores que afectan el proceso de adsorción de herbicidas en SDCV [1, 7, 120].

Factor	Efecto	Consideraciones y posibles mecanismos
<i>Composición del suelo</i>		
CO o MO	Suele tener una correlación positiva con la adsorción, como se ha observado en 2,4 – D, CS, AT y DI. Aunque suele asociarse solo a la forma neutra de los herbicidas, en PC se ha demostrado que a pH 5,9 (PC en forma aniónica) la adsorción ocurre principalmente en la MO, seguida por óxidos de Fe y finalmente arcilla. MC se adsorbe principalmente en la MO, siendo mayor su afinidad por las sustancias húmicas	La adsorción de AT en la MO podría ser disminuida por la acidez superficial de las arcillas mediante repulsión electrostática. Por otro lado, su $K_{oc}$ varía enormemente entre suelos, lo que supone su adsorción en otros componentes del mismo. Para DI (neutro), se ha sugerido la formación de puente hidrógeno con el oxígeno carbonílico o el hidrógeno del grupo amino adyacente
Carga variable	La adsorción de 2,4 – D es favorecida a mayor contenido de Al intercambiable. La adsorción de 2,4 – D en $Fe_{DCB}$ es poco relevante frente a $Fe_{OX}$ . En el caso de PC, los óxidos son los principales adsorbentes solo cuando $MO < 0,3\%$ .	En su forma aniónica, PC es capaz de quelarse a iones metálicos con el nitrógeno piridínico y el grupo carboxilo. Se sugiere también la complejación de cationes polivalentes del suelo ( $Fe^{3+}$ , $Al^{3+}$ , $Cu^{2+}$ , $Zn^{2+}$ , $Ca^{2+}$ ) con coloides orgánicos y PC
Mineral	En muchos casos suele ser poco relevante, como ocurre con 2,4 – D y CS (ácidos). Por otro lado, correlaciona positivamente con la adsorción de AT (base). Por su parte, la adsorción de DI (neutro) correlaciona positivamente con el área superficial solo a bajo contenido de MO (0,1 a 1,7%). La adsorción de MC (neutro) aumenta con el contenido de arcilla y el área superficial, pero está mucho más relacionada con el contenido de MO	La poca relevancia de las arcillas puede asociarse al contenido de CO presente en los suelos donde se realizan los estudios, o en el caso de herbicidas ácidos, a la repulsión electrostática entre la forma aniónica y la superficie de las arcillas. La adsorción del herbicida neutro DI se ha relacionado con la interacción del grupo carbonilo con cationes intercambiables en la superficie de montmorillonitas y grupos Al
<i>Solución</i>		
pH	En herbicidas ionizables, como 2,4 – D, CS y AT, la adsorción correlaciona negativamente con el pH, es decir, se favorece en suelos ácidos. Esto se ha atribuido a fuerzas de van der Waals y grupos SiOH. En herbicidas no ionizables, como DI, el pH no tiene un efecto significativo	En herbicidas ácidos, la correlación negativa entre adsorción y pH se asocia la repulsión electrostática entre la especie aniónica y la superficie del suelo. En herbicidas básicos, como AT, el efecto del pH se ha atribuido a intercambio aniónico y catiónico de la especie catiónica con la superficie negativa de las arcillas
Fuerza iónica	En general el aumento de la fuerza iónica favorece la adsorción de compuestos ionizables, como 2,4 – D y PC. En herbicidas no ionizables no tiene un efecto significativo	En el caso del DI, la baja dependencia de la adsorción con la fuerza iónica y con el tipo de iones presentes en solución ha sido relacionada a su baja polaridad
CIC	Favorece la adsorción de cationes. Suele desfavorecer la adsorción de aniones, aunque a veces no tiene efecto significativo	Se relaciona con la capacidad de adsorción de MO y arcilla

La posible interdependencia entre estos coeficientes es considerada en cada tipo de interacción.

Para representar de mejor manera el sistema herbicida-SDCV, se tienen en cuenta:

- (i) Grado de ionización, representado como fracción neutra ( $\phi_{\text{neu}}$ , equivalente a la definición de la sección 5.2.1) y fracción iónica ( $\phi_-$  o  $\phi_+$  según corresponda, con un valor de  $1 - \phi_{\text{neu}}$ ).
- (ii) Densidad de carga superficial, que da cuenta de la fracción de carga positiva ( $\sigma_+$ ) y negativa ( $\sigma_-$ ) en la superficie de cada componente del suelo. La cantidad de carga es calculada con las ecuaciones 14 y 15, suponiendo que  $\sigma_- + \sigma_+ = 1$  para minerales y MO, mientras que  $\sigma_0 + \sigma_+ = 1$  en óxidos de Fe y contextualizando los valores de  $\text{PIE}_{\text{suelo}}$  a cada componente en SDCV.
- (iii) Fuerza iónica: Constante en todos los casos (soluciones  $\text{CaCl}_2$  0,01 M).

### **Interacción herbicida-MO**

Suponiendo que la adsorción en la MO involucra a la especie neutra ( $K_{f(\text{CO})}^{\text{neu}}$ ) y la iónica ( $K_{f(\text{CO})}^{\text{ion}}$ ), y que ambas no interactúan entre sí, se tiene:

$$K_{f(\text{CO})} = K_{f(\text{CO})}^{\text{neu}} + K_{f(\text{CO})}^{\text{ion}} \quad (\text{F4})$$

Considerando que:

- (i) Superficie de adsorción: La adsorción ocurre en la MO, por lo que aumenta proporcionalmente con el contenido de CO. La cantidad fraccional de CO será representada por  $f_{\text{CO}}$ .

- (ii) Similitud estructural entre herbicidas y fracción orgánica y grado de ionización: La adsorción sería proporcional al coeficiente de partición octanol/agua de cada especie ( $K_{ow}^{neu}$  y  $K_{ow}^{ion}$ ) y la proporción en que estas se encuentran ( $\phi_{neu}$  y  $\phi_{+,-}$ ).
- (iii) Densidad de carga superficial: La adsorción de la especie iónica depende de la interacción electrostática con la superficie de adsorción, que puede presentar carga positiva ( $\sigma_+^{MO}$ ), favoreciendo la adsorción de aniones, o negativa ( $\sigma_-^{MO}$ ), lo que favorece la adsorción de cationes.
- (iv) CIC: Proporcional a la adsorción de cationes e inversamente proporcional a la adsorción de aniones, debido a la relación existente entre CIC y carga superficial negativa del suelo.
- (v) Interacción MO-mineral: Reduce la adsorción de herbicidas no ionizables, tales como DI, debido al bloqueo mutuo de sitios de adsorción [74, 121, 122]. Sin embargo, no se cuenta con información suficiente para cuantificar el bloqueo de sitios de adsorción en este modelo. Por otro lado, la adsorción de herbicidas ionizables en la MO es función de la carga superficial de las arcillas ( $\sigma_-^{min}$  y  $\sigma_+^{min}$ ), debido a sus efectos atractivos/repulsivos.

A partir de lo mencionado anteriormente, la adsorción de herbicidas en forma neutra

( $K_{f(CO)}^{neu}$ ), aniónica ( $K_{f(CO)}^{ion(-)}$ ) y catiónica ( $K_{f(CO)}^{ion(+)}$ ) queda definida como:

$$K_{f(CO)}^{neu} = \alpha_{neu} * K_{ow}^{neu} * \phi_{neu} * f_{CO} \quad (F5)$$

$$K_{f(CO)}^{ion(-)} = \alpha_- * K_{ow}^{ion(-)} * f_{CO} * \phi_- * \frac{\sigma_+^{MO} * (1 - \sigma_-^{min})}{CIC} \quad (F6)$$



$$K_{f(CO)}^{ion(+)} = \alpha_+ * K_{ow}^{ion(+)} * f_{CO} * \phi_+ * CIC * \sigma_{-}^{MO} * (1 - \sigma_{+}^{min}) \quad (F7)$$

Con  $\alpha_{neu}$ ,  $\alpha_-$  y  $\alpha_+$  como factores de proporcionalidad para la adsorción de las diferentes especies herbicidas en CO.

Con base en la especiación anteriormente considerada, estas ecuaciones serán representativas a priori para ácidos o bases monopróticos tales como los herbicidas usados en esta investigación bajo las condiciones experimentales antes mencionadas. No obstante, su validez se verá afectada para los siguientes casos: (i) especies polipróticas, suelos cuya MO posea densidad de carga superficial local variable (alta heterogeneidad), (ii) herbicidas sensibles al tipo de MO, e (iii) interacción MO-mineral significativa producida por bloqueo mutuo de sitios de adsorción.

Analizando  $K_{f(CO)}$  en relación a los herbicidas considerados en este trabajo y a la información proporcionada por la Tabla F1:

- (i) La adsorción de herbicidas no ionizables es independiente del pH ( $\phi_{neu} = 1$  y  $\phi_- = \phi_+ = 0$  para todo pH).
- (ii) La adsorción de herbicidas aniónicos será baja respecto a las especies básicas y neutras, debido a la repulsión electrostática reflejada en la magnitud de  $\sigma_{+}^{min}$  (valor pequeño,  $\sigma_{-}^{min} \sim 1$ ).
- (iii) La CIC afecta al proceso de adsorción de herbicidas ionizables solo en su forma iónica, con efecto diferenciado para ácidos y bases.
- (iv) La adsorción de herbicidas ionizables se produce incluso a disociación completa.
- (v) La dependencia del tipo de MO no es contemplada por las ecuaciones F5, F6 y F7, y podría convertirse en una fuente de error significativa. Tampoco se

considera la menor accesibilidad hacia los sitios de carga positiva en SDCV producto de la repulsión electrostática con la MO. A medida que estos fenómenos sean estudiados, podrían ser incorporados como parte del ciclo PDCA (Figura 4) para promover la perfectibilidad del modelo.

### **Interacción herbicida-mineral**

Se considera una separación por granulometría, tal que:

$$K_{f(\text{min})} = K_{f(\text{arena})} + K_{f(\text{lamo})} + K_{f(\text{arcilla})} \quad (\text{F8})$$

Considerando:

- (i) Superficie de adsorción: La adsorción es proporcional a la superficie de adsorción de cada componente ( $S_{A(\text{arena})}$ ,  $S_{A(\text{lamo})}$ ,  $S_{A(\text{arcilla})}$ ). En herbicidas ionizables también es proporcional a la densidad de carga: positiva ( $\sigma_+^{\text{min}}$ ) para aniones y negativa ( $\sigma_-^{\text{min}}$ ) para cationes.
- (ii) Grado de ionización: La adsorción depende de la proporción en que se encuentra cada especie ( $\phi_{\text{neu}}$  y  $\phi_-$  o  $\phi_+$ ).
- (iii) Disimilitud estructural entre herbicidas y composición mineral para herbicidas neutros: La adsorción depende de la polaridad, y se asume proporcional a  $1/K_{\text{ow}}$ . Para herbicidas ionizables, se asume adsorción únicamente en superficies de carga opuesta a la especie, de modo que se corrige la proporcionalidad a  $\sigma_+^{\text{min}}/K_{\text{ow}}^{\text{ion}(-)}$  para aniones, y  $\sigma_-^{\text{min}}/K_{\text{ow}}^{\text{ion}(+)}$  para cationes.
- (iv) CIC: Proporcional a la adsorción de cationes e inversamente proporcional a la adsorción de aniones (idéntico al sugerido para la MO).

- (v) Interacción MO-mineral y mineral-mineral: En herbicidas no ionizables, la adsorción se favorece a baja densidad de carga superficial [123]. En herbicidas ionizables de determinada carga, la adsorción es desfavorecida al aumentar la magnitud de la densidad de carga opuesta en el suelo.

Con base en lo anterior, se contemplan tres tipos de interacción: herbicida catiónico con superficie de carga negativa, herbicida aniónico con superficie de carga positiva, y herbicida neutro afectado por la carga superficial total. Así, para cada caso se tendrá:

$$K_{f(i)}^{neu} = \beta_{neu}^i * \frac{S_{A(i)} * \phi_{neu} * (1 - |\Delta\sigma|)}{K_{ow}^{neu}} \quad (F9)$$

$$K_{f(i)}^{ion(-)} = \beta_{-,+}^i * \frac{S_{A(i)} * \phi_{-} * (\sigma_{+}^{min})^2}{K_{ow}^{ion(-)} * CIC} \quad (F10)$$

$$K_{f(i)}^{ion(+)} = \beta_{+,-}^i * \frac{S_{A(i)} * \phi_{+} * (\sigma_{-}^{min})^2 * CIC}{K_{ow}^{ion(+)}} \quad (F11)$$

Donde los coeficientes  $\beta$  son factores de proporcionalidad equivalentes a  $\alpha$  para MO.

Considerando que  $(\sigma_{+}^{min} + \sigma_{-}^{min}) = 1$ , el término  $(1 - |\Delta\sigma|)$  puede ser reescrito a:

$$1 - |\Delta\sigma| = 1 - |\sigma_{-}^{min} - \sigma_{+}^{min}| = 1 - |1 - \sigma_{+}^{min} - \sigma_{+}^{min}| = 1 - |1 - 2\sigma_{+}^{min}| \quad (F12)$$

Si se asumen composición mineralógica y afinidad constantes para diferente granulometría, entonces los coeficientes  $\beta$  en sus distintas variantes serán independientes de  $i$ .

Luego, la superficie de adsorción  $S_{A(i)}$  puede ser expresada como:

$$S_{A(i)} = n^{\circ} \text{partículas } i * \text{Área partícula } i \quad (F13)$$

Considerando la cantidad fraccional  $f_i$  de cada componente ( $f_{\text{arena}}$ ,  $f_{\text{limo}}$ ,  $f_{\text{arcilla}}$ ), con definición equivalente a la ecuación E4:

$$n^{\circ}\text{partículas } i = \left( \frac{V_{\text{total } i}}{V_{\text{partícula } i}} \right) = \left( \frac{V_{\text{suelo}} * f_i}{V_{\text{partícula } i}} \right) \quad (\text{F14})$$

Donde  $V_{\text{suelo}}$  corresponde al volumen de la fracción sólida del suelo (partículas).

Suponiendo partículas esféricas representadas por un radio promedio  $\bar{r}_{\text{partícula } i}$ :

$$S_{A(i)} = \left( \frac{V_{\text{suelo}} * f_i}{\frac{4\pi}{3} * \bar{r}_{\text{partícula } i}^3} \right) * (4\pi * \bar{r}_{\text{partícula } i}^2) = 3 * f_i * \left( \frac{V_{\text{suelo}}}{\bar{r}_{\text{partícula } i}} \right) \quad (\text{F15})$$

Otras formas de expresar el volumen son:

$$V_{\text{suelo}} = V_{\text{suelo sa}} (1 - \phi) = \frac{m_{\text{suelo}} (1 - \phi)}{\rho_{\text{bulk}}} = \frac{m_{\text{suelo}}}{\rho_{\text{partícula}}} \quad (\text{F16})$$

Donde  $V_{\text{suelo sa}}$  corresponde al volumen de suelo seco al aire, de porosidad  $\phi$ ,  $m_{\text{suelo}}$  es la masa de suelo seco al aire,  $\rho_{\text{bulk}}$  es la densidad aparente del suelo seco al aire y  $\rho_{\text{partícula}}$  es la densidad real de las partículas, con  $\rho_{\text{partícula}}$  independiente de la granulometría.

Incorporando las ecuaciones F15 y F16 en las ecuaciones F9, F10 y F11, y unificando luego las expresiones asociadas a cada componente (arena, limo, arcilla):

$$K_{f(\text{min})}^{\text{neu}} = 3\beta_{\text{neu}} \left( \frac{\Phi_{\text{neu}}}{K_{\text{ow}}^{\text{neu}}} \right) \left( \frac{m_{\text{suelo}}}{\rho_{\text{partícula}}} \right) \left( \frac{f_{\text{arena}}}{\bar{r}_{\text{arena}}} + \frac{f_{\text{limo}}}{\bar{r}_{\text{limo}}} + \frac{f_{\text{arcilla}}}{\bar{r}_{\text{arcilla}}} \right) (1 - |1 - 2\sigma_+^{\text{min}}|) \quad (\text{F17})$$

$$K_{f(\text{min})}^{\text{ion}(-)} = 3\beta_{-,+} \left( \frac{\phi_- (\sigma_+^{\text{min}})^2}{K_{\text{ow}}^{\text{ion}(-)} \text{CIC}} \right) \left( \frac{m_{\text{suelo}}}{\rho_{\text{partícula}}} \right) \left( \frac{f_{\text{arena}}}{\bar{r}_{\text{arena}}} + \frac{f_{\text{limo}}}{\bar{r}_{\text{limo}}} + \frac{f_{\text{arcilla}}}{\bar{r}_{\text{arcilla}}} \right) \quad (\text{F18})$$

$$K_{f(\text{min})}^{\text{ion}(+)} = 3\beta_{+,-} \left( \frac{\phi_+ (\sigma_-^{\text{min}})^2}{K_{\text{ow}}^{\text{ion}(+)}} \right) \left( \frac{m_{\text{suelo}}}{\rho_{\text{partícula}}} \right) \left( \frac{f_{\text{arena}}}{\bar{r}_{\text{arena}}} + \frac{f_{\text{limo}}}{\bar{r}_{\text{limo}}} + \frac{f_{\text{arcilla}}}{\bar{r}_{\text{arcilla}}} \right) \text{CIC} \quad (\text{F19})$$

Finalmente, analizando  $K_{f(\min)}$  en relación al tipo de herbicida y a la información proporcionada por la Tabla F1:

- (i) En herbicidas no ionizables la adsorción es independiente del pH ( $\phi_{\text{neu}} = 1$  y  $\phi_- = \phi_+ = 0$  para todo pH).
- (ii) En general,  $\sigma_+^{\min} \sim 0$ . Suponiendo que  $\beta_{+,-}$  y  $\beta_{-,+}$  son mayores a  $\beta_{\text{neu}}$  debido a la interacción electrostática asociada, y teniendo en cuenta que  $K_{\text{ow}}^{\text{neu}}$  es mayor que  $K_{\text{ow}}^{\text{ion}(-)}$  y  $K_{\text{ow}}^{\text{ion}(+)}$ , entonces  $K_{f(\min)}^{\text{ion}(-)} \ll K_{f(\min)}^{\text{neu}} < K_{f(\min)}^{\text{ion}(+)}$ , con lo cual a pH ácido se favorece la adsorción de herbicidas ácidos ( $\phi_{\text{neu}} > \phi_-$ ) y básicos ( $\phi_+ > \phi_{\text{neu}}$ ), lo que se corresponde con la interpretación del MBD [30].
- (iii) La correlación positiva entre adsorción y área superficial es explicada por la granulometría, donde  $\bar{r}_{\text{arcilla}} < \bar{r}_{\text{lilimo}} < \bar{r}_{\text{arena}}$ , con lo cual el impacto de  $f_i/\bar{r}_i$  frente a cambios en  $f_i$  será mayor para la arcilla. Esto explicaría además el mayor aporte de las arcillas al proceso de adsorción.
- (iv) Fuentes de error importantes son el supuesto de esfericidad y la ausencia de interacciones específicas entre grupos funcionales de los herbicidas y la superficie de los distintos minerales, por ejemplo complejación, lo que podría generar errores en la predicción.

### **Interacción herbicida-óxidos de Fe**

La constante  $K_{f(\text{Fe})}$  considera el efecto de los óxidos de Fe, asociado a su mayor PIE en comparación a MO y minerales, de modo que su carga superficial podría permanecer positiva o neutra incluso bajo una densidad de carga total negativa, ofreciendo así microambientes para la adsorción de aniones y herbicidas no

ionizables en condiciones desfavorables. En tal caso, los sitios de carga positiva y negativa no necesariamente se neutralizan, sino que pueden coexistir en la superficie heterogénea del suelo [8].

Luego, considerando:

- (i) Superficie de adsorción: La adsorción es proporcional a la cantidad de óxidos de Fe en SDCV. Dicha cantidad será representada por  $f_{Fe}$ .
- (ii) Grado de ionización: La adsorción depende de la proporción en que se encuentra cada especie ( $\phi_{neu}$  y  $\phi_-$  o  $\phi_+$ ).
- (iii) Interacciones posibles: Los óxidos de Fe se consideran microambientes de carga positiva ( $\sigma_+^{Fe}$ ) o neutra ( $\sigma_0^{Fe}$ ), es decir,  $\sigma_0^{Fe} + \sigma_+^{Fe} = 1$ . Luego, la interacción herbicida-óxidos de Fe se supone favorable para todas las especies en  $\sigma_0^{Fe}$ , y solo para aniones y especies neutras en  $\sigma_+^{Fe}$ .
- (iv) Disimilitud estructural entre herbicidas y óxidos de Fe: De forma equivalente a los minerales, la adsorción se asume proporcional a  $1/K_{ow}$ . Luego, considerando la distribución de carga en los óxidos de Fe, el factor de proporcionalidad se corrige a  $\sigma_+^{Fe}/K_{ow}$  en sitios de carga positiva, y  $\sigma_0^{Fe}/K_{ow}$ , o bien,  $(1 - \sigma_+^{Fe})/K_{ow}$  en sitios sin carga.
- (v) Interacción MO-óxidos de Fe y mineral-óxidos de Fe: La presencia de CO y el contenido de arcilla disminuyen la accesibilidad a sitios positivos de los óxidos, debido a su aporte a la magnitud de la carga superficial negativa [8]. Esto se corrige restando la fracción de CO y arcilla a la potencial adsorción en sitios de carga positiva.

Por lo tanto:

$$K_{f(\text{Fe})}^{\text{neu}} = \beta_{\text{neu}}^{\text{Fe}^0} \left( \frac{1 - \sigma_+^{\text{Fe}}}{K_{\text{ow}}^{\text{neu}}} \right) f_{\text{Fe}} \Phi_{\text{neu}} + \beta_{\text{neu}}^{\text{Fe}^+} \left( \frac{\sigma_+^{\text{Fe}} (1 - f_{\text{CO}} - f_{\text{arcilla}})}{K_{\text{ow}}^{\text{neu}}} \right) f_{\text{Fe}} \Phi_{\text{neu}} \quad (\text{F20})$$

$$K_{f(\text{Fe})}^{\text{ion}(-)} = \beta_{\text{neu}}^{\text{Fe}^0} \left( \frac{1 - \sigma_+^{\text{Fe}}}{K_{\text{ow}}^{\text{ion}(-)}} \right) f_{\text{Fe}} \Phi_{-} + \beta_{\text{neu}}^{\text{Fe}^+} \left( \frac{\sigma_+^{\text{Fe}} (1 - f_{\text{CO}} - f_{\text{arcilla}})}{K_{\text{ow}}^{\text{ion}(-)}} \right) f_{\text{Fe}} \Phi_{-} \quad (\text{F21})$$

$$K_{f(\text{Fe})}^{\text{ion}(+)} = \beta_{\text{neu}}^{\text{Fe}^0} \left( \frac{1 - \sigma_+^{\text{Fe}}}{K_{\text{ow}}^{\text{ion}(+)}} \right) f_{\text{Fe}} \Phi_{+} \quad (\text{F22})$$

Donde  $\beta_j^i$  son factores de proporcionalidad para la adsorción de herbicidas en óxidos de Fe, con  $i = \text{Fe}^0$  para representar la superficie sin carga en los óxidos de Fe,  $i = \text{Fe}^+$  para la superficie cargada positivamente, y  $j$  para representar la especiación del herbicida, es decir, no ionizable (neu), anión (-) o catión (+).

Notar que, si bien el impacto de MO y minerales a la adsorción en óxidos de Fe se relaciona con la diferencia de magnitud entre las cargas positiva y negativa en SDCV, la cuantificación del término positivo utiliza  $\sigma_+^{\text{Fe}}$ , que corresponde a la fracción de carga y no da cuenta de la magnitud de la misma.

Con lo anterior presente, se propone  $(1 - f_{\text{CO}} - f_{\text{arcilla}})$  como factor de corrección, en vez de un término asociado a la interacción electrostática del herbicida con CO y minerales, debido a la falta de información para cuantificar esto último, lo que supone incerteza y posibles errores en la predicción.

Adicionalmente, la ecuación propuesta no considera la formación de complejos en la superficie de los óxidos de Fe, debido a la especificidad de sus mecanismos asociados. Esto podría ser importante al momento de explicar desviaciones.

Con base en la ecuación propuesta, los escenarios de mayor similitud a SCP son:

- (i) Herbicidas no ionizables con baja afinidad por los óxidos de Fe, tal que  $K_{f(Fe)}^{neu}$  sea poco significativos frente a  $K_{f(CO)}^{neu}$  y  $K_{f(min)}^{neu}$ . Esto será función de  $f_{CO}$ ,  $f_{Fe}$  y  $\left(\frac{f_{arena}}{\bar{r}_{arena}} + \frac{f_{limo}}{\bar{r}_{limo}} + \frac{f_{arcilla}}{\bar{r}_{arcilla}}\right)$ .
- (ii) Herbicidas ácidos completamente ionizados ( $\phi_- = 1$ ) en suelos con  $pH > PIE_{Fe}$  ( $\sigma_+^{Fe} \rightarrow 0$ ), bajo el supuesto de que los aniones son adsorbidos en menor intensidad por óxidos de Fe sin carga superficial, con lo cual  $K_{f(Fe)}$  alcanza un valor mínimo.
- (iii) Herbicidas básicos en condiciones de  $pH \leq pKa - 2 \leq PIE_{Fe}$ , de modo que  $K_{f(Fe)} \sim 0$  debido a la repulsión electrostática entre el herbicida catiónico y la carga positiva de los óxidos de Fe ( $\sigma_+^{Fe} = 1$ ).



## **ANEXO G:**

### **Calidad estadística de los modelos (Q)SAR generados**

**Tabla G1.** Calidad estadística de los modelos (Q)SAR para predecir  $\log K_f$  o  $K_f$  de herbicidas no ionizables.

	Ecuación 26 <sup>a</sup> ( $\log K_f$ )		Ecuación 27 <sup>a</sup> ( $\log K_f$ )		Ecuación 28 <sup>b</sup> ( $K_f$ )	
<b>Bondad del ajuste</b>	$n_e = 26$	$\rho = 8,7$	$n_e = 26$	$\rho = 6,5$	$n_e = 26$	$\rho = 6,5$
	$R^2 = 0,7570$	$F^{3,22} = 22,8475$	$R^2 = 0,7968$	$F^{4,21} = 20,5896$	$R^2 = 0,7012$	$F^{4,21} = 11,7358$
	$SEC_e = 0,9896$	$EE_e = 0,2121$	$SEC_e = 0,8275$	$EE_e = 0,19851$	$SEC_e = 4419,043$	$EE_e = 14,8645$
	$EEN = 0,156$	$EA \ll EMA$	$EEN = 0,146$	$EA \ll EMA$	$EEN = 0,149$	$EA \ll EMA$
	$FIV(\log f_{arcilla}) = 1,0707$		$FIV(\log f_{arcilla}) = 1,1491$		$FIV(K_{f(CO)}) = 1,4004$	
	$FIV(\log f_{Fe\ DCB}) = 1,0064$		$FIV(\log f_{Fe\ PYRO}) = 1,1194$		$FIV(K_{f(min)}) = 1,9426$	
	$FIV(Weta2. unity) = 1,0119$		$FIV(\log f_{Fe\ DCB}) = 1,1614$		$FIV(K_{f(Fe^0)}) = 1,2303$	
		$FIV(Weta2. unity) = 1,0126$		$FIV(K_{f(Fe^+)}) = 2,3534$		
<b>Robustez</b>	$cR_p^2 = 0,7200$	$Q_{LOO}^2 = 0,6666$	$cR_p^2 = 0,7308$	$Q_{LOO}^2 = 0,6966$	$cR_p^2 = 0,6148$	$Q_{LOO}^2 = 0,2957$
	$\bar{r}_{m(LOO)}^2 = 0,5618$	$\Delta r_{m(LOO)}^2 = 0,1508$	$\bar{r}_{m(LOO)}^2 = 0,6004$	$\Delta r_{m(LOO)}^2 = 0,1244$	$\bar{r}_{m(LOO)}^2 = 0,4607$	$\Delta r_{m(LOO)}^2 = 0,0506$
<b>Predictividad</b>	$Q_{F1}^2 = 0,7258$	$Q_{F2}^2 = 0,6012$	$Q_{F1}^2 = 0,65498$	$Q_{F2}^2 = 0,62406$	$Q_{F1}^2 = -0,3011$	$Q_{F2}^2 = -0,3895$
	$r_{prueba}^2 = 0,7731$	$\bar{r}_{m(prueba)}^2 = 0,7070$	$r_{prueba}^2 = 0,71835$	$\bar{r}_{m(prueba)}^2 = 0,6407$	$r_{prueba}^2 = 0,2861$	$\bar{r}_{m(prueba)}^2 = 0,1815$
		$\Delta r_{m(prueba)}^2 = 0,0741$		$\Delta r_{m(prueba)}^2 = 0,1286$		$\Delta r_{m(prueba)}^2 = 0,1299$
	Aprueba los criterios de Golbraikh y Tropsha		Aprueba los criterios de Golbraikh y Tropsha		No aprueba los criterios de Golbraikh y Tropsha	
<b>Parámetros de comparación</b>	$R_{adj}^2 = 0,7238$	$FIT = 4,0319$	$R_{adj}^2 = 0,7581$	$FIT = 8,2357$	No aplica <sup>c</sup>	
	$CIA = 0,0543$		$CIA = 0,0513$			

<sup>a</sup> Ecuación generada a partir de la mejor combinación de descriptores.

<sup>b</sup> Ecuación generada usando descriptores basados en la interacción herbicida-SDCV.

<sup>c</sup> No se consideran, pues el *endpoint* en esta ecuación no presenta la misma escala respecto a las ecuaciones con las que debe ser comparada.

**Tabla G2.** Calidad estadística de los modelos (Q)SAR para predecir log K<sub>f</sub> o K<sub>f</sub> de herbicidas ácidos.

	Ecuación 31 <sup>a</sup> (log K <sub>f</sub> )		Ecuación 32 <sup>a</sup> (log K <sub>f</sub> )		Ecuación 34 <sup>b</sup> (K <sub>f</sub> )	
<b>Bondad del ajuste</b>	n <sub>e</sub> = 43	ρ = 10,75	n <sub>e</sub> = 43	ρ = 10,75	n <sub>e</sub> = 43	ρ = 21,5
	R <sup>2</sup> = 0,5353	F <sup>4,38</sup> = 10,9428	R <sup>2</sup> = 0,5340	F <sup>4,38</sup> = 10,8844	R <sup>2</sup> = 0,5532	F <sup>2,40</sup> = 24,7579
	SEC <sub>e</sub> = 2,5881	EE <sub>e</sub> = 0,2610	SEC <sub>e</sub> = 2,5955	EE <sub>e</sub> = 0,2614	SEC <sub>e</sub> = 1649,4242	EE <sub>e</sub> = 6,4215
	EEN = 0,156	EA << EMA	EEN = 0,146	EA << EMA	EEN = 0,126	EA << EMA
	FIV(log f <sub>arcilla</sub> ) = 2,3580		FIV(log f <sub>CO</sub> ) = 4,3339		FIV(K <sub>f(CO)</sub> ) = 1,0094	
	FIV(CIC) = 2,8801		FIV(CIC) = 2,0468		FIV(K <sub>f(Fe+)</sub> ) = 1,0094	
	FIV(log f <sub>Fe OX</sub> ) = 1,5680		FIV(log f <sub>Fe DCB</sub> ) = 4,9254			
	FIV(nAtomLC) = 1,0318		FIV(nAtomLC) = 1,0402			
<b>Robustez</b>	cR <sub>p</sub> <sup>2</sup> = 0,4897	Q <sub>LOO</sub> <sup>2</sup> = 0,4269	cR <sub>p</sub> <sup>2</sup> = 0,4861	Q <sub>LOO</sub> <sup>2</sup> = 0,4291	cR <sub>p</sub> <sup>2</sup> = 0,5322	Q <sub>LOO</sub> <sup>2</sup> = 0,4229
	$\bar{r}_{m(LOO)}^2 = 0,2752$	$\Delta r_{m(LOO)}^2 = 0,2583$	$\bar{r}_{m(LOO)}^2 = 0,2774$	$\Delta r_{m(LOO)}^2 = 0,2574$	$\bar{r}_{m(LOO)}^2 = 0,2754$	$\Delta r_{m(LOO)}^2 = 0,2741$
<b>Predictividad</b>	Q <sub>F1</sub> <sup>2</sup> = 0,1228	Q <sub>F2</sub> <sup>2</sup> = 0,1226	Q <sub>F1</sub> <sup>2</sup> = 0,3430	Q <sub>F2</sub> <sup>2</sup> = 0,3428	Q <sub>F1</sub> <sup>2</sup> = 0,0036	Q <sub>F2</sub> <sup>2</sup> = -0,0582
	r <sub>prueba</sub> <sup>2</sup> = 0,1656	$\bar{r}_{m(prueba)}^2 = -0,1352$	r <sub>prueba</sub> <sup>2</sup> = 0,5795	$\bar{r}_{m(prueba)}^2 = -0,1824$	r <sub>prueba</sub> <sup>2</sup> = 0,0505	$\bar{r}_{m(prueba)}^2 = 0,0101$
	$\Delta r_{m(prueba)}^2 = 0,5409$		$\Delta r_{m(prueba)}^2 = 0,9973$		$\Delta r_{m(prueba)}^2 = 0,0563$	
	No aprueba los criterios de Golbraikh y Tropsha		No aprueba los criterios de Golbraikh y Tropsha		No aprueba los criterios de Golbraikh y Tropsha	
<b>Parámetros de comparación</b>	R <sub>adj</sub> <sup>2</sup> = 0,4864	FIT = 1,6212	R <sub>adj</sub> <sup>2</sup> = 0,4849	FIT = 1,6125	No aplica <sup>c</sup>	
	CIA = 0,0800		CIA = 0,0802			

<sup>a</sup> Ecuación generada a partir de la mejor combinación de descriptores.

<sup>b</sup> Ecuación generada usando descriptores basados en la interacción herbicida-SDCV.

<sup>c</sup> No se consideran, pues el *endpoint* en esta ecuación no presenta la misma escala respecto a las ecuaciones con las que debe ser comparada.

**Tabla G3.** Calidad estadística de los modelos (Q)SAR para predecir n de herbicidas no ionizables.

	Ecuación 36 <sup>a</sup> (log K <sub>f</sub> )		Ecuación 37 <sup>a</sup> (log K <sub>f</sub> )	
<b>Bondad del ajuste</b>	n <sub>e</sub> = 26	ρ = 8,7	n <sub>e</sub> = 26	ρ = 6,5
	R <sup>2</sup> = 0,6008	F <sup>3,22</sup> = 11,0376	R <sup>2</sup> = 0,6041	F <sup>4,21</sup> = 8,0102
	SEC <sub>e</sub> = 1,3192	EE <sub>e</sub> = 0,2449	SEC <sub>e</sub> = 1,3084	EE <sub>e</sub> = 0,2496
	EEN = 0,189	EA << EMA	EEN = 0,193	EA << EMA
	FIV(log f <sub>arena</sub> ) = 1,0012		FIV(log f <sub>limo</sub> ) = 1,5366	
	FIV(log f <sub>arcilla</sub> ) = 1,0159		FIV(log f <sub>arcilla</sub> ) = 2,3438	
	FIV(SC - 5) = 1,0170		FIV(CIC) = 2,3764	
			FIV(SC - 5) = 1,0867	
<b>Robustez</b>	cR <sub>p</sub> <sup>2</sup> = 0,5464	Q <sub>LOO</sub> <sup>2</sup> = 0,4856	cR <sub>p</sub> <sup>2</sup> = 0,5173	Q <sub>LOO</sub> <sup>2</sup> = 0,4778
	r̄ <sub>m(LOO)</sub> <sup>2</sup> = 0,3657	Δr <sub>m(LOO)</sub> <sup>2</sup> = 0,1847	r̄ <sub>m(LOO)</sub> <sup>2</sup> = 0,3624	Δr <sub>m(LOO)</sub> <sup>2</sup> = 0,1674
<b>Predictividad</b>	Q <sub>F1</sub> <sup>2</sup> = 0,8654	Q <sub>F2</sub> <sup>2</sup> = 0,8628	Q <sub>F1</sub> <sup>2</sup> = 0,9083	Q <sub>F2</sub> <sup>2</sup> = 0,9065
	r <sub>prueba</sub> <sup>2</sup> = 0,8881	r̄ <sub>m(LOO)</sub> <sup>2</sup> = 0,6885	r <sub>prueba</sub> <sup>2</sup> = 0,9232	r̄ <sub>m(prueba)</sub> <sup>2</sup> = 0,7835
	Δr <sub>m(prueba)</sub> <sup>2</sup> = 0,1176		Δr <sub>m(prueba)</sub> <sup>2</sup> = 0,0734	
	Aprueba los criterios de Golbraikh y Tropsha		Aprueba los criterios de Golbraikh y Tropsha	
<b>Parámetros de comparación</b>	R <sub>adj</sub> <sup>2</sup> = 0,5464	FIT = 2,3651	R <sub>adj</sub> <sup>2</sup> = 0,5287	FIT = 1,7837
	CIA = 0,0511		CIA = 0,0472	

<sup>a</sup> Ecuación generada a partir de la mejor combinación de descriptores.

<sup>b</sup> Ecuación generada usando descriptores basados en la interacción herbicida-SDCV.