Deconstructing Adsorption Variability: Investigating Controls on Pesticide Adsorption in Soil and Modelling Koc

CHRISTIE, EMMA,LOUISE

How to cite:

CHRISTIE, EMMA,LOUISE (2013) *Deconstructing Adsorption Variability: Investigating Controls on Pesticide Adsorption in Soil and Modelling Koc*, Durham theses, Durham University. Available at Durham E-Theses Online: http://etheses.dur.ac.uk/7347/

Deconstructing Adsorption Variability:

Investigating Controls on Pesticide Adsorption in Soil and Modelling K_{oc}

Emma Louise Christie

Department of Earth Sciences

Durham University

A thesis submitted in partial fulfilment of the requirements of the Council of the University of Durham for Degree of Doctor of Philosophy (PhD)

2012

ABSTRACT

Due to potential environmental risks of pesticides, it is important that the fate of pesticides is known and that safer pesticides are developed in the future. This thesis focused on identifying controls on the K_{oc} of pesticides in soil based on their structural parameters. This thesis also developed quantitative structure-activity relationships (QSAR) models to predict the environmental fate of new pesticides.

To understand the controls on K_{oc}, a range of multivariate statistical techniques were used including; principal component analysis, and analysis of variance. Predictive models were created using logistic regression, and multiple linear regression.

The study found adsorption of pesticides in soil is controlled by a combination of size and solubility parameters. Logistic regression models were able to predict the adsorption potential of metabolites, relative to their parent based on metabolite structures. This study found that adsorption behaviour of pesticides was fairly specific to different chemical groups. A QSAR model for K_{oc} was constructed for a group of early stage compounds and could predict K_{oc} to just over an order of magnitude.

The results of this study have implications for the pesticide development process. If developed further to include a wider range of chemical groups then the models have the potential to reduce the dependence on laboratory tests in the early stages of the development process. However, this study also questions the use of K_{oc} as a predictive parameter and offers alternative solutions to predicting environmental fate of pesticides.

DECLARATION AND COPYRIGHT

I confirm that no part of the material presented in this thesis has previously been submitted by me or any other person for a degree in this or any other university. In all cases material from the work of others has been acknowledged.

ACKNOWLEDGEMENTS

This thesis was funded by a BBSRC CASE award in collaboration with Syngenta UK.

Firstly, my thanks go to my supervisor Fred Worrall. I have learnt so much from Fred and he has been incredibly supportive throughout my PhD, offering encouragement from the very start, I literally couldn't have done this without you. Equal thanks go to my industry supervisor Andrew Eatherall. Thank you for offering a helpful alternative perspective on my thesis and for keeping an eye on me during my visits to Jealott's Hill. Thanks to Sui Kuet from Syngenta for her help and eternal patience when it came to my lab work. Finally thanks to those, both in Earth Sciences and Syngenta, whose advice I have sought and help I have received over the last 4 years.

I want to acknowledge my Van Mildert friends, old and new, for motivating me with fun times throughout my (many) years in Durham, and for helping in their own little ways: Vicky Brown (chief map maker), Gary Elphick and Martin Brader (the IT crowd). Also, although they are no longer in Durham, they are still equally as important to me: Rachel Hope (journal retriever extraordinaire), and Amanda Levinson, thank you for letting me take advantage of your generous hospitality when it came to conferences!

On a personal note, thank you to my Dad for looking after me in some difficult years and especially in this final one. Hopefully it has been worth it and I have made you proud. Finally, I dedicate this thesis to my Mum.

TABLE OF CONTENTS

Abstract ii
Declaration and Copyright iii
Acknowledgements iv
Contents v
List of Figures viii
List of Tables x

1. Introduction **1**
1.1 Pesticides; their benefits and risks *1*
1.2 Pesticides in the Environment *4*
1.2.1 Pesticides in Soil 6
1.2.2 Pesticide Chemistry 11
1.3 Predicting Environmental Fate of Pesticides *14*
1.4 Study Outline *15*

2. Methodology **17**
2.1 Data Collection *17*
2.1.1 Constructing the Dataset 22
2.1.2 Justification of Parameters 25
2.2 Data Analysis *27*
2.2.1 Principal Component Analysis (PCA) 28
2.2.2 Multiple Regression 29
2.2.3 Logistic Regression 32
2.2.4 Analysis of Variance (ANOVA) 34
2.3 Experimental Methodology *34*
2.3.1 Selection of Soil Properties 35
2.3.2 Batch Equilibrium Adsorption Study 35

3. Preliminary Results – All Data **40**
3.1 Introduction *40*
3.2 Study Approach *40*
3.3 Results *40*
3.3.1 PCA: Connectivity Parameters and Molecular Fragments 40
3.3.2 PCA: Soil Properties 43
3.3.3 PCA: Molecular Properties 45
3.4 Analysis by Chemical Group *47*
3.4.1 Connectivity and Molecular Fragments 48
3.4.2 Soil Properties 51
3.4.3 Molecular Properties 54
3.5 Group A Compounds *57*
3.5.1 PCA: Connectivity Parameters and Molecular Fragments 58
3.5.2 PCA: Molecular Properties 61
3.5.3 Multiple Regression 63

3.5.4 Logistic Regression 64
3.6 Summary 66

4. Parent/Metabolite Study 68
4.1 Introduction 68
4.2 Study Approach 69
4.3 Glyphosate and AMPA Adsorption Study 69
4.3.1 Soil Choice 69
4.3.2 Experimental Study Results 71
4.4 Parent and Metabolite Analysis: Methodology 72
4.5 Parent and Metabolite Analysis: Results 73
4.5.1 PCA: Connectivity Parameters and Molecular Fragments 73
4.5.2 PCA: Molecular Properties 77
4.5.3 PCA: Soil Properties 80
4.6 Modelling Parent and Metabolite Compounds 81
4.6.1 Multiple Regression Model 81
4.6.2 Logistic Regression 82
4.7 Summary 85

5. Soil Study 87
5.1 Introduction 87
5.2 Study Approach 87
5.3 Fieldwork: Collecting Soil Samples 87
5.4 Soil Characteristics 89
5.5 Hexaconazole Study 90
5.5.1 Adsorption Study Methodology 91
5.5.2 Hexaconazole Experimental Study Results 96
5.6 Hexaconazole Analysis: Methodology 97
5.7 Hexaconazole Analysis: Results 97
5.7.1 PCA: Soil Properties 97
5.7.2: Multiple Linear Regression: Modelling Soil Properties 99
5.7.3: Analysis of Variance (ANOVA) 101
5.8 Summary 101

6. Stage 1 Compound Study 103
6.1 Introduction 103
6.2 Stage 1 Data 104
6.2.1 Stage 1 Reports 104
6.3 Stage 1 Data Analysis 105
6.3.1 PCA: Connectivity Parameters and Molecular Fragments 105
6.3.2 PCA: Molecular Properties 108
6.3.3 PCA: PCA Soil Properties 110
6.4 Multiple Regression Model 113
6.5 Summary 114
6.6 Further Thought 115

7. Validation and Discussion **116**
7.1 Introduction *116*
7.2 Stage 1 Model Validation *116*
7.2.1 Stage 1 Model Discussion 119
7.3 Comparison of Stage 1 Model against EPA Model *119*
7.4 Suitability of K_{oc} Models *122*
7.4.1 Is it Possible to Predict K_{oc}? 123
7.5 Summary *124*

8. Conclusion **125**

Appendix A: How to Make your Model **127**

References **132**

Appendix B: (Disc)
Folder 1 – Batch Equilibrium Methodology
Subfolder – 'AMPA' Containing:
AMPA 24hr Bq Count
AMPA Adsorption Spreadsheet
AMPA Calculations
AMPA Treatment Solution Spreadsheet

Subfolder – 'Glyphosate' Containing:
Glyphosate 24hr Bq Count
Glyphosate Adsorption Spreadsheet
Glyphosate Calculations
Glyphosate Treatment Solution Spreadsheet

Subfolder – 'Hexaconazole' Containing:
Hexaconazole 24hr Bq Count
Hexaconazole Adsorption Spreadsheet
Hexaconazole Calculations
Hexaconazole Treatment Solution Spreadsheet

Folder 2 – HPLC Results

FIGURES

Figure 1.1 Pathways of a pesticide once applied to crops **4**
Figure 1.2a 2:1 or TOT clay structure **8**
Figure 1.2b Representation of the process of cation exchange **8**
Figure 1.3 Common functional groups featured in the pesticides analysed in this study **11**
Figure 1.4 Carbamate pesticide with the ethyl carbamate functional group circled **14**

Figure 2.1 Structures of the three compounds used in the batch equilibrium experiments **24**

Figure 3.1 Scatter plot of principal component 4 versus principal component 5 for connectivity parameters and molecular fragments **43**
Figure 3.2 Scatter plot of principal component 3 versus principal component 4 for soil properties **45**
Figure 3.3 Scatter plot of principal component 4 versus principal component 5 for molecular properties **47**
Figure 3.4 Scores from PCA plotted by chemical group for connectivity parameters and molecular fragments **49**
Figure 3.5 Scores from PCA plotted by chemical group for connectivity parameters and molecular fragments, simplified into two groups **50**
Figure 3.6 Scores from PCA plotted by chemical group for soil properties **52**
Figure 3.7 Scores from PCA plotted by chemical group for soil properties, simplified into two groups **53**
Figure 3.8 Scores from PCA plotted by chemical group for molecular properties **55**
Figure 3.9 Scores from PCA plotted by chemical group for molecular properties, simplified into two groups **56**
Figure 3.10 Scores from PCA plotted by compound for connectivity parameters and molecular fragments **60**
Figure 3.11 Scores from PCA plotted by compound for molecular properties **62**

Figure 4.1 Scatter plot of principal component 3 versus principal component 5 for all parent and metabolite data, analysed for connectivity parameters and molecular fragments **75**
Figure 4.2 Scores from PCA analysed for connectivity parameters and molecular fragments, plotted by compound and by parent and metabolite **76**
Figure 4.3 Scatter plot of principal component 2 versus principal component 3 for all parent and metabolite data, analysed for molecular properties **78**
Figure 4.4 Scores from PCA analysed for molecular properties, plotted by compound and by parent and metabolite **79**
Figure 4.5 Scatter plot of principal component 3 versus principal component 4 for all parent and metabolite data, analysed for soil properties **81**
Figure 4.6 Probability plot of Equation 7, showing metabolites that are less adsorbing and more adsorbing than their parents **84**

Figure 5.1 Locations of sites for new soil samples **88**
Figure 5.2 Scan of developed TLC plate after one week exposure **95**

Figure 5.3 Scatter plot of principal component 2 versus principal component 4 for Hexaconazole, analysed for soil properties **98**

Figure 6.1 General anonymised structure for Stage 1 benzazole compounds used in this study **104**
Figure 6.2 Scatter plot of principal component 3 versus principal component 5 for Stage 1 compounds, analysed by connectivity parameters and molecular fragments **107**
Figure 6.3 Scatter plot of principal component 3 versus principal component 4 for Stage 1 compounds, analysed by molecular properties **109**
Figure 6.4 Scatter plot of principal component 1 versus principal component 3 for Stage 1 compounds, analysed by soil properties **111**

Figure 7.1 Predicted Log K_{oc} versus Observed Log K_{oc} for Equation 10 **117**
Figure 7.2 Recalculated predicted Log K_{oc} versus Observed Log K_{oc} for Equation 10, having removed data for Compound 10 **118**
Figure 7.3 EPA predicted Log K_{oc} values, compared to observed Log K_{oc} values and the Log K_{oc} values predicted by the Stage 1 model **120**

TABLES

Table 1.1 Average total potential losses and total actual losses to six major crop types worldwide in 2001 – 2003 **1**

Table 2.1 Structures and chemical classes of the most commonly discussed pesticides in this study **17**
Table 2.2 Descriptions of the molecular (semi-empirical and quantum) properties used in the data analysis **23**

Table 3.1 The first five principal components for connectivity parameters and molecular fragments with the cumulative proportion of variance explained **40**
Table 3.2 The first four principal components for soil properties with the cumulative proportion of variance explained **43**
Table 3.3 The first five principal components for molecular properties with the cumulative proportion of variance explained **45**
Table 3.4 The first five principal components for Group A compounds, for connectivity parameters and molecular fragments **58**
Table 3.5 The first four principal components for Group A compounds, for molecular properties **61**
Table 3.6 The significant parameters of MLR model for predicting K_{oc} of Group A compounds, the regression coefficient and R^2 value of the model **64**
Table 3.7 The significant parameters of the Group A versus Group B model **65**

Table 4.1 List of soils tested and their characterisation **70**
Table 4.2 Results for 24 hour adsorption step, by soil type for Glyphosate **71**
Table 4.3 Results for 24 hour adsorption step, by soil type for AMPA **72**
Table 4.4 The first five principal components for Parent and Metabolite compounds, analysed for connectivity parameters and molecular fragments **73**
Table 4.5 The first five principal components for Parent and Metabolite compounds, analysed for molecular properties **77**
Table 4.6 The first four principal components for Parent and Metabolite Compounds, analysed for soil properties **80**
Table 4.7 The significant properties of MLR model for predicting K_{oc} of parent and metabolite compounds, the regression coefficients and R^2 value for the model **82**
Table 4.8 The significant parameters of model of model classifying parent and metabolite compounds **83**
Table 4.9 The significant parameters of the adsorption potential model **84**

Table 5.1 Geology, soil type, and land use of locations taken for soil sampling **89**
Table 5.2 Soil characterisation, as determined by NRM Laboratories Ltd **90**
Table 5.3 HPLC conditions used for analysis of Hexaconazole aqueous phase **92**
Table 5.4 Soils selected for TLC analysis **93**
Table 5.5 K_{oc} values for Hexaconazole after 72 hour adsorption **96**

Table 5.6 The first four principal components for Hexaconazole, analysed for soil properties **97**
Table 5.7 The significant parameters of MLR model for predicting K_{oc} of Hexaconazole based on soil properties, and the regression coefficients and R^2 value for the model **99**
Table 5.8 ANOVA results testing for significance of soils and compounds on K_{oc} **101**

Table 6.1 Soil Properties for Benazoles Stage 1 Study **105**
Table 6.2 First five principal components for Stage 1 compounds, tested for connectivity parameters and molecular fragments **105**
Table 6.3 First five principal components for Stage 1 compounds, tested for molecular properties **108**
Table 6.4 First five principal components for Stage 1 compounds, tested for soil properties **110**
Table 6.5 The significant parameters of MLR model for predicting K_{oc} of Stage 1 compounds, and the regression coefficients and R^2 of the model **113**

Table 7.1 Compounds from validation set and their K_{oc} values and percentage differences between predicted and observed values **121**
Table 7.2 Calculated root mean average percentage errors for Hexaconazole and other compounds from the azoles chemical group **123**

1. INTRODUCTION

1.1 Pesticides; their benefits, and risks

There are many opinions on the use of pesticides so it makes sense to review some of these opinions and issues to put this study into context. The term 'pesticide' is a broad non-specific term that covers a large number of substances including; insecticides, herbicides and fungicides (Wilson and Tisdell, 2001). It is estimated that 2.5 million metric tons of pesticides are applied each year worldwide (van der Werf, 1996). The main advantage of pesticides or crop protection products is ensuring food security for a growing population. Oerke and Dehne, (2004) stated that human population is projected to increase to 7.7 billion by 2020. The increased population density, combined with changes in dietary habits of developing countries and increased use of grains for livestock feed is projected to cause the demand for grain production to more than double (Oerke and Dehne, 2004). An increasing population will mean that there is competition for land. Land suitable for agricultural production is limited and most soils with high productivity potential are already under cultivation (Oerke and Dehne, 2004).

Over the last 60 years there have been changes in the way farmers and growers produce food, to meet the demands of consumers, governments, food processors and retailers, which have included the extensive use of pesticides (Cooper and Dobson, 2007). It is estimated that each dollar invested in pesticide control returns approximately four dollars in crops saved (Pimentel *et al*., 1992). The use of pesticides is to prevent or reduce agricultural losses to pests, which results in improved yield and greater availability of food, at a reasonable price all year round (Cooper and Dobson, 2007). Oerke (2006) estimated the global total potential and actual losses of crops to pests in wheat, rice, maize, potatoes, soybeans and cotton for the period 2001-2003 (Table 1.1).

Crop	Total Potential Loss	Total Actual Loss
Wheat	49.8%	28.20%
Rice	77.0%	37.40%
Maize	68.5%	31.20%
Potatoes	74.9%	40.30%
Soybeans	60.0%	26.30%
Cotton	82.0%	28.80%

Table 1.1 Total average potential and total losses to six major crop types worldwide in 2001-2003 (Oerke, 2006). Potential loss is the loss of crops without physical, biological or chemical crop protection. The actual loss is the crop loss occurring despite crop protection practices. Total losses include loss from weeds, animal pests, pathogens and viruses.

Although crop protection products do not eliminate all crop losses (Pimentel *et al*., 1992), using the work of Oerke, (2006) it can be calculated that there was a 54% reduction from potential to actual losses across the six major crop types.

As well as protecting the existing major crop types from losses, the use of pesticides can broaden the range of viable crop options that farmers can grow at particular times of the year. Cooper and Dobson, (2007) used the example of tomatoes that can be grown in the Zimbabwean rainy season by using fungicides to protect against late blight. By extending the growing season and range of crops that can be grown it is financially beneficial for farmers, particularly for countries that export to the USA and Europe. Similarly, consumers in developed countries gain from the wider range of imported crops that are available for a greater proportion of the year (Cooper and Dobson, 2007). Additional benefits of pesticides are observed when compared to alternative control methods. For example, Edwards-Jones, (2008) suggested that when compared to other weed control methods, using herbicides is more beneficial as it reduces the labour associated with hand weeding. Pesticides also have the potential to control human diseases such as malaria and fevers. Cooper and Dobson, (2007) review some of the studies of malaria incidence and bed nets treated with an insecticide and suggested that using treated bed nets reduced the number of infective bites by 75%.

There is much work in the literature that highlights the negative effects and risks of pesticides. The largest volume of pesticide use is in developed countries, although use is growing in developing countries (Wilson and Tisdell, 2008). Many problems relating to pesticide use and particularly pesticide poisoning are found in developing countries. Products that are under regulatory control or even banned in places such as Europe are still available in developing countries (Matthews, 2008). Tariq *et al*., (2007) reported that in Pakistan the chances of human exposure to pesticides are relatively high due to low awareness of the safe use of pesticides and low literacy rates.

Environmental risks of pesticides include poisoning of non-target organisms, particularly mammals. Organisms may take up pesticides through ingestion of contaminated food and water, respiration or direct contact (van der Werf, 1996). If the pesticides are slow to metabolise, or are fat soluble there is the potential for the compound to move up the food chain in a process call biomagnification. Also of concern is the effect of pesticides in the soil on microorganisms and invertebrates. Organisms in the soil such as earthworms, bacteria, fungi, *etc,* are vital to ecosystems because they dominate the structure and function of natural systems (Pimentel *et al*., 1992). Earthworms in particular contribute to soil fertility and are an important part of terrestrial food webs (van der Werf, 1996).

The main pathway of earthworm exposure to pesticides is via contaminated soil pore water, with damage often being observed when there is heavy rainfall after application (van der Werf, 1996). Along with many other processes, microorganisms play an important role in the soil due to their ability to fix nitrogen, making it available for plants (Pimentel *et al*., 1992).

The negative effects of pesticides that are most relevant to this study are the risk of pesticide pollution to groundwater and to a certain extent, surface waters. Owing to the environmental and human risks, there has been a change from persistent organochloride pesticides to less persistent pesticides that are more rapidly removed from the environment. However, the newer compounds are generally more water soluble and create the risk of contaminating ground and surface water (Ward and Robinson, 2000). Pesticides usually enter lakes and streams by water runoff and soil erosion (Wilson and Tisdell, 2001). Once in the aquatic system, pesticides can cause fish losses through high concentrations that directly kill fish, lower level doses that eliminate essential fish food like insects, or by reducing dissolved oxygen levels in water due to decomposition of aquatic plants (Pimentel *et al*., 1992). However, the use of herbicides can also be a benefit in clearing waterways of invasive plant species that can out-compete other plants for space and can clog rivers and dams (Cooper and Dobson, 2007).

Groundwater pollution by pesticides is a concern as groundwater is used as a supply of drinking water. In the US, approximately half of the population obtain their water from wells (Pimentel *et al*., 1992). A survey of groundwater in major hydrological basins in the US by Koplin *et al*., (2000), found that one or more pesticides (including metabolites) were found in nearly half of the sites sampled. The European Union have adopted a maximum concentration in groundwater of 0.1 µg/L for any individual pesticide; and 0.5µg/L for the sum of all individual pesticides (Hiscock, 2005). In the UK the Cretaceous Chalk aquifer in South East England contributes to over half of the groundwater use in the UK (Haria *et al*., 2003). There have been studies that have detected pesticide and metabolite concentrations in the chalk aquifer *e.g.* Johnson *et al*., (2001).

A guideline document issued by the European Commission defined metabolite as “a term used for all reaction or breakdown products of an active substance of a plant protection product, which are formed in the environment after the application, be it by biotic (microbials, other taxa) or abiotic processes (hydrolysis, photolysis)” (European Commission, 2003). As in the EC guideline document, this study uses the terms “metabolite”, “breakdown product”, and “degradation product” interchangeably, as there is a combination of these compounds within the data analysed. Pesticide contamination of groundwater is a concern because pesticide residues can remain in groundwater for extended periods of time. Lapworth and Gooddy, (2006) found metabolites of Diuron® (3-(3,4-dichlorophenyl)-1,1-dimethylurea) were present in the chalk aquifer several years after application. A possible explanation for concentrations of metabolites appearing in groundwater is that there are only a few microorganisms in groundwater with the potential to degrade pesticides, and groundwater recharge rates can be slow. A study by Pimentel *et al*., 1992 suggested that for some areas recharge rate is on average less than 1% per year, but this will vary with geology and rainfall. There is also a concern that some pesticides, such as Isoproturon® (3-(4-isopropylphenyl)-1,1-dimethylurea), may persist in the soil for a number of years and be mobilised by rainwater leading to groundwater contamination in the future (Johnson *et al*., 2001).

Despite the risks associated with pesticides, it has been argued that a ban on pesticides would be detrimental to the well being of humans. Guattuso, (2000) reports that by banning some pesticides it would reduce the availability and affordability of fruits and vegetables, which are needed as an important part out our diets. Therefore it is important that 'safer' pesticides are developed by ensuring that the environmental fate of pesticides is better understood.

1.2 Pesticides in the Environment

Once the pesticide is applied to the crop it is estimated that around 0.1% of the pesticide actually reaches the target pest (Arias-Estévez *et al*., 2008). Figure 1.1 demonstrates the pathways of a pesticide in the environment once applied to a crop.

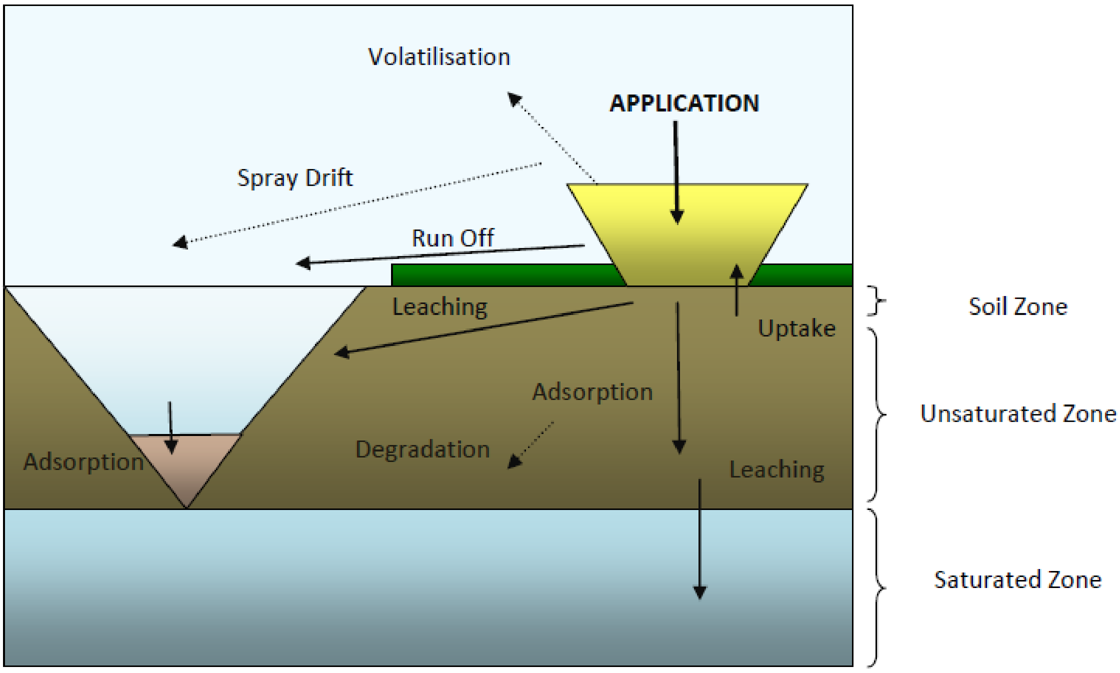


Figure 1.1 Pathways of a pesticide once applied to crops. Pathways listed account for over 99% of pesticide applied.

Pesticides are generally applied to the crop or soil as a liquid spray, but can also be applied as a seed treatment. During the spray application process the pesticides may volatilise into the air or be transported by spray drift (Figure 1.1). Volatilisation is one of the processes that contribute to pesticide dissipation from surfaces, including the crops and soil, with volatilisation being generally higher from crop surfaces than from soil (Guth *et al*., 2004). Once in the atmosphere, pesticides can travel and redeposit over large distances (van der Werf, 1996). Spray drift droplets can deposit pesticides on soil or can travel and be deposited into surface water (De Schampheleire *et al*., 2007). Although pesticide deposits on soil from spray drift can reach surface water from runoff and soil erosion (Figure 1.1), there are ways to mitigate the risk of spray drift for example, by leaving buffer zones of unsprayed land along surface waters. The use of buffer zones and other mitigation techniques is used in many countries across Europe (De Schampheleire *et al*., 2007).

Uptake of the pesticide (Figure 1.1) depends on the characteristics of the compound and the crop (Wang and Liu, 2007). For example, Behrendt and Brüggemann, (1993), reported that compounds with medium water solubility would be taken up systemically by the roots, whereas Wang and Liu, (2007) reported on foliar uptake as a diffusion process across the leaf. Uptake of pesticides to the crops would reduce effects of environmental pollution (Wang and Liu, 2007). However, foliar uptake across the leaf could potentially increase pesticide residues on the surface of plants (Juraske *et al*., 2009)

There are a number of pathways for pesticides to reach surface water from both point and diffuse pollution sources. In addition to spray drift and volatilisation, other diffuse paths for pesticides to surface water include field drain flow, and surface runoff and soil erosion (Reichenberger *et al*., 2007). Point sources of pesticide pollution are usually related to spills, such as spills during filling or cleaning of spraying equipment, which are examples of bad management practice (Holvoet *et al*., 2007). Preferential flow is the key contributor to the rapid transfer of pesticides to field drainage systems (Novak *et al*., 2001). Relatively rapid movement of pesticide loaded water through only a portion of the available pore space, decreases the residence time of the pesticide in the upper soil layers, where adsorption is usually faster than in the subsoil, which causes higher pesticide concentrations in drainage systems (Reichenberger *et al*., 2007). Surface runoff occurs when the rate of water application to the ground surface exceeds the rate of infiltration and the surface storage capacity is exceeded (Holvoet *et al*., 2007). Soil erosion consists of detachment of the soil particles from the soil surface and their subsequent transport down slope. Detachment is caused by the abrasive power of surface runoff and by raindrop impact (Morgan, 2001). The pesticides lost in runoff and erosion events leave the field either dissolved in runoff water or adsorbed to eroded soil particles (Reichenberger *et al*., 2007). Work by Kjaer *et al*., (2011) on pesticide transport and pathways found that particle-facilitated transport, such as that involving soil erosion and detachment, only accounted for a small proportion of observed pesticide leaching. Pesticides that were strongly bound to the soil, such as Glyphosate, were more likely to leach via drain connected macropores (Kjear *et al.,* 2011).

1.2.1 Pesticides in Soil

The environmental pathways that are most relevant to this study are those that occur in the soil, such as adsorption, degradation and leaching (Figure 1.1). As mentioned above, pesticides can enter surface water through surface runoff and erosion from treated fields (Reichenberger *et al*., 2007) and one of the ways pesticides can enter groundwater via leaching through the soil (Arias-Estévez *et al*., 2008). Therefore it is important that the processes involving pesticides in the soil are understood.

Factors that can affect the leaching of pesticides from the soil include the timing of application and the physical and chemical properties of the soil and compound (Hiscock, 2005). Pesticide leaching is the downward movement of pesticides through the soil profile and the unsaturated zone into groundwater (Figure 1.1). Leaching is highest in pesticides that are weakly adsorbing (Reichenberger *et al*., 2007). One of the most important processes in this study is adsorption. Adsorption is a process that removes a compound from the bulk solution and therefore affects the behaviour of the compound in the soil environment (Yaron, 1989). Adsorption can be defined as either chemical or physical adsorption. Chemi-sorption generally involves the formation of strong chemical bonds between the solute molecules and specific surface chemical groups (Weber et al., 1991). In comparison, physi-sorption generally involves weak intermolecular forces that leave the chemical structure of the surface intact (Weber et al., 1991). Covalent bonding is generally indicative of chemis-sorption, whereas van der Waals forces are involved in physio-sorption. Both of these sorption mechanisms, and others, will be discussed in more detail below. For pesticides that strongly adsorb to the soil, the risk of leaching from the root zone is generally considered to be low (Kjear *et al*., 2011). Most of the adsorption and degradation processes occur in the soil zone and unsaturated zone (Figure 1.1) as this is where soil organic matter and microbial activity are high (Pykh and Malkina-Pykh, 1997). Below the soil zone, pesticide mobility is affected by the availability of sorption sites (Hiscock,, 2005).

The extent of adsorption depends on the soil properties and the properties of the compound (Senesi, 1992). Suitable adsorption sites in soil for pesticides include organic matter and clays, in general, the degree of adsorption increases with increasing surface area and with decreasing grain size, which is why organic matter and clay particles make good adsorption sites (Hiscock, 2005). For soils with high (>5%) organic matter the mobility of pesticides has been related to total organic matter content, while in soils with low organic matter the mobility is often related to active components of the inorganic fraction, which is predominantly the clay sized fractions (Spark and Swift, 2002).

Clay Structure

Clay sized fractions can be defined as particles <0.002mm, and clay minerals are layered silicate minerals that occur in the clay sized fraction of soils and sediments (Nesse, 2000). Figure 1.2a visualises the structure in a 2:1 clay, such as a smectite.

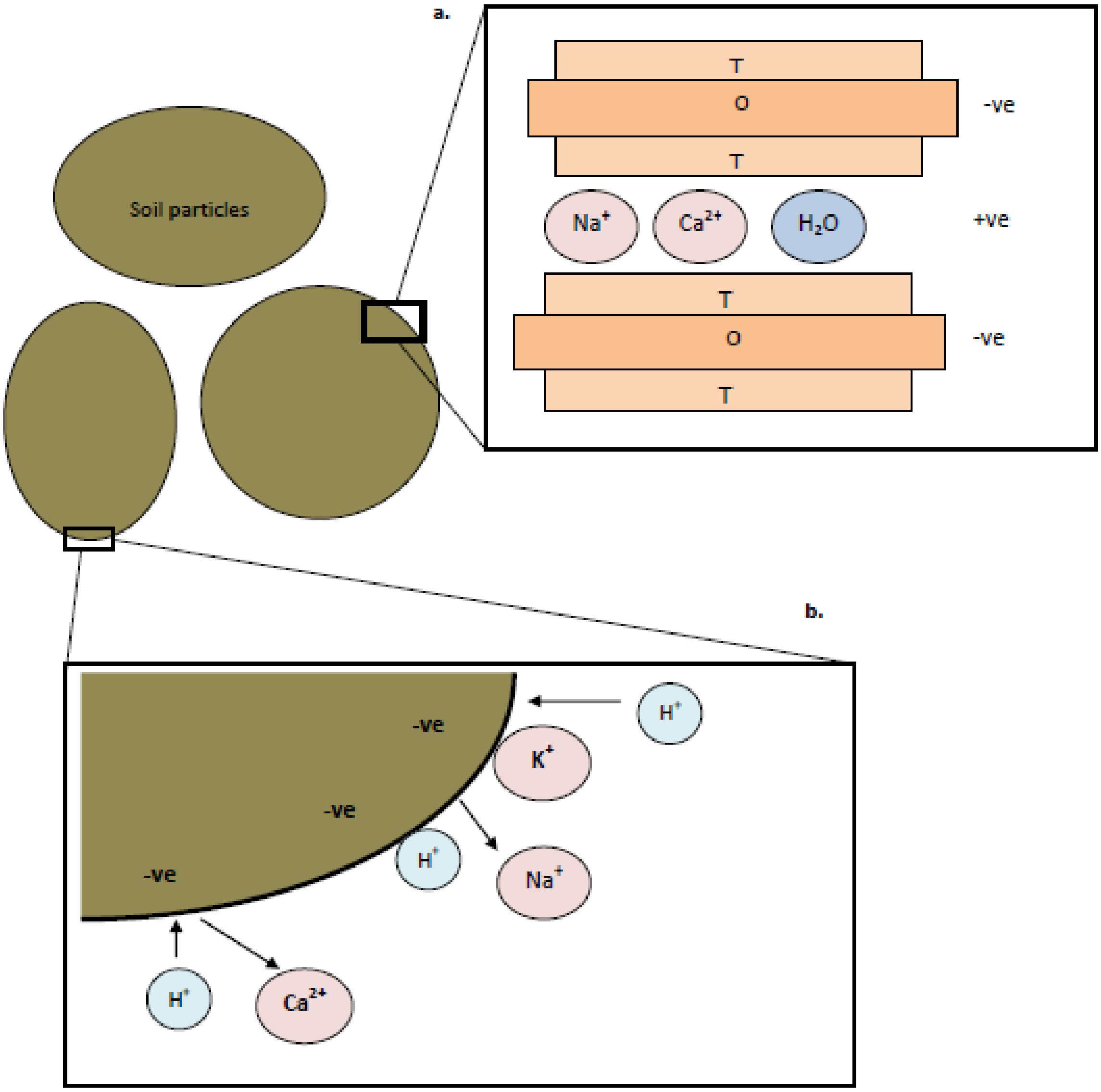


Figure 1.2 (a) 2:1 or TOT clay structure. TOT layers have negative charge, with interlayer space filled with cations or water. (b) Representation of the process of cation exchange. Hydrogen ions are strongly adsorbed to the negative sites, displacing other cations.

A 2:1 clay has repeating layers, each of which is composed of a tetrahedral (T) and octahedral sheets (O) sheets arranged in a TOT sequence. The tetrahedral sheet contains oxygen and silicon, with aluminium also replacing silicon to contribute to a negative charge. The octahedral sheet contains magnesium or aluminium between hydroxyl planes, which also had a negative charge (Yariv and Cross, 2002). The most abundant group of 2:1 clay minerals is known as smectite. The negative layer charge in a smectite is balanced by an interlayer space filled with cations, usually calcium and sodium. The relatively low negative charge, coupled with the presence of cations in only about a third of the interlayer spaces, allows water to easily move into the interlayer, causing the clay structure to swell and expand (Nesse, 2000). The process of swelling clays will be also discussed in the methodology section (Chapter 2.3.2).

The ability of pesticides to adsorb to the soil is related to types of bonding and charges. Both clays (Hiscock, 2005) and organic matter (Spark and Swift, 2002) have negatively charged adsorption surfaces onto which cations can adsorb, in a process known as cation exchange. The process of cation exchange can be represented above in Figure 1.2b. In cation exchange the hydrogen ions that are in the soil water or part of a pesticide molecular structures are strongly attracted to the negative charge on the clay layers. The hydrogen ions adsorb to the surface and in doing so, displace the other cations present in the interlayer space (Ward and Robinson, 2000).

Not to be confused with cation exchange, ionic bonding involves ionised or easily ionisable carboxylic and phenolic hydroxyl groups of humic substances (Senesi, 1992). Bipyridylium pesticides, such as Paraquat® and Diquat®, bind to soil humic substances via their cationic group, forming highly stable and unreactive bonds with the carboxyl groups of the humic substances (Gevao *et al*., 2000). Soil pH can have an effect on adsorption of pesticide (Yaron, 1989). Kah and Brown, (2007) reported that the adsorption of ionisable pesticides is strongly influenced by pH, with a decrease in adsorption often observed with increasing pH. Pesticides like triazine herbicides become cationic depending on their basicity and the pH of the soil which governs the degree of ionisation of acidic groups of the humic substances (Gevao *et al*., 2000).

The presence of numerous oxygen and hydroxyl-containing functional groups on humic substances in soil form hydrogen bonds with suitable complementary functional groups on pesticides, although pesticide molecules compete with water for these binding sites on the humic substances (Senesi, 1992). There has also been hydrogen bonding observed for adsorption to clay minerals, where hydrogen bonding is dependent on the strength of interactions between the sorbent, sorbate and water, and the pH value of the solution (von Oepen *et al*., 1991).

Chemisorption, where a bond between the pesticide molecule and the surface atoms occurs, usually takes the form of a covalent bond (von Oepen *et al*., 1991). Covalent bonds lead to stable and mostly irreversible incorporation of pesticides into humic substances (Senesi, 1992). The pesticides that are most likely to covalently bond to the soil humic matter are those that have similar functionalities to the components of the humus *i.e.* structurally resemble phenolic compounds such as the phenylcarbamates chemical group of pesticides (Gevao *et al*., 2000).

Van der Waals interactions are weak short range dipolar or induced-dipolar attractions that occur, in addition to stronger binding forces in all adsorbent-adsorbate interactions. The dipole induces small dipoles in other molecules of opposite charge, and attracts each other for a short time (von Oepen *et al*., 1991). The van der Waal forces are additive and their adsorption contribution increases with the size of the interacting pesticide (Gevao *et al*., 2000). They are of particular relevance to interactions between non-ionic and non-polar pesticides on humic acid molecules, with van der Waal forces shown to be the major adsorption mechanism for 2,4-D (2,4-Dichlorophenoxyacetic acid) (Senesi, 1992).

Humic substances contain electron deficient structures or electron rich moieties, and charge transfer complexes are formed via electron donor-acceptor mechanisms with pesticides possessing alternatively electron donor or electron acceptor properties (Gevao *et al*., 2000). Charge transfers interactions involve the overlap of their respective molecular orbitals and a partial exchange of electron density (von Oepen *et al*., 1991). Bipyridlium pesticides are thought to form a charge transfer complex with soil humic acids (Gevao *et al*., 2000).

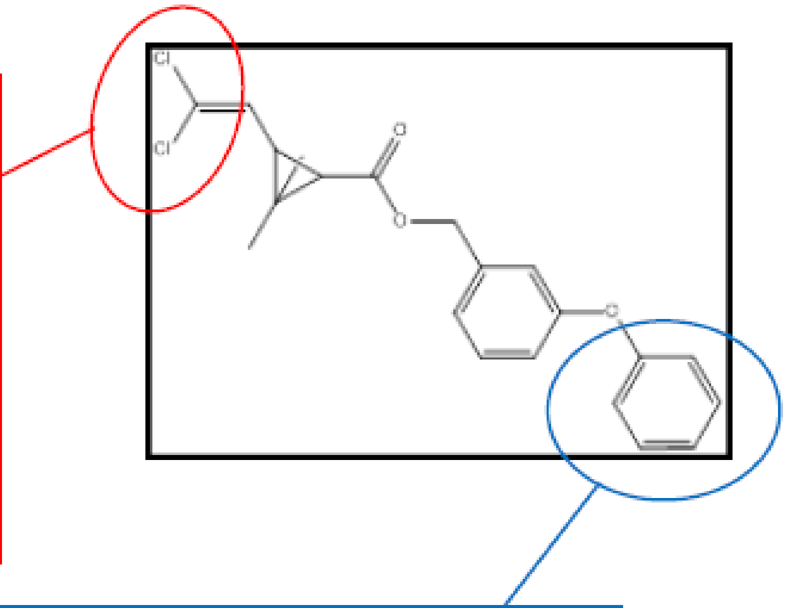
Correlations between organic matter content and the soil adsorption coefficient K_d has led to the assumption that soil organic matter is the main sorbent in soils (Wauchope *et al*., 2002). While the presence of organic matter, and clay mechanisms for adsorption are generally accepted, individual pesticides can demonstrate different adsorption mechanisms, for example one of the adsorption mechanisms of Glyphosate [N-(phosphonomethyl)glycine] is to bind to mineral sites (Kjear *et al*, 2011), but other studies have also shown that the pH of the solution is important, and adsorption of Glyphosate is not strongly correlated with organic matter content (Coupe et al., 2011). Coupe et al., (2001) also stated that Glyphosate binds to surface sites through its phosphonate group (Figure 2.1a) and therefore sorption would also be negatively correlated with phosphate concentration as it competes for the same binding sites. Using Glyphosate as an example, this highlights the complexities involved in understanding adsorption mechanisms of pesticides.

While compounds that are strongly adsorbing to soil will have less opportunity to leach to groundwater, their persistence in the soil is still an issue. Pesticides that are very persistent in soil will slowly break down and can potentially result in groundwater contamination (Tariq *et al*, 2007). For example, Glyphosate is strongly adsorbing but has a soil half life of between 1.7 – 142 days, and its metabolite AMPA has a soil half life of 76 – 240 days (Coupe et al., 2011). When compounds like

Glyphosate and AMPA are compared with compounds such as the triazole fungicides some, of which have half lives of more than 400 days (Singh, 2005) then the half life of the triazoles would obviously be more concerning for potential groundwater contamination. Pesticide degradation in soils is influenced by factors such as microbial availability (Ghafoor *et al.*, 2011). The concern regarding degradation products or metabolites is that the metabolites could potentially be more toxic or more mobile than its parent compound (van der Werf, 1996), which could then create potential problems for groundwater contamination.

1.2.2 Pesticide Chemistry

The chemistry of a pesticide has an effect on its environmental fate and its mode of action to the desired target. Figure 1.3 represents some of the common functional groups found in pesticide structures and details their potential effect on adsorption.



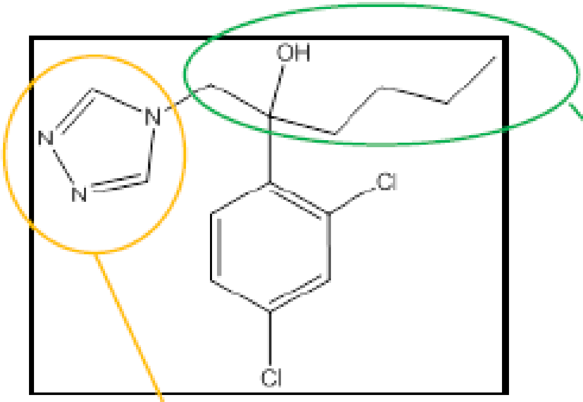
Chloroalkane

Shows strong polarity between the C-Cl bond, with the Cl atom being more electronegative. The polar molecules can interact with the water molecule, but the attractive forces are not as strong as the hydrogen bonds in water, resulting in a low solubility in water.

The same polarity also applies to fluoroalkane bonds.

Phenyl group

The phenyl group is generally a hydrophobic, non-polar stable compound, which would usually result in low solubility in water. The phenyl group is stable due to the aromatic orbital.

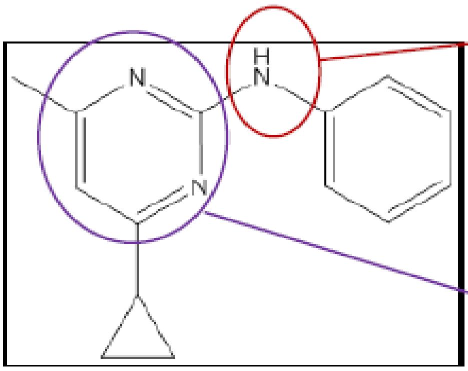


Alcohol group

In this case a 2-hexanol. The hexanol chain is slightly soluble in water. Other alcohols with lower molecular masses are soluble in water, as the alcohol molecules displace the water molecules in the hydrogen bonded association of water molecules. Alcohols with higher molecular mass, such as the hexanol chain form fewer hydrogen bonds compared to the water molecules they displace, therefore are less soluble.

Triazole group

Triazoles are very soluble, with a low pK_a value. The low pK_a means that it is an acid and easily deprotonates, with less opportunity to bond to negative adsorption sites in soils.

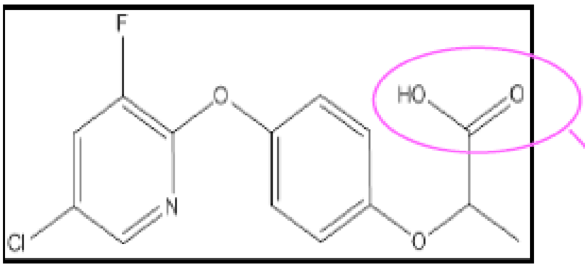


Amines

Soluble in water, as they form hydrogen bonds with water molecules

Pyridine

The pyridine group is weakly basic as it protonates at the N atom. The hydrogen ion would then be available to adsorb to negative sites in soil minerals



Carboxyl group

Carboxlic acids ionise to give hydrogen ions. The carboxylic acids dissolve in water as they form hydrogen bonds to water molecules.

Figure 1.3 Common functional groups featured in the pesticides analysed in this study

Many of the functional groups featured in Figure 3.1 that are common throughout the structures are related to the polarity of the bonds and in turn, how soluble in water it makes the structure. The issue of solubility on the environmental fate of pesticides is reoccurring in the results and validation chapters and appears to play an important role in understanding how pesticides adsorb in soil. Another common feature relating to the structure of a pesticide is the acidity or basicity of compounds and their ability to gain or lose hydrogen ions which would play an important role in ionic bonding.

Mode of Action

Crop protection products have different modes of action depending on their 'target'. The most common crop protection products featured in this study included herbicides, insecticides and fungicides, for example Glyphosate and AMPA (Chapter 4) are herbicides, and Hexaconazole (Chapter 5) is a fungicide.

In the case of herbicides, application can be pre-emergence (applied before the weed and in some cases the crop grows), or post-emergence (applied after the weed has grown). Post-emergence herbicides can be applied in the form of contact herbicides, where the entire foliage of the crop is covered; or as a systemic herbicide where the compound is taken up by the weed and translocated throughout the entire plant (Copping and Hewitt, 1998). Pre-emergence herbicides are usually taken up by the roots of the weed so the compound needs to be slightly water soluble so that it is available to the germinating weed, but no so soluble that is can be leached away from the weed germination zone. They also need to be persistent in the soil so that weeds that germinate over a period of time are controlled (Copping and Hewitt, 1998).

Fungicides, such as the triazoles group have different action against fungi. Protective action has the aim of preventing infection; curative action, where the effect of the fungicide is in the early stages of infection; and eradicant action which is used in the later visible stages of infection (Singh, 2005).

Insecticides have many different modes of action and it is not within the realm of this study to detail them all. However, using the example of the organophosphorus and carbamate groups of pesticides, can show how the molecular structure of an insecticide can assist its mode of action. Figure 1.4 shows a member of the carbamate group of pesticides. Both carbamate and organophosphorus mode of action is to inhibit the acetylcholinesterase (AChE) enzyme, which affects the nervous system of insects (Moris et al., 1995).

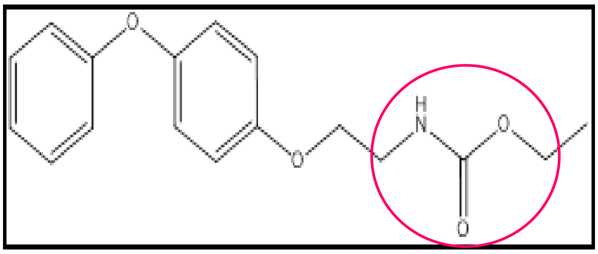


Figure 1.4 Carbamate pesticide with the ethyl carbamate functional group circled.

Both organophosphorus and carbamate pesticides inhibit AChE with the hydroxyl group in the enzyme attacking either the carbamate group or the relatively positive phosphorus atoms in the insecticides (Copping and Hewitt, 1998).

1.3 Predicting Environmental Fate of Pesticides

It is important to know the environmental fate of pesticides due to potential contamination of groundwater, surface water, and effects on humans and microorganisms *etc*. To ensure pesticide safety to humans and the environment, pesticides have a strict registration procedure and are regulated. In the European Union, the Pesticide Authorisation Directive (PAD) 91/414/EEC, which became effective in 1993, has the aim of reviewing active ingredients that had been used as pesticides in the EU before 1993. One of the main objectives of the PAD was to improve safety standards for consumers and operators; and to decrease environmental contamination (Hillocks, 2012). Annex 1 of the PAD is a positive list of active ingredient that are currently authorised to for use in plant protection within the EC. Before an active ingredient can be included in Annex 1, agrochemical companies must submit a complete dossier on both the active substance and at least one plant protection product containing the active ingredient (Hillock, 2012). Included in this dossier are studies on the environmental fate. The PAD has been superseded by the Plant Protection Products Regulation 1107/2009 which came into force in November 2009, and assesses the impact of pesticides on people and the environment. Also due to concerns about pesticide run off into water courses the Water Framework Directive (WFD) was transposed in 2003 (Directive 2000/60/EC). One of the implications the WFD has for UK farming is that certain herbicides could potentially be banned under the WFD (Hillocks, 2012).

There is much work in the literature focused on trying to model environmental fate of pesticides. Some studies on modelling pesticide inputs into ground and surface waters, for example Reichenberger *et al*., (2007) or Holvoet *et al*., (2007) while others use QSARs (Quantitative Structure Activity Relationships) to predict physical and chemical characteristics of pesticides as indicators of environmental fate. The assumption behind QSARs is that variations in activity within structurally similar compounds can be correlated with changes in parameters which reflect molecular properties (Reddy and Locke, 1994a). QSARs are widely used as alternatives to experimental determinations of environmental fate parameters such as K_{ow} and $K_{oc,}$, because experimental determinations are time-consuming and expensive (Bintein and Devillers, 1994). The K_{ow} parameter is the octanol/water partition coefficient, which is the ratio of chemical concentration in the octanol phase compared to its concentration in the aqueous phase; the higher the K_{ow} value; the greater the degree of predicted hydrophobicity. Hydrophobicity determines the amount of pesticide present in the aqueous phase and therefore available for degradation (Reddy and Locke, 1994b). The most accepted measure of sorption of pesticides to soil is the soil sorption coefficient, normalised to organic carbon, K_{oc}, (Chapter 2). K_{oc} values provide a relative measure of mobility in aqueous/soil systems; where compounds with higher K_{oc} values will be less mobile than those with lower values (Sabljić *et al*., 1995). This study is focused on the modelling and prediction of K_{oc}, as it is the parameter that is favoured by industry risk assessments (Wauchope *et al*., 2002).

1.4 Study Outline

The overall aim of the study was to *understand the controls on the K_{oc} of pesticides in soils, with a view to developing predictive QSAR models*. The overall aim was broken down into two objectives

1. Understand controls on adsorption.

Objective 1 was to investigate how K_{oc} varies between different soils and different compounds, specifically using compounds molecular properties to identify a correlation with K_{oc}. It was hypothesised that certain structural and soil properties would correlate significantly with K_{oc}. A better understanding of K_{oc} has the advantage of potentially preventing pesticide movement from soils into groundwater, which is important for groundwater quality and pollution.

2. Develop models for predicting K_{oc}.

Objective 2 was to investigate if, once a correlation with K_{oc} was identified, could those parameters be used to predict K_{oc} for other compounds. Predictive models have the potential to improve the pesticide development process, as they could reduce the need for experimental K_{oc} values in the early stages of the pesticide development process.

The outline of this study is as follows:

Following the introduction, **Chapter 2** explains the process of obtaining the structural properties of the compounds to use in the QSAR models. The methodology explains the statistical analysis used to build the predictive models and explains the reasons why the particular structural parameters were selected for analysis. Chapter 2 also outlines the batch equilibrium procedure for experimental K_{oc} values that were used to obtain additional data for Chapter 4 and Chapter 5.

Chapter 3 reports the preliminary analysis of a diverse range of compounds, obtained from industry reports. The preliminary analysis showed that adsorption behaviour of pesticides could generally be modelled into 'high' and 'typical' K_{oc} values. Chapter 3 reports on the structural parameters that are important for high adsorption compounds.

The analysis of parent and metabolite compounds is reported in **Chapter 4**. The parent and metabolite compounds were analysed separately. Additional data was obtained for this study by experimental work. Logistic regression models were able to identify between parent compounds and metabolite compounds based on their molecular structures. More importantly, logistic regression models could predict the adsorption potential of metabolites relative to their parent, which has relevance for risk assessments of the environmental fate of metabolites.

The importance of soil properties on adsorption of pesticides is analysed in **Chapter 5**. Chapter 5 outlines the details of the fieldwork and experimental work to obtain K_{oc} values for a compound, tested on 24 soil types. The experimental results are analysed and a predictive K_{oc} model based on soil type is presented.

Chapter 6 presents the results of a study of modelling Stage 1 compounds, which are compounds in the early stages of compound development. Chapter 6 also discusses some of the practical issues in predicting K_{oc}. Chapter 6 concludes that prediction of K_{oc} is possible for this particular situation and suggests options for future work.

The Stage 1 model is validated in **Chapter 7**. Chapter 7 also compares K_{oc} predictions from this Stage 1 model against predictions of the current US EPA model. The suitability of this model and of using K_{oc} as a predictor of environmental fate is discussed and suggestions for alternative predictors are made. The main results of this study are concluded in **Chapter 8**.

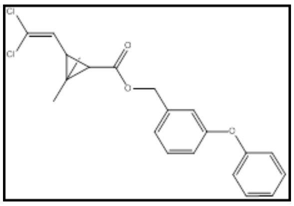
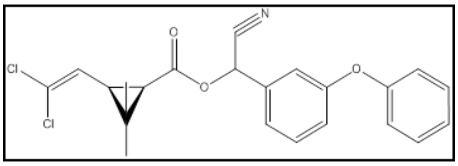
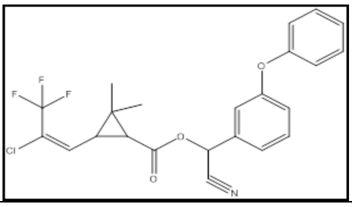
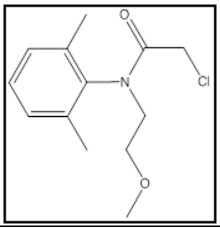
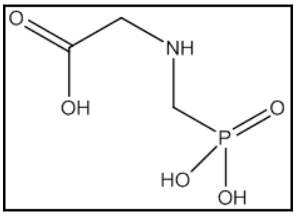
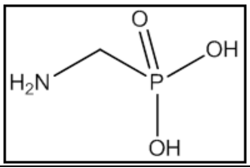
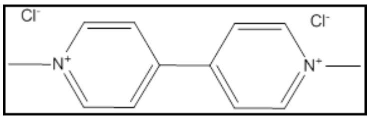
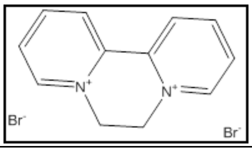
2. METHODOLOGY

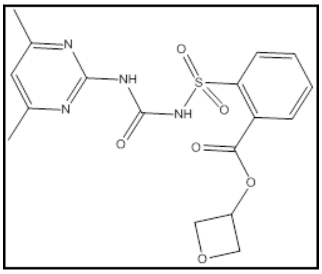
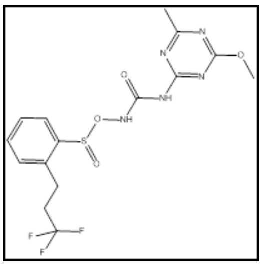
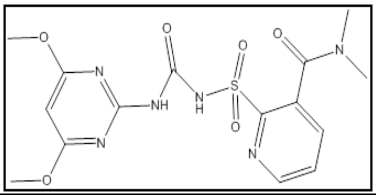
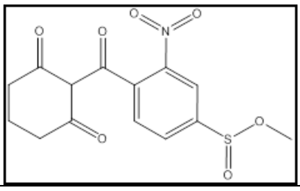
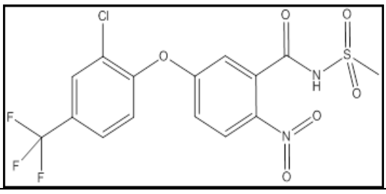
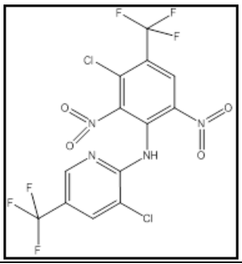
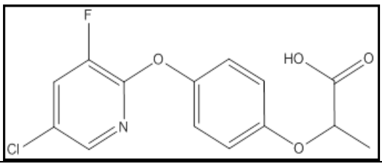
2.1 Data Collection

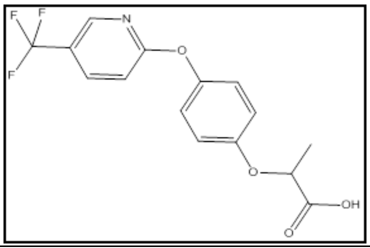
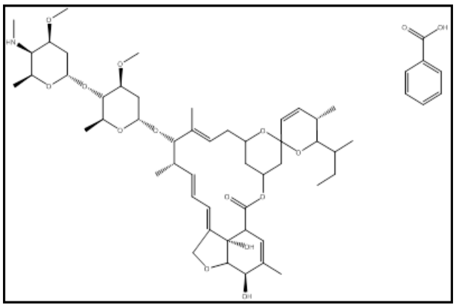
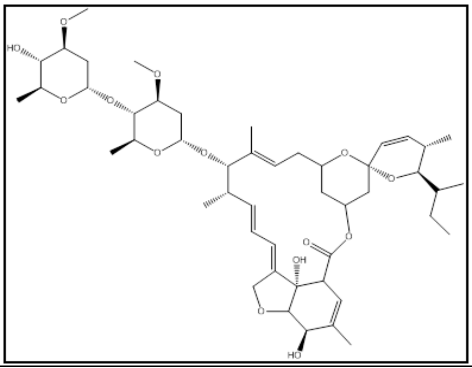
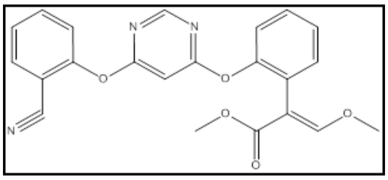
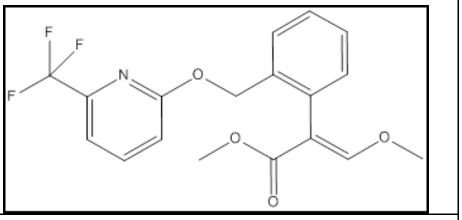
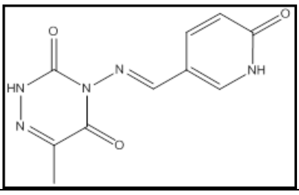
The data supplied to this project were obtained from Syngenta's SmartDoc database. The data were in the form of industry reports with the results of adsorption studies performed at various Syngenta laboratories and contract laboratories. The reason for using these reports is that there were large amounts of data available in the SmartDoc system, so a large dataset of adsorption, compound and soil properties could be constructed.

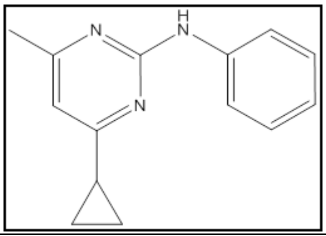
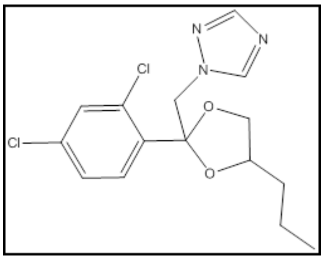
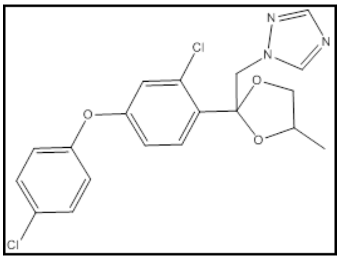
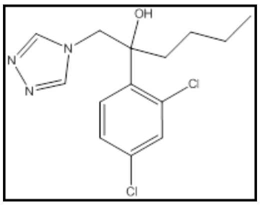
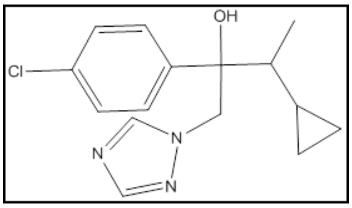
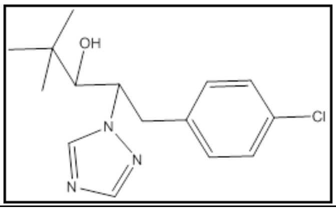
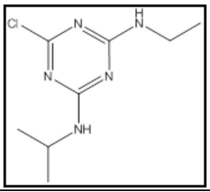
The advantage of using these reports is that the studies were carried out using a relatively consistent methodology across all the laboratories, and the majority of studies in the database had to follow Good Laboratory Procedure (GLP). This meant that there was reduced variability in the results obtained in the studies, which were then used in constructing the dataset. However, by only using reports obtained from Syngenta, it led to some restriction in the Active Ingredients (AIs) that could be used in constructing the dataset, i.e. only compounds and reports that were recorded in the SmartDoc system could be used in the study. This means that there may be some reduced variability in the range of AIs and also quantity of adsorption data available for each AI, for example, some compounds only had one study, whereas others had three or four, giving more soil data for each compound. A table of the main compounds discussed in this study is given in Table 2.1.

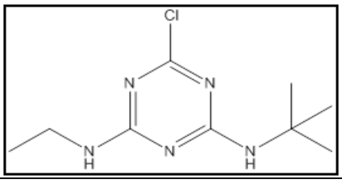
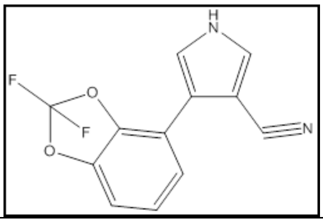
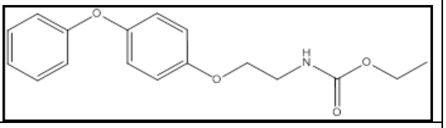
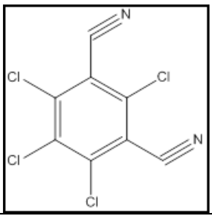
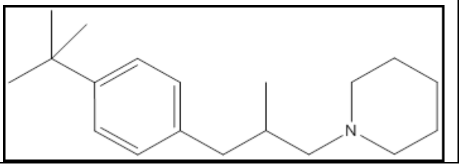
Table 2.1 Structures and chemical classes of the most commonly discussed pesticides in this study. Structures are taken from the original reports or from The Pesticide Manual (Tomlin, 1997). Unless stated all chemical structures are given as their neutral structures.

Chemical Class	Example Compounds	Structure
Pyrethoid	Permethrin	
	Cypermethrin	
	Lambda-cyhalorthrin	
Choloacetonilide	Dimethachlor	
Aminophosphonic	Glyphosate	
	AMPA	
Bipyridylum	Paraquat	
	Diquat	

Sulfonylurea	Oxasulfuron	
	Prosulfuron	
	Nicosulfuron	
Benzoylcyclohexanedione	Mesotrione	
Diphenyl-ether	Fomesofan	
2,6 - dinitroaniline	Fluazinam	
Aryloxphenoxy propionic acid	Clodinafop	

	Fluazifop	
Avermectin	Emamectin B1a	
	Abamectin B1a	
Stobilurin	Azoxystrobin	
	Picoxystrobin	
Azomethine	Pymetrozine metabolite	

Pyrimide	Cyprodinil	
Triazole	Propiconazole	
	Difenconazole	
	Hexaconazole	
	Cyproconazole	
	Paclobutrazol	
1,3,5-Triazine	Atrazine	

	Terbuthylazine	
Phenylpyrrole	Fludoxonil	
Carbamate	Fenoxycarb	
Chloronitrile	Chlorothalonil	
Unclassified	Fenpropidin	

2.1.1 Constructing the Dataset

The information within the industry reports that was of interest in creating the dataset included:

- Adsorption information - This would be a table of K_{oc} and K_d values for the compound by each soil type. The K_{oc} values collected from the reports are the linear distribution coefficient not K_{foc} which refers to the Freundlich distribution coefficient; the difference between K_{oc} and K_{foc} and the reason for using the K_{oc} parameter will be explained in more detail in Section 2.1.2.
- Compound information - This just included the name or development code for the compound and usually the structure or molecular formula for the compound.
- Soil data - This comprised the name of the soil, its USDA textural classification, the country it was sampled from and depth sampled in soil profile (if applicable). The soil data that was available and would be analysed comprised of its pH, Cation Exchange Capacity (CEC),

organic matter (OM) and/or organic carbon (OC) content and the percentage of sand, silt and clay.

The compound's structure was modelled into the commercial software package HyperChem 8.0 (Hypercube Inc.), where the molecular properties were obtained and recorded in the dataset. Molecular properties will be the term used to cover quantum and semi-empirical properties. The structure was geometry optimised by HyperChem so that the bonds in the structure were at the correct angles. The geometry optimisation feature employs energy minimisation algorithms to locate the stable structure (Hypercube Inc.) The structure could then be redrawn into ISIS Draw (MDL Information Systems) so that the structure could be imported into the software package Topix 1.0 (http://www.lohninger.com/topix.html), where atom and bond counts, connectivity parameters and molecular fragments were collected and recorded. Table 2.2 lists the molecular properties that were calculated from the structure:

Table 2.2 Descriptions of molecular properties used in the data analysis. PCA Abv. refers to how the structural property is listed in the tables of PCA loadings in results chapters.

Symbol	Structural property	PCA Abv.	Definition/Physical interpretation
μ	Debye dipole moment	dipolm	solubility
totalE	Total energy	totalE	size descriptor
HOMO	Energy of Highest Occupied Molecular Orbital	HOMO	Acidity
NHOMO	Energy of the Next Highest Occupied Molecular Orbital	NHOMO	
LUMO	Energy of the Lowest Unoccupied Molecular Orbital	LUMO	Basicity
NLUMO	Energy of the Next Lowest Unoccupied Molecular Orbital	NLUMO	
V_{sav}	Surface accessible volume	Vsav	size descriptor
A_{sas}	Solvent accessible surface area	Asas	size descriptor
V_{vdW}	van der Waals volume	Vvdw	size descriptor
A_{vdW}	van der Waals surface area	Avdw	size descriptor
ΔHhyd	Hydration enthalpy	Δhhyd	solubility
LogP	Octanol-Water Partition Coefficient	logp	
n	molecular refractivity	refractivity	refractive index value
π	polarisability	polarisability	solubility
M	relative molecular mass	mass	size descriptor

It should be noted that unless stated all the molecular structures were assumed to be neutral structures. The problem with assuming a neutral structure is that the electronic molecular properties, such as the hydration enthalpy, the dipole moments, the HOMO and LUMO, may change if the structure was then to become charged and may therefore have an effect on the pesticide behaviour in the environment. The molecular properties were calculated using HyperChem using the set up menu and selecting molecular mechanics (MM+) to calculate V_{sav}, A_{sas}, V_{vdW}, A_{vdW}, ΔHhyd, logP, µ,n, π, and M_r. Molecular mechanics (MM+) method uses a general purpose force field to treat atoms as an individual unit to calculate the quantum properties (Hypercube Inc). Then by selecting semi-empirical (extended Hückel method) the total energy and the HOMO, NHOMO, LUMO and NLUMO were calculated. The extended Hückel method uses both the pi and sigma bonds to calculate the energies of the molecular orbitals (Hypercube Inc.)

Additional properties that were calculated from the structures were:

- Bond and Atom counts - This is the number of the following atoms in the molecule: carbon, oxygen, nitrogen, fluorine, chlorine, phosphorous and silicon. Number of: single, double and triple bonds and carbon-carbon single bonds, carbon-carbon double bonds, carbon-carbon triple bonds and number of aromatic bonds.
  - The bond and atom counts could be calculated by counting the relevant bonds and atoms in the compounds molecular structure, but for ease this was obtained as part of calculations performed by Topix.

- Connectivity parameters - In this report the connectivity parameters will be expressed as $h_{\chi}{}^{v}_{m}$ where: χ is connectivity, *h* is calculated from zero to ninth order; *m* is designated p for path fragments, c for cluster fragments and pc for path-cluster fragments; and v denotes that the value is valence corrected. The connectivity parameters used in this study were valance corrected (Kier-Hall Connectivity parameters). There were 29 connectivity parameters calculated for each compound in the dataset. 0_{χ} through to 9_{χ} (zero order connectivity to ninth order connectivity) were classed as low order connectivity, and were interpreted as a size or volume descriptor. ${}^{3}_{0}\chi^{c}_{p}$ through to ${}^{4}_{9}\chi^{c}_{p}$ (third order cluster, zero order path connectivity through to fourth order cluster, ninth order path connectivity). Path and cluster connectivities can be related to the degree of branching in a molecular structure.

  - Molecular connectivity can be defined as:

$$h_{\chi} = \Sigma\,(\delta^{v}_{i} * \delta^{v}_{j} \dots \delta^{v}_{h+1})^{-1/2}$$

(Equation 1)

Where: δ^v = the vertex degree from the adjacency matrix (valence corrected), i, j ... refer to pairs of adjacent atoms, and the reciprocal square root products are summed across all possible subgraphs of the hydrogen suppressed graph of the molecule (Worrall, 2001). The connectivity parameters are calculated through the input of .mol files of the compounds into Topix. The structure of the compound was drawn in ISIS Draw and saved as a .mol file. A .mol file is a text file that can be read in Notepad and shows rows of numbers and letters than are identifiable as elements of the compounds structure. In Topix the compound file was loaded and the parameters to be calculated were selected from tick boxes. The output file, .asc, can be opened in a program such as Excel and lists the calculated values from the parameters.

- The molecular fragments: - these are sometimes referred to as augmented atoms and were calculated as: C1C1C, C1C, O1C, C1O1C1C, C1C1C1C, and C2C1C. This is where the first atom is the augmented atom and the others are those that are bound to it by the bond of order given by the number between. For example, O1C is oxygen bonded to a carbon by a single bond.
  - The molecular fragments are one of the parameters that can be calculated by Topix.

2.1.2 Justification of Parameters

Soil Adsorption Coefficients

Soil sorption is characterised by a partition coefficient K, with subscript d for distribution (Wauchope *et al*., 2002). K_d is defined by:

$$K_d = \frac{x/m_s}{C_e}$$

(Equation 2)

Where: x/m_s = concentration of pesticide in the solid phase and C_e = concentration of pesticide in liquid phase. There is a generalisation that K_d shows a high correlation with organic matter content of soils (Wauchope *et al*., 2002). Therefore sorption is often characterised by K_{oc}, the sorption distribution normalised to organic carbon, which is defined as:

$$K_{oc} = \frac{K_d}{OC} \times 100$$

(Equation 3)

Where: K_d is the linear sorption distribution coefficient, as determined by batch equilibrium experiments (see Section 2.3), and OC is the organic carbon content.

Sorption data originating from batch equilibrium studies are often described using the non-linear Freundlich equation:

$$\frac{x}{m} = K_f(C_e)^{1/n}$$

(Equation 4)

Where: x/m = amount adsorbed; C_e = concentration of pesticide in solution; K_f = Freundlich distribution coefficient; and 1/n = exponent (Dubus *et al.*, 2003). The value of 1/n usually lies between 0.7 and 1 (Wauchope *et al.*, 2002). Dubus et al. (2003) reported that it is common practice to normalise the K_f coefficient to organic matter and report this new value as K_{oc}. However the initial K_{oc} value relates to the linear distribution of the coefficient (Equation 3). Therefore there could be possible confusion in the literature surrounding the term K_{oc}, as it could refer to the linear or Freundlich distribution. In terms of this potential confusion, Dubus *et al*. (2003) states that the differences in the values between K_{oc} and Kf_{oc} will be maximum for compounds that exhibit strong non-linear sorption. Wauchope *et al*. (2002) follows this by stating that the most important consequence of isotherm non-linearity of the Freundlich type with 1/n < 1, is that mobilities for compounds at very high concentrations will be under predicted by K_d values measured at lower concentrations and vice versa. This project used the linear K_{oc} coefficient as this is the parameter that tends to be used by modellers (Dubus *et al*., 2003) and allowed the models produced in this project to be compared to models used in the regulatory process *e.g.* EPIWIN (http://www.epa.gov/oppt/exposure/pubs/episuitedl.htm).

Structural Properties

The structural properties in the data base generally have some kind of physical interpretation and there has been much work in the literature around the subject of predicting soil adsorption coefficients based on structural properties. Meylan *et al*. (1992) and Worrall (2001) are among those who have used molecular topology to predict K_{oc}. The molecular topology relates to the connectivity parameters as described above in Section 2.1.1. Worrall (2001) states the molecular topological approach has an advantage as it is simple and that no more information than is present in the structural formula is required as an input. The work of Meylan *et al*. (1992) and Worrall (2001) showed lower order connectivity parameters to be successful in building K_{oc} models. The lower order connectivities can be related to molecular volume and molecular size, which directly relate to the solubility of the compound, with higher order connectivities related to the degree of branching (Worrall, 2001). Meylan *et al*. (1992) and Lohninger (1994) used connectivity parameters alongside molecular fragments to model K_{oc}. Lohninger (1994) also stated that using molecular fragments are an advantage as they can be readily calculated before any experiments, providing a good method for screening future pesticides. Both Meylan *et al*. (1992) and Lohninger (1994) found that including molecular fragments with connectivity parameters in the model improved the estimation of sorption coefficients. Molecular fragments are known to generally have some influence on the chemical and physical properties of a compound (Lloyd, 1989). For example, Lohninger (1994) found polar fragments such as hydroxy groups (O1C fragment) decreased the sorption coefficient.

Other authors such as Reddy and Locke (1994a) and Gramatica *et al*. (2000) have used quantum parameters to predict K_{oc}. The quantum parameters relate to the molecular properties listed above in Section 2.1.1. Dai *et al*. (2000) found that using molecular properties like the ones listed above had advantages as they were easily obtained by computer calculations and the parameters had "clear chemical sense". The molecular parameters can be interpreted in terms of their affinity to adsorb pesticides to soil. Worrall and Thomsen (2004) included quantum and molecular properties, as well as topological properties in their model to predict groundwater pollution. By including connectivity parameters in the model the explained variance in the data increased (Worrall and Thomsen, 2004).

As outlined in Section 2.1.1, this project chose to use a mixture of connectivity parameters, molecular fragments and molecular properties to build the models. The reason these parameters were selected was to attempt to increase the explained variance in the data as shown by Worrall and Thomsen (2004) and to attempt to get many parameters that could explain most of the adsorption behaviour of the compounds.

2.2 Data Analysis

The software used to analyse the data was Microsoft Excel 2003 and 2007 and statistics software Minitab v14.

There was potentially a problem with having such a large dataset. There are around 80 variables to consider and with so many variables being analysed together it would be difficult to see what is actually causing a variation in adsorption. A way of resolving this was to group the variables into similar categories which were then analysed separately. The categories for analysis were: soil properties; quantum properties; and connectivity parameters with molecular fragments - the variables in each category have been described in section 2.1.1. Dividing the dataset up into categories like this can still create an issue of co-linearity. Co-linearity is where some variables, or predictors, show a high correlation with each other and would suggest that the variable shares something in common with the other variable. The problem of co-linearity means that if these variables are included in the data set then it would not explain much of the variation in the data (Howell, 1997). An easy way to resolve this problem is by identifying predictors in the database with opposite coefficients, which are predictors that have a similar coefficient but one variable has a negative value and the other is positive. If predictors with opposite coefficients are found then one is removed.

2.2.1 Principal Component Analysis (PCA)

Principal Component Analysis was used first to analyse the full dataset of 700 K_{oc} values because it is a way of converting a large multivariate dataset into a smaller number of variables. Principal Component Analysis transforms the original, observed, variables into principal components that are a linear function of the original variable set (Wang, 2009). The new principal components are independent of each other and thus avoiding the issue of co-linearity (Wang, 2009). The PCA identifies which combinations of variables explain the largest amount of variation in the dataset (Fowler *et al*., 1998), but is also useful for identifying groups of data and multiple trends. The principal components are ordered so that the first principal component is chosen to explain the largest possible amount of information in the data and the second principal component explains the second largest amount of information in the data and so on (Wang, 2009). The first few principal components therefore, account for a large proportion of the total variance in the dataset.

Principal Components (PCs) were calculated from Minitab. Principal Components were selected for examination if their eigenvalues >1.00 and including the first PC with an eigenvalue <1.00. The reason for using 1.00 as a cut off point is because any variable that is more or less independent of all other variables will have an eigenvalue close to 1.00 but will still be important when explaining the overall variance (Jolliffe, 2010). The reason only a few principal components are chosen was because the idea of PCA was that the first few principal components will jointly explain a reasonably large proportion of the information in the original sample set (Fowler *et al*., 1998). When the principal components were calculated a set of loadings were produced, with a value for each variable in the dataset. The loadings were interpreted as if it were a regression coefficient, although it cannot be tested for significance. As one of the aims of the project was to determine controls on adsorption, the principal components were selected by choosing the two principal components that had the highest loadings for K_{oc}.

Scores for the principal components were also calculated by Minitab. The scores are the transformed variables values and correspond to a particular data point (Shaw, 2003). Therefore the scores were used to create scatter plots for the first two principal components relating to K_{oc}. The PCA was run a number of times for each category of predictors as outliers on the scatter plots were removed, following visual inspection, at each run to help make the trends clearer and reduce noise in the dataset.

An advantage of PCA is that it possible to visualise multiple trends within a dataset. Therefore, scatter plots of PCs were examined and redrawn by chemical group classification of the compounds. The chemical group classification of each compound was determined either from information given in the industry studies or, if that information was not available, from its classification in 'The Pesticide Manual' (Tomlin, 1997).

2.2.2 Multiple Regression

The principle of multiple regression is to find an equation to predict Y on the basis of *j* number of predictors or variables. The general equation for multiple regression is:

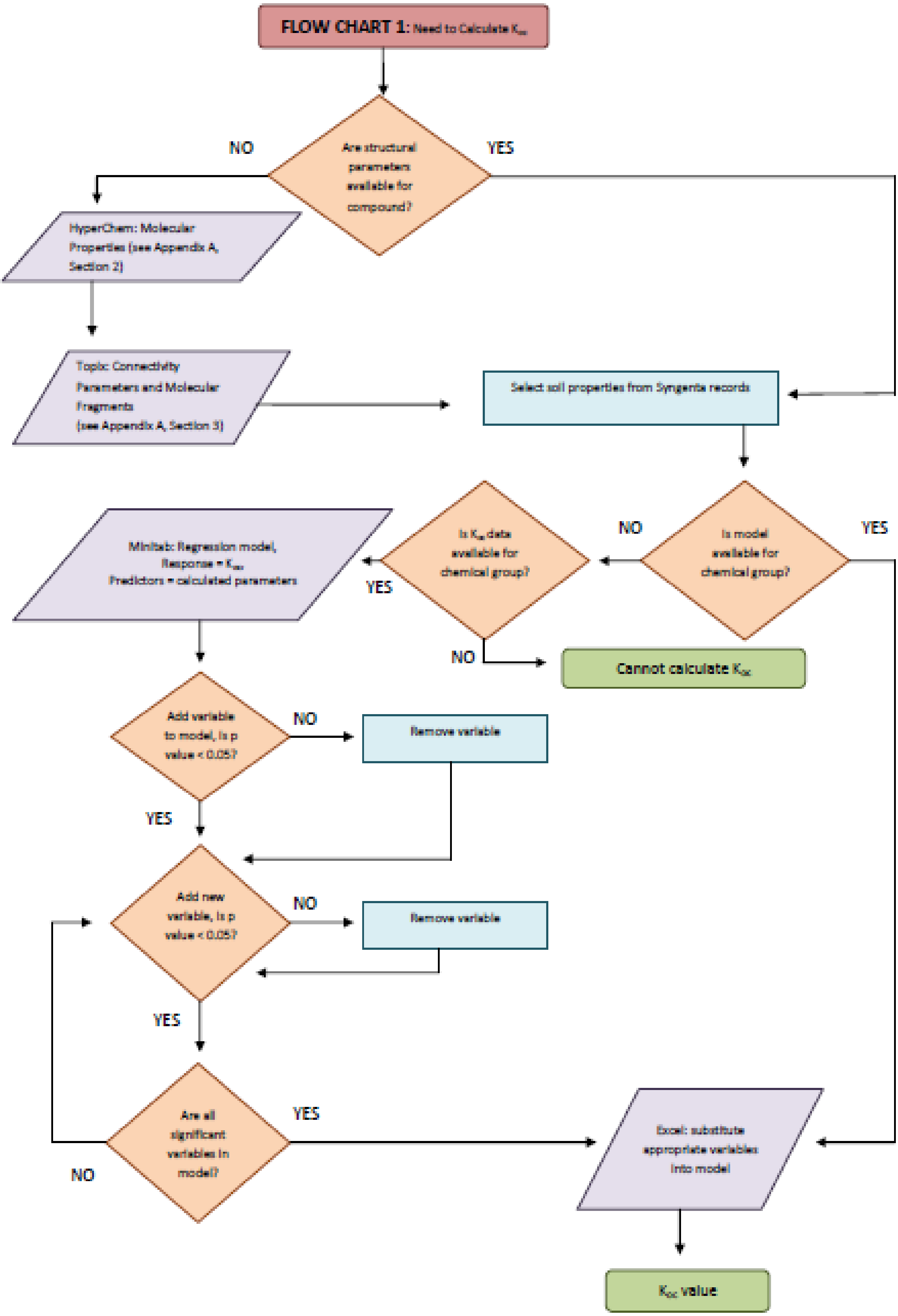
$$\mathrm{Y_i} = \mathrm{A} + \sum_{\mathrm{j=1}}^{\mathrm{j=V}} \mathrm{M_j} * \mathrm{X_{ji}}$$

(Equation 5)

Where: Y_i = the dependent variable; X_j = the jth explanatory variable; A = constant; M_j = the jth regression coefficient; and V = the number of explanatory variables (Shaw, 2003). For this project, the aim was to see if the variation in K_{oc} can be predicted from different soil or compound properties so $Y_i = K_{oc}$ and X_j = variables from the dataset *e.g*. dipole moment or %OC.

The dataset was analysed in the subcategories as outlined in section 2.2. Multiple regression was tried in three ways so as to make sure the model was as accurate as possible and to avoid any co-linearity. Firstly, variables were added to the model individually, if the variable was found to be significant at the 95% level then other variables were added one at a time in a stepwise manner until no more significant variables were found. Variables were added or removed according to their significance, and therefore their influence on the R^2 value. The R^2 was calculated by Minitab and can be directly interpreted in terms of the percentage of variance in the original dataset that was accounted for by the model (Howell, 1997) – this is a forward stepwise regression method. The second method for creating, and checking the model was to add all the variables from the subcategories at once then remove them one at time in a backward stepwise manner, according to their P values. In this case a variable would be removed if the P value was above 0.05 by convention and in accordance with other studies (Pires *et al*. 2008). Thirdly, in addition to the forward and backward stepwise methods, all of the variables were added in the model individually then removed to record their effects on the R^2 value: this was also a check for co-linearity.

The process for collecting the molecular properties and analysisng the data to build multiple regression models is shown below in Flow Chart 1. Full details of the steps needed to obtain the compounds structural properties are found in Appendix A.

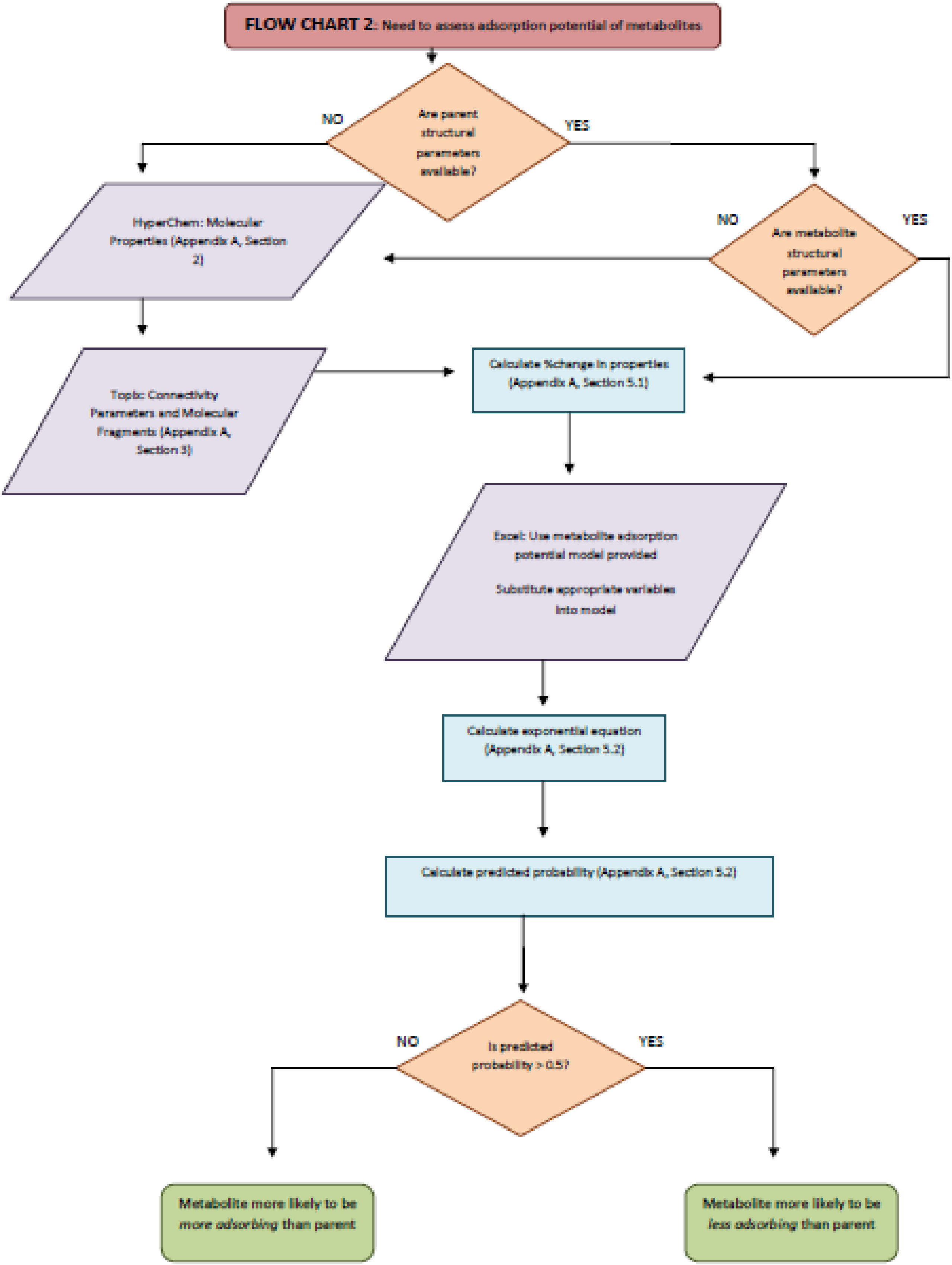

FLOW CHART 1: Need to Calculate K_{oc}
Are structural parameters available for compound?
NO
YES
HyperChem: Molecular Properties (see Appendix A, Section 2)
Topic: Connectivity Parameters and Molecular Fragments (see Appendix A, Section 3)
Select soil properties from Syngenta records
Is model available for chemical group?
NO
YES
Is K_{oc} data available for chemical group?
YES
NO
Cannot calculate K_{oc}
Minitab: Regression model, Response = K_{oc} Predictors = calculated parameters
Add variable to model, is p value < 0.05?
NO
Remove variable
YES
Add new variable, is p value < 0.05?
NO
Remove variable
YES
Are all significant variables in model?
YES
NO
Excel: substitute appropriate variables into model
K_{oc} value

2.2.3 Logistic Regression

Logistic regression is a technique for predicting a binary outcome from continuous explanatory variables (Worrall, 2001). For example, in this project it was used for analysis of the parent and metabolite compounds, so a binary outcome of parent vs. metabolite from the continuous variables measured in the database. The method transforms from a probability scale (0, 1) to the scale of continuous variables (∞, -∞) (Worrall, 2001). The transformation used is the logit transform $y = \log\left(\frac{\theta}{1-\theta}\right)$. For example in this project θ = the probability of a compound being a parent compound. The transformed parameter y can be linearly related to the chosen explanatory variables. This regression uses a maximum likelihood estimation rather than least squares estimation of coefficients as would be used in multiple regression.

The dataset was analysed in the subcategories (as described in Section 2.2) but only parent compounds and their respective metabolites were analysed. The analysis of the parent and metabolite compounds using logistic regression is described in detail in chapter 4. Logistic regression was also used during the preliminary analysis to segregate compounds with 'typical' adsorption behaviour from those compounds that displayed more unusual adsorption behaviour (the model is shown in Chapter 3).

An example of the process for collecting the molecular data, and using the logistic regression models is represented by Flow Chart 2, which in this case has been used to assess the adsorption potential of metabolites (the results of this study are described fully in Chapter 4). The method for building the logistic regression models is similar to building multiple regression models, so Flow Chart 2 instead assumes a logistic regression model has been built and demonstrates how the model is used to calculate the adsorption potential. Again, full details for obtaining the structural properties are in the instructions in Appendix A.


FLOW CHART 2: Need to assess adsorption potential of metabolites
NO
Are parent structural parameters available?
YES
HyperChem: Molecular Properties (Appendix A, Section 2)
NO
Are metabolite structural parameters available?
YES
Topic: Connectivity Parameters and Molecular Fragments (Appendix A, Section 3)
Calculate %change in properties (Appendix A, Section 5.1)
Excel: Use metabolite adsorption potential model provided
Substitute appropriate variables into model
Calculate exponential equation (Appendix A, Section 5.2)
Calculate predicted probability (Appendix A, Section 5.2)
NO
Is predicted probability > 0.5?
YES
Metabolite more likely to be more adsorbing than parent
Metabolite more likely to be less adsorbing than parent

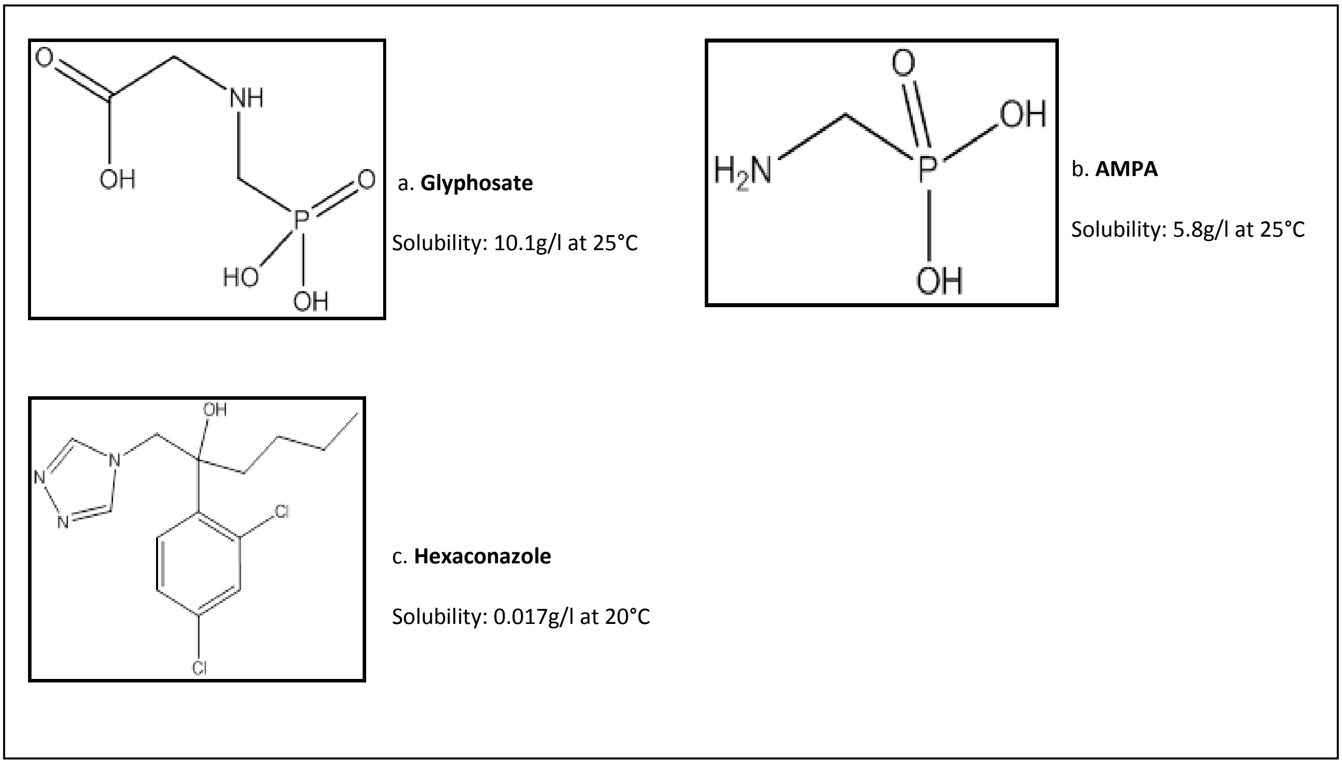
2.2.4 Analysis of Variance (ANOVA)

The technique of ANOVA was used to support the multiple linear regression. The advantage of using ANOVA is that it can model two independent variables as well as modelling the individual effects of each variable separately: the interacting effects of the variables can be analysed (Howell, 1997). For this project ANOVA was used to determine whether soil properties or compound properties were more important in controlling adsorption.

The ANOVA experiment was set up in Minitab as a two factor experiment with compound and soil as the factors. It was not found necessary to transform the data in order to meet the assumptions of normality and homogeneity of variance within ANOVA.

2.3 Experimental Methodology

Additional K_{oc} data were obtained for three compounds using a batch equilibrium experiment. The three compounds tested were an aminophosphonic compound (Glyphosate), its metabolite (AMPA), and a compound from the triazoles group (Hexaconazole) Figure 2.1.



a. **Glyphosate**

Solubility: 10.1g/l at 25°C

b. **AMPA**

Solubility: 5.8g/l at 25°C

c. **Hexaconazole**

Solubility: 0.017g/l at 20°C

Figure 2.1. Structures of the three compounds used in the batch equilibrium experiments. All structures are given as neutral structures and were obtained from Tomlin (1997). Glyohosate and AMPA solubility obtained from Coupe et al., (2011), Hexaconazole solubility as given in Tomlin (1997).

The experiments were performed at Syngenta laboratories' at their Jealott's Hill research site and were carried out according to OECD guidelines (OECD 2000), and therefore the K_{oc} results from the experiments should be comparable to the K_{oc} values already in the dataset. Although, the adsorption experiment was performed to OECD standards, unfortunately due to the time restraints, no preliminary study (Tier 1 – OECD 2000) was made and no determination of adsorption isotherms (Tier 3 – OECD 2000) was carried out. However, any specific details of a Tier 1 study that was needed to determine the soil/solution ratio and equilibrium time for adsorption for the three compounds was taken from previous reports of those compounds made by Syngenta. Also the experiment was carried out under the guidance of one of Syngenta's study directors who ensured the methodology was appropriate. Tier 3 was not necessary for these experiments as all the values taken from the reports were K_{oc} values, therefore calculating K_{foc} values were not a priority.

2.3.1 Selection of Soil Properties

The reason for the adsorption experiments was to expand the number of K_{oc} values per compound in the database and expand the range of soils included. The additional soils came from the collection of standard soils that Syngenta use in batch equilibrium experiments, and therefore already met OECD guidelines. The three compounds chosen had not been tested on these soils before. Full details of the characteristics of the soils can be found in Chapters 4 and 5. The study based around Hexaconazole (Chapter 5) used seven new UK soils in addition to the Syngenta soils available. These new soils were selected for specific reasons based on their underlying geology, and as much as possible tried to meet OECD guidelines (OECD 2000). The seven new soils were prepared in the same way as the Syngenta soils. For full characteristics of the collected soil see Chapter 5.

2.3.2 Batch Equilibrium Adsorption Study

The following outlines the general methodology that was followed for each of the three compounds. All three compounds were available in a radio-labelled form. All calculations and serial dilutions are given in Appendix B – Folder 1.

Experiment Preparation

By studying the previous reports for the three compounds appropriate soil: aqueous ratios for each compound were selected. The soil: aqueous ratio for Glyphosate was 1:20 (1g soil: 20mls aqueous solution), and for AMPA and Hexaconazole was 1:10 (2g soil: 20mls aqueous solution). As 1ml of the aqueous solution will be the radioactive treatment solution (i.e. contains the radio-labelled compound) and to maintain the correct ratios, 19ml will be made up of 0.01M $CaCl_2$. 0.01M $CaCl_2$

was used as opposed to deionised water as it minimises cation exchange and improves centrifugation (OECD 2000). Although not explicitly stated in the OECD guidelines, a reason for minimising cation exchange may be to create a more stable soil system, as without the Ca^{2+} ions there is the potential for the clay structure to become too hydrated and swell (de Jonge and de Jonge, 1999). Divalent cations such as the Ca^{2+} ions, although very hydrated have a small hydrated radius compared to monovalent ions. Being a hydrated ion has the effect of increasing its size and reducing its mobility so it is strongly adsorbed to a negatively charged surface. Whereas the small hydrated radius means there are less water molecules present to fill the interlayer spaces, stabilising the clay structure (Hiscock, 2005). The adsorption step was given as 24 hours for all three compounds.

Syngenta reports for Hexaconazole showed that there was adsorption of this compound onto the Teflon tubes to be used in the adsorption experiment. However, in the presence of soil, Hexaconazole preferentially adsorbed onto the soil (Oliver and Kuet, 1999). Therefore, Teflon tubes were still appropriate for this study. Teflon tubes were also appropriate for Glyphosate and AMPA as these compounds showed no signs of adsorption on to the tubes (Thomas and Lane, 1996). For each soil type there were two replicates and for each experiment a blank tube was included. The blank tube contained only the soil and $CaCl_2$ mix. This blank tube was used to check the analytical method and for matrix effects caused by the soil (OECD 2000). For the Glyphosate and AMPA experiments a control tube was also used. The control tube only has the aqueous solution with no soil and was used to confirm that there was no adsorption of the two compounds onto the Teflon tubes. For all three compounds, 1g or 2g of the appropriate soil was weighed into each tube, then 19mls 0.01M $CaCl_2$ was added. The lids were applied and once secured the samples were placed on an end-over-end shaker for 24hrs for pre-equilibration.

Preparation of Treatment Solution

Each chemical came with a dispense sheet that provided the radioactivity data for the chemical. The radioactivity data is important because it identifies the levels of radioactivity that has been added to the compound and needs to be known to be able to correctly dilute the compound to use in the adsorption experiment. The compounds had been labelled with ^{14}C. The radioactive ^{14}C isotope acted as a tracer, which meant that it was easy to see whether the compounds were mostly present in the soil phase or the aqueous phase.

The level of radioactivity determines if a serial dilution of the stock was needed to make the treatment solution. A serial dilution step was used to check that the activity dispensed value given with the chemical is correct. It was important to know the correct activity of the chemical because it was used to calculate the size of the aliquot taken from the stock solution to use in the treatment solution. The activity was quantified by Liquid Scintillation Counting (LSC). Also, serial dilution may have to occur so any aliquots taken are at an appropriate concentration for the LSC machine to read. As the Glyphosate had a relatively low concentration it could be used in the LSC machine without needing a serial dilution first. However AMPA and Hexaconazole had higher activity and needed a serial dilution. Full details of the serial dilution calculations are in Appendix B – Folder 1.

Glyphosate

Without the need for a serial dilution, the Glyphosate treatment solution could be made directly from the chemical that was provided. The Glyphosate radiochemical had arrived in liquid in a glass vial. Using a pipette the Glyphosate was transferred into a clean 100ml volumetric flask. The size of the flasks had been chosen based on the number of tubes that had to be treated with each chemical. This pipette was referred to as 'hot' as it had been in direct contact with the radio labelled Glyphosate. Using a different pipette, 1-2mls of $CaCl_2$ was pipetted into the vial that contained the Glyphosate to rinse around the walls of the vial. This pipette was the 'cold' pipette as it should not have had any contact with the radio labelled Glyphosate. With the hot pipette, the liquid that was used to rinse the walls was transferred back into the volumetric flask. Using the cold pipette, more $CaCl_2$ was washed around the walls of the vial, and then transferred back to the volumetric flask using the hot pipette. This process was repeated to ensure any radiochemical on the walls of the vial was washed into the 0.01M $CaCl_2$ and none was lost. This rinsing was repeated until 100ml volume of the flask is reached. The Gyphosate treatment solution was now at a concentration of 20,000Bq/ml.

AMPA

The radioactivity of AMPA was higher than Glyphosate and would have been too concentrated to use in the LSC machine so had to be diluted. A stock solution had to be prepared first. To make the stock 10MBq of AMPA was diluted into a 10mls of ultra pure water giving a stock concentration of 1MBq/ml (or 1000Bq/µl). Ultra pure water was used in the stock creation process to dissolve the chemical as, unlike Glyphosate, the AMPA had arrived at the lab in the form of a solid. The stock was prepared by using the hot and cold pipette technique to rinse and transfer the dissolved AMPA into the volumetric flask.

To make the treatment solution an aliquot of the stock solution had to be taken and diluted with 0.01M $CaCl_2$. The calculation and quantifications in appendix B show how the volume of the aliquot of stock was determined. It was calculated that 3.4ml of stock should be taken. Using a pipette 3 x 1ml aliquots of the AMPA stock were taken and dispensed into a clean 50ml volumetric flask. Then a 400µl aliquot of the stock was taken and dispensed into the 50ml volumetric flask. The volumetric was made up to the 50ml line with 0.01M $CaCl_2$ giving the AMPA treatment solution a concentration of around 75,000Bq/ml.

Hexaconazole

As with AMPA, Hexaconazole also had to be made into a stock solution prior to experimental use. The reason for making a stock was that the Hexaconazole radiochemical arrived dissolved in 2mls acetonitrile, and therefore had to undergo a solubility check to ensure it was appropriate to use with the $CaCl_2$ in the experiment. To make the stock, Hexaconazole was diluted into 10mls of acetonitrile giving the stock a concentration of 230,000Bq/ml (230Bq/µl). A solubility test was made (appendix B) and based upon the result it was decided that it was appropriate to use with the $CaCl_2$ and the treatment solution could be made. The Hexaconazole treatment solution was prepared in the same way as the AMPA treatment solution by calculating the volume of stock that was needed to be diluted (appendix B). It was calculated that 3.5ml of the stock was needed. Using a pipette 3 x 1ml aliquots of the stock were taken and dispensed into a clean 100ml volumetric flask. Then a 500µl aliquot was taken and dispensed into the 100ml volumetric flask. The volumetric was made up to the 100ml line with 0.01M $CaCl_2$ giving the treatment solution a concentration of around 9000Bq/ml. After each treatment solution was made it was quantified by LSC to ensure the concentration was correct before applying it to the samples (Appendix B – Folder 1_Hexaconazole).

Treating and Quantifying the Samples

Further quantifications of each of the treatment solutions were made during the process of treating the samples. A sample of each treatment solution was taken before the tubes were treated, once during the treatment process and then once after all the tubes were treated. The reason for quantifying during the treatment process was to test exactly how much radioactivity had been applied to the tubes and also to check that there was continuity in the pipetting technique (Appendix B – Folder 1). In each experiment, all the tubes were treated with 1ml of the treatment solution, except for the blanks. The tubes were placed back in the end-over-end shaker for a planned 24hours for the adsorption step.

After 24 hours, the tubes were removed from the shaker and centrifuged for 10 minutes at 3000rpm. This centrifugation allowed the soil to settle at the bottom of the tube and the supernatant to be sampled. The supernatants were removed from the tubes with a pipette and transferred into individual labelled glass vials for storage. From each of the glass vials a 2 x 1ml aliquot of the supernatant was removed with pipette and dispended into separate large LSC vials. 10mls of scintillation fluid was added to each vial. The concentrations of compounds in the vials were measured overnight by LSC. For a full list of the counts see Appendix B – Folder 1. Using the Bq counts from the LSC, the K_d and K_{oc} values were calculated for each compound. The results of the Glyphosate and AMPA study are given in Chapter 4 and the results for Hexaconazole in Chapter 5. The full adsorption spreadsheets for each compound can be found the relevant files in Appendix B – Folder 1.

For Hexaconazole, additional analysis of the aqueous solution was made using High Pressure Liquid Chromatography (HPLC) and the soil phase using Thin Layer Chromatography (TLC) before calculating K_d and K_{oc}. The reason for the extra analysis was because the Compound A experiment featured new soils that had not been used in a batch equilibrium study before and there was a risk that degradation of the compound may have occurred. The HPLC and TLC methodology is described fully in Chapter 5.

3. PRELIMINARY RESULTS –ALL DATA

3.1 Introduction

There have been many studies attempting to predict K_{oc} based on a compounds molecular descriptors (*e.g.*, Gramatica *et al*., 2000), connectivity indices (*e.g.* Baker *et al*., 2001) and fragment data (*e.g*. Meylan *et al*., 1992). This study analyses a wide range of molecular descriptors to predict the adsorption behaviour of a large group of pesticides. Unlike similar work in the literature, this study had access to the original experimental lab reports to obtain the K_{oc} data.

The full dataset is comprised of 700 K_{oc} values from 80 compounds; this includes the parent compounds and their relative metabolites. When the compounds are assigned to their chemical groups (according to The Pesticide Manual (Tomlin, 1997)) the full dataset has 17 different chemical groups. This full dataset will be referred to as 'Dataset 1'.

3.2 Study Approach

In order to understand the variation in K_{oc} across the range of compounds, this study used a range of multivariate statistical techniques. The study focused firstly on analysing the complete dataset of 80 compounds (Dataset 1). The techniques used to analyse Dataset 1 were principal component analysis and analysis by chemical class with the methodology as described in Section 2.2.1. The results of the data analysis are presented and discussed. The preliminary analysis of Dataset 1 identified that the compounds fell into two distinct trends and therefore modelling the full range of compounds was unsuccessful.

Based on the preliminary results, Dataset 1 was split into two groups based on the two trends identified in the analysis. The split was made in an attempt to produce more suitable models. This study then focused on what has been referred to as 'Group A' compounds (Dataset 2). The modelling techniques to analyse the Group A data were multiple regression and logistic regression. The methodology was outlined in Sections 2.2.2 and 2.2.3.

3.3 Results

3.3.1. PCA: Connectivity Parameters and Molecular Fragments

Table 3.1 The first five principal components for connectivity parameters and molecular fragments with the cumulative proportion of variance explained. Eigenvalues: PC1: 26.04, PC2: 3.51, PC3: 1.75, PC4: 1.045, PC5: 0.93.

	PC1	**PC2**	**PC3**	**PC4**	**PC5**
K_{oc}	0.005	0.015	-0.065	-0.866	-0.444
Ki0	0.183	-0.134	-0.083	0.064	-0.005
Ki1	0.187	-0.103	-0.115	0.026	0.007
Ki2	0.190	-0.080	-0.530	0.033	-0.005
Ki3	0.189	-0.125	-0.045	0.019	-0.024
Ki4	0.189	-0.128	-0.023	0.011	-0.027
Ki5	0.189	-0.135	-0.002	0.006	-0.030
Ki6	0.189	-0.126	0.037	0.008	-0.045
Ki7	0.189	-0.132	0.051	-0.004	-0.035
Ki8	0.190	-0.114	0.058	-0.027	-0.020
KiCP30	0.188	0.027	-0.140	0.030	-0.027
KiCP31	0.189	0.006	-0.107	-0.033	-0.002
KiCP32	0.192	-0.051	-0.007	-0.002	-0.008
KiCP33	0.193	-0.037	-0.026	-0.028	0.001
KiCP34	0.194	-0.027	0.018	-0.013	-0.030
KiCP35	0.193	-0.064	0.047	-0.016	-0.007
KiCP36	0.192	-0.067	0.098	-0.009	-0.014
KiCP37	0.190	-0.085	0.108	-0.029	0.012
KiCP38	0.188	-0.109	0.109	-0.036	0.009
KiCP39	0.186	-0.122	0.114	-0.042	0.017
KiCP40	0.126	0.299	-0.165	0.066	-0.091
KiCP41	0.112	0.356	-0.196	-0.100	0.062
KiCP42	0.157	0.290	0.000	-0.017	0.055
KiCP43	0.170	0.234	-0.023	-0.067	0.076
KiCP44	0.138	0.357	0.078	0.004	0.035
KiCP45	0.145	0.317	0.037	0.042	0.013
KiCP46	0.128	0.313	0.174	0.083	-0.025
KiCP47	0.141	0.234	0.086	0.016	0.043
KiCP48	0.180	0.136	0.053	0.015	-0.013
KiCP49	0.183	-0.017	-0.017	-0.029	-0.024
C1C1C	0.056	-0.027	-0.630	-0.109	0.180
C1C	0.172	0.056	0.180	-0.022	0.058
O1C	0.156	-0.183	0.128	0.012	0.074
C1O1C1C	0.088	-0.103	-0.489	0.023	0.307
C1C1C1C	0.164	-0.130	0.133	-0.018	0.046
C2C1C	0.049	0.010	-0.254	0.440	-0.795
Variance Explained (%)	72	82	87	90	93

The eigenvalues suggest there are five principal components that are worth exploring. The first five principal components explain 93% of the variance in the data (Table 3.1). The first principal component shows a positive loading for KiCP34: *0.194* (the 3rd order cluster, 4th order path connectivity). The second principal component shows a positive loading for KiCP41: *0.356* (the 4th order cluster, 1st order path connectivity). The third principal component has a high negative loading for the molecular fragment C1C1C: *-0.630*. The fourth principal component has a high negative loading for K_{oc}: *-0.866*. The fifth principal component shows a high negative loading for the molecular fragment C2C1C: -0.795. By looking at the loadings on the first five principal components, it suggests a combination of connectivity parameters and molecular fragments as being important for K_{oc}.

The higher order path and cluster connectivity parameters are associated with the structural complexity of the molecule like the degree of branching (Worrall and Thomsen, 2004). Increased branching will also restrict microbial degradation (Worrall, 2001). The molecular fragments C1C1C and C2C1C (an alkane and an alkene) can be related to solubility.

As the aim of this study was to understand the controls on K_{oc}, the two principal components with the highest loadings on K_{oc} were selected to make a scatter plot so any trends could be visualised (Figure 3.1). Figure 3.1 represents the interaction between K_{oc} and molecular fragments C2C1C and C1C1C. The data points plotted in the scatter graphs from Minitab are the scores from the PCA. This means that the values in the tables have been transformed and the data points are representative of the variables and not the actual values of the variables.

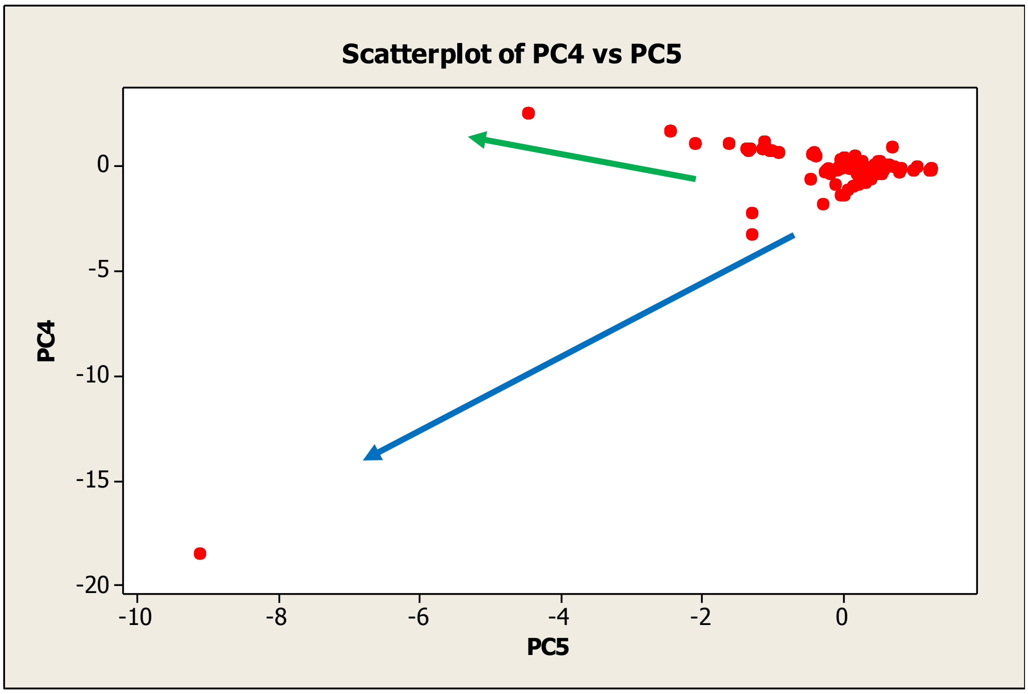


Figure 3.1 Scatter plot of principal component 4 versus principal component 5 for connectivity parameters and molecular fragments. Blue arrow demonstrates direction of negative trend defined by K_{oc}. Green arrow demonstrates influence on data from alkene molecular fragment.

As K_{oc}, on the fourth and fifth principal components has a negative value (Table 3.1), the data will follow this negative trend, as shown by the blue line in Figure 3.1. However, the molecular fragment C2C1C has a positive loading for principal component 4 that is causing other compounds to move away from the original negative trend (as shown by the green line). The interaction between the loadings and variables in the two principal components can be represented in Figure 3.1. The blue line (Figure 3.1) shows the direction of the negative trend created by the loadings on the variables. The blue line is therefore representing a mix of K_{oc} and the C1C1C and C2C1C molecular fragments. It would be expected that the highest K_{oc} values would be along this line. The green line is showing the effect of the positive loading on C2C1C for principal component 4, resulting in a 'V' shaped split in the data points.

3.3.2 PCA: Soil Properties

Table 3.2 The first four principal components for soil properties with the cumulative proportion of variance explained. Eigenvalues are: PC1: 3.68, PC2: 1.62, PC3: 1.03, PC4: 0.98.

	PC1	**PC2**	**PC3**	**PC4**
K_{oc}	-0.003	0.000	0.684	0.729
% OM	0.406	-0.479	0.012	-0.003
%OC	0.400	-0.476	0.004	0.000
%Sand	-0.427	-0.443	0.040	-0.038
%Silt	0.381	0.394	-0.108	0.072
%Clay	0.366	0.383	0.079	-0.029
pH	-0.046	-0.049	-0.716	0.680
CEC	0.460	-0.210	-0.019	0.004
Variance Explained (%)	46	66	79	91

The eigenvalues suggests there are four principal components worth exploring. The first four principal components explain 91% of the variation in the data (Table 3.2). Principal component 1 has a negative loading on % sand content: *-0.427* and positive loadings on organic carbon content: *0.400* and organic matter content: *0.406*. The second principal component has negative loadings for organic matter content: *-0.479* and organic carbon content: *-0.476*. The third principal component shows a high negative loading for soil pH: *-0.716* and a high positive loading for K_{oc}*: 0.684*. The fourth principal component shows high positive loadings for both pH: *0.680* and K_{oc}: *0.729*. As above, the two principal components with the highest loadings on K_{oc} were selected to make a scatter plot so any trends could be visualised. This produced a scatter plot of PC 3 v PC4 (Figure 3.2).

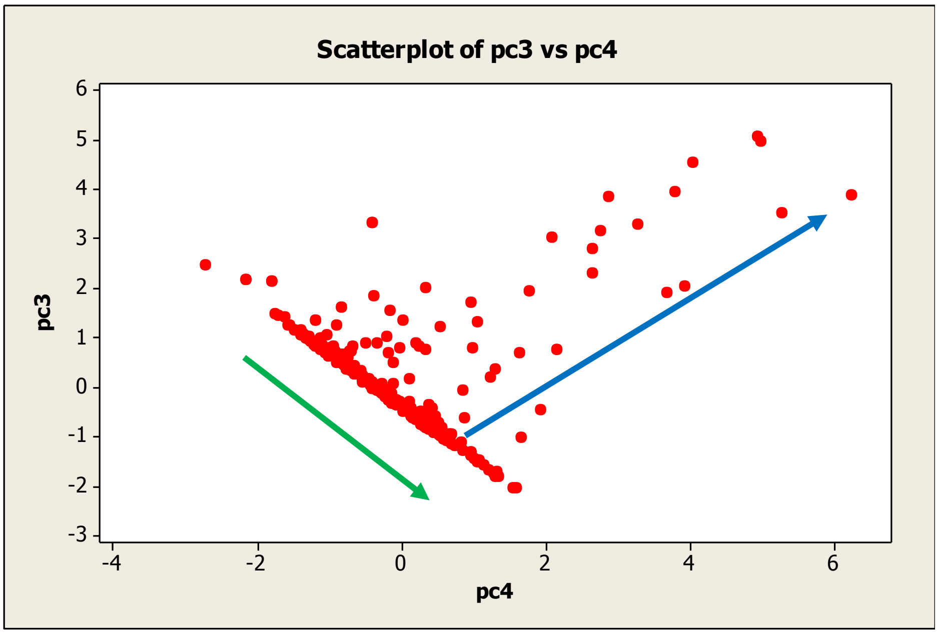


Figure 3.2 Scatter plot of principal component 3 versus principal component 4 for soil properties. Blue arrow represents direction of trend defined by K_{oc}*. Green arrow represents the influence of pH on the data.*

Figure 3.2 visualises the interaction between K_{oc} and pH. The blue line in Figure 3.2 shows the positive trend that represents K_{oc}. The green line is showing the effect the negative loading for pH has on the data points. The positive and negative loadings for PC 3 seem to be showing that there is almost a ‘barrier’ in the data (along the green line) where no data points seem to plot below it. As PC 3 relates to soil pH Figure 3.2 is suggesting that soil pH is acting as a control on K_{oc} and is linked to higher K_{oc} values.

3.3.3 PCA: Molecular Properties

Table 3.3 The first five principal components for molecular properties with the cumulative proportion of variance explained. Eigenvalues are: PC1: 9.37, PC2: 2.83, PC3: 1.49, PC4: 1.18, PC5: 0.98

	PC1	PC2	PC3	PC4	PC5
K_{oc}	0.024	-0.090	-0.001	0.766	0.320
dipolm	0.100	0.074	-0.031	-0.231	-0.350
totalE	-0.313	-0.145	0.017	0.039	-0.004
NHOMO	0.211	-0.351	0.058	0.005	-0.110
HOMO	0.188	-0.307	0.028	0.352	-0.009
LUMO	-0.137	0.361	0.502	0.145	0.039
NLUMO	-0.153	0.446	-0.127	0.177	0.044
V_{sav}	0.315	0.154	-0.015	0.001	0.026
A_{sas}	0.303	0.169	-0.115	0.047	0.050
V_{vdW}	0.314	0.155	0.009	-0.007	0.019
A_{vdW}	0.312	0.166	-0.019	-0.004	0.019
ΔHhyd	-0.069	0.289	-0.601	0.132	0.059
logP	0.216	-0.139	0.432	0.035	0.032
refractivity	0.318	0.123	0.040	-0.015	0.023
polarisability	0.317	0.120	0.049	-0.008	0.027
mass	0.315	0.132	-0.032	-0.004	0.016
Variance Explained (%)	52	68	76	83	88

The first five principal components explain 88% of the variation in the data. Principal component 1 has positive loadings for refractivity: *0.318* and polarisability: *0.317*. Principal component 2 has a positive loading on NLUMO: *0.446*. Principal component 3 has a high negative loading for ΔHhyd: -*0.601*. Principal component 4 shows a high positive loading for K_{oc}: *0.766*. Principal Component 5 has a negative loading on the dipole moment: *-0.530*. The results of the PCA in Table 3.3 did show that some of the eigenvalues were very close together for example, mass, V_{sav}, V_{vdW}. However, in all cases it was the highest eigenvalues that were selected for importance.

The polarisability, ΔHhyd and dipole moment are related to aqueous solubility. The ΔHhyd is usually inversely related to the aqueous solubility and the polarisability and dipole moment are normally proportional to the solubility (Worrall and Thomsen, 2004). Reddy and Locke, (1994a) found the LUMO to be significant in establishing a relationship with K_{oc}. The high loadings on the principal components for change in hydration energy and the dipole moment suggest that in this case solubility is important for K_{oc}. As above, the two principal components with the highest loadings on

K_{oc} were selected to make a scatter plot so any trends could be visualised. This produced a scatter plot of PC 4 v PC5 (Figure 3.3).

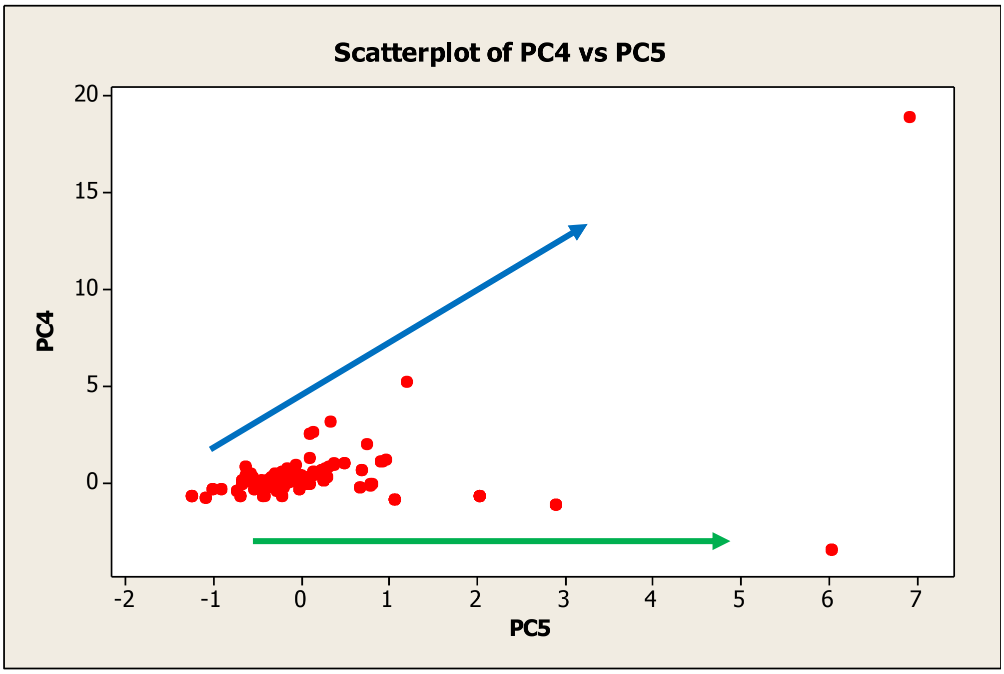


Figure 3.3 Scatter plot of principal component 4 versus principal component 5 for molecular properties. Blue arrow represents direction of positive trend influenced by HOMO parameter. Green arrow represents the positive trend influenced by dipole moment parameter.

The arrows indicate the split in the data points created by the interaction between the loadings on the variables. The blue line shows the influence the positive loading for HOMO has on K_{oc}. The green line shows the influence the positive loading for the dipole moment has on K_{oc}. The result is a positive 'V' shaped split in the data. This split implies that there are two groups of compounds, being controlled by different molecular properties. One group is controlled by HOMO and one controlled by the dipole moment.

3.4 Analysis by Chemical Group

The principal component analysis found that there were multiple trends present in the data set. To determine which compounds lay on which trend the scores from the PCA were plotted by chemical category as well as by category of molecular descriptor. The compounds were categorised according to their chemical group, as defined in The Pesticide Manual (Tomlin, 1997) or the original Syngenta report. Each chemical group was then added as an individual series in the graph. This created copies of the scatter plots from Figures 3.1, 3.2 and 3.3 but sorted by chemical group type. The red and blue circles in Figure 3.4, 3.6 and 3.8 have been added to show the general locations of the groups of compounds that are showing the separate trends identified in the PCA.

3.4.1 Connectivity Parameters and Molecular Fragments

Figure 3.4 shows the interaction between K_{oc} and the molecular fragments C1C1C and C2C1C when broken down into the different chemical group types. The original scatter plot (Figure 3.1) showed that there was a split in the data set; from analysis by chemical group type it identifies the composition of the two trends, which have been identified in Figure 3.4 by the red and blue circles. By thinking of the compounds within these two trends as two different groups, the scatter plot can be simplified (Figure 3.5).

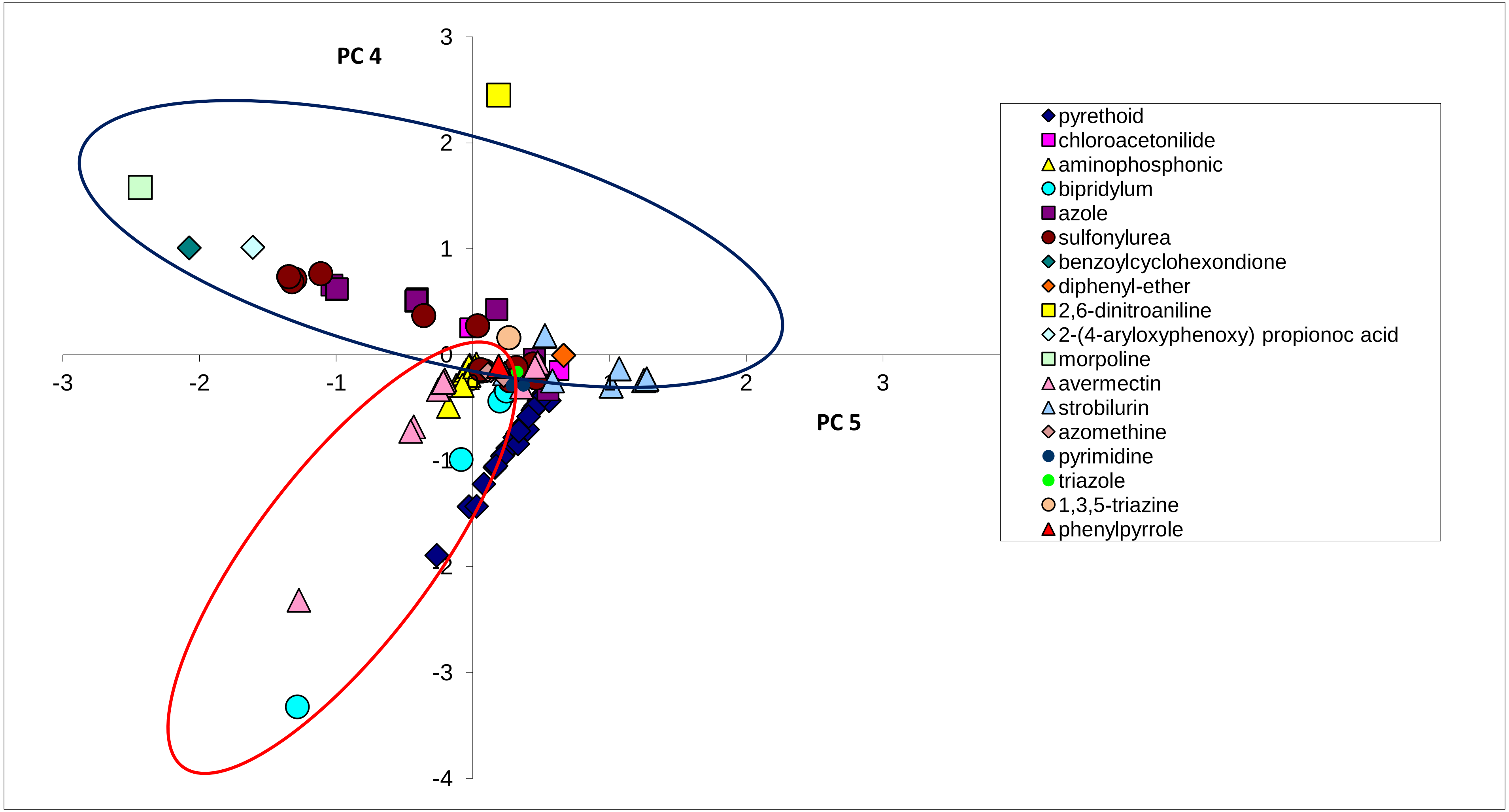


Figure 3.4 Scores from PCA plotted by chemical group for connectivity parameters and molecular fragments, with groupings illustrated as per Figure 3.5.

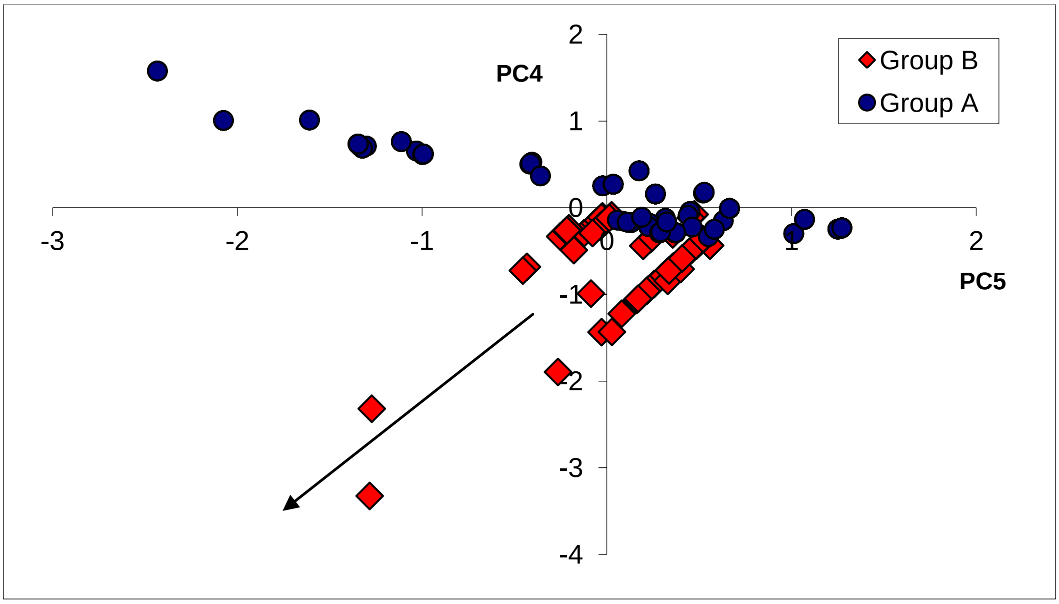


Figure 3.5 Scores from PCA plotted by chemical group for connectivity parameters and molecular fragments simplified into two groups. Arrow indicates direction of increase in K_{oc} as identified by loadings from PCA.

In Figure 3.5 the main trend, shown by the blue circles, will be referred to as Group A. Plotting away from the main trend is the second trend, which will be referred to as Group B. Group A comprises the majority of the dataset and a diverse range of chemical groups. Group B comprises only compounds from the pyrethoid, aminophosphonic, bipyridylium and avermectin chemical groups. The compounds in these four chemical groups are displaying an adsorption behaviour that is different to the majority of the compounds in the data set. The PCA would suggest that the presence of particular molecular fragments is important (Table 3.1).

Table 3.1 has identified that the molecular fragment C2C1C, an alkene, has having a high negative loading on principal component 5. The negative loading relates to the direction of the trend in the scatter plots, so in Figure 3.5 the trend for alkenes plots in the direction of the left half of the graph. An alkene fragment will decrease the solubility of the compound (Lloyd, 1989) and with decreased solubility, adsorption will increase. By considering the influence of the C2C1C fragment, Figure 3.5 suggests that the compounds with the lowest solubility are found in the negative portions of the graph, which is dominated by the Group B compounds. It is important to note that the negative portions of Figure 3.5 also includes some of the Group A compounds, at the very end point of the Group A line, which has been caused by the positive loading for C2C1C on PC 4. Figure 3.5 shows that based on the loadings from the PCA, the Group B compounds are plotting in the direction of the highest K_{oc} values (as indicated by the arrow). Therefore, Figure 3.5 would suggest that the Group B compounds are more adsorbing than the majority of the Group A compounds, because the Group B compounds are more insoluble.

However, when comparing the original fragment data obtained by Topix for the compounds, the average number of alkene fragments in the Group B compounds was the same as the Group A compounds. Individual compounds will have different fragment counts but generally speaking the original data would imply that although Group B compounds are more adsorbing than the majority of the Group A compounds, as shown by the higher K_{oc} values, the alkene fragment is not controlling adsorption in this case. In Figure 3.5 the majority of the data points from both groups are clustered around the lower right portion of the graph. If this grouping of data points were interpreted in terms of the eigenvalues from Table 3.1 then this data cluster may also correspond to the C1C1C molecular fragment.

3.4.2 Soil Properties

When the data for the soil properties are plotted by their chemical groups it produces two groups of data points (Figure 3.6). The data points show similarities with the connectivity parameters and molecular fragments graph (Figure 3.4). The main cluster of data points, blue oval, again comprises the majority of compounds and a range of chemical groups in the data set. There is again also a sub group of compounds, red oval, trending away from the main group of data. Like the previous figure, the subgroup is again only comprised of compounds from pyrethoid, aminophosphonic and avermectin chemical groups.

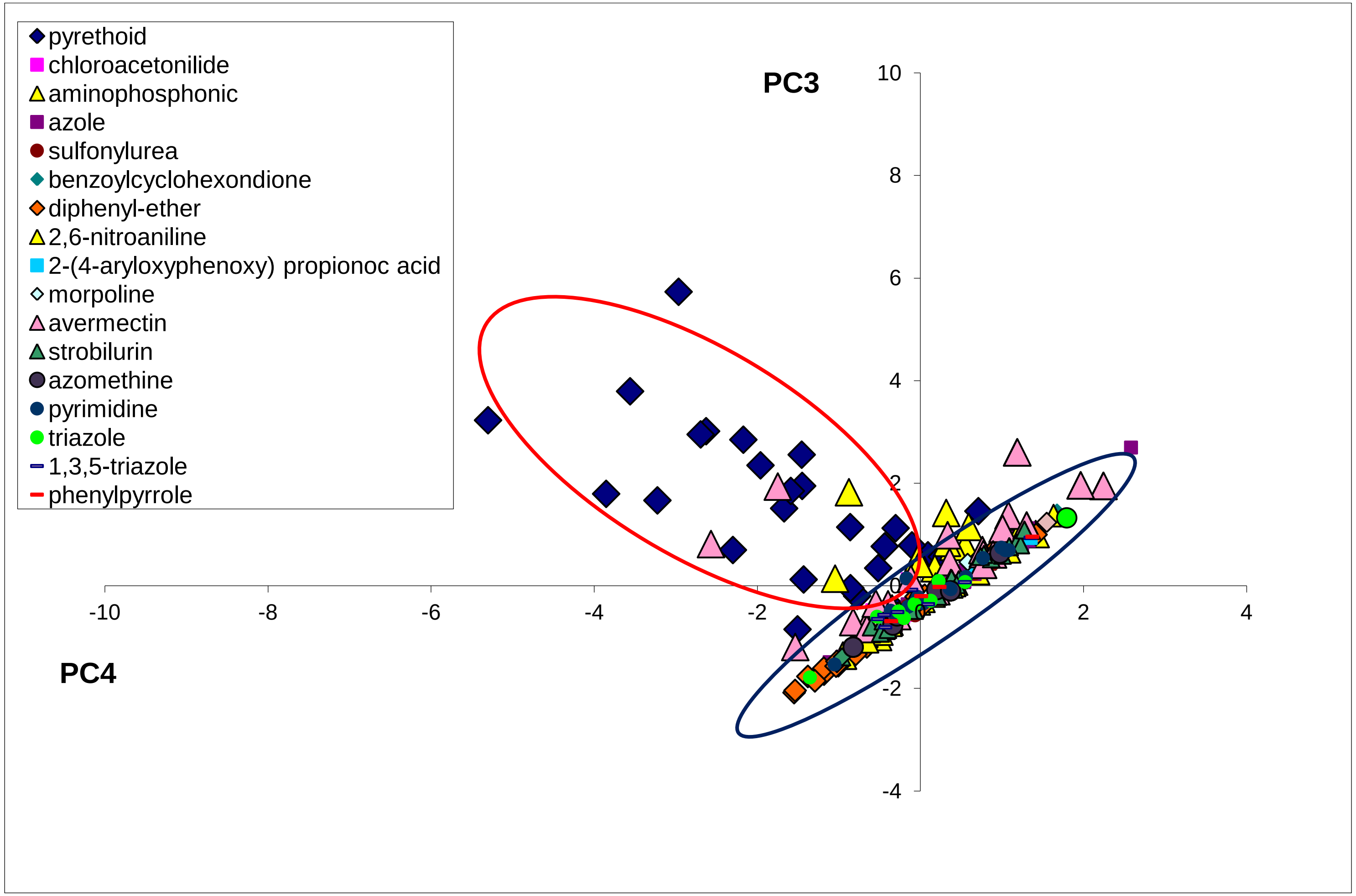


Figure 3.6 Scores from PCA plotted by chemical group for soil properties, with groupings illustrated as per Figure 3.7

To simplify the scatter plot, the individual chemical groups in the two trends can be grouped together (Figure 3.7). The main trend in will again be referred to as Group A and the sub group of compounds in will be referred to as Group B.

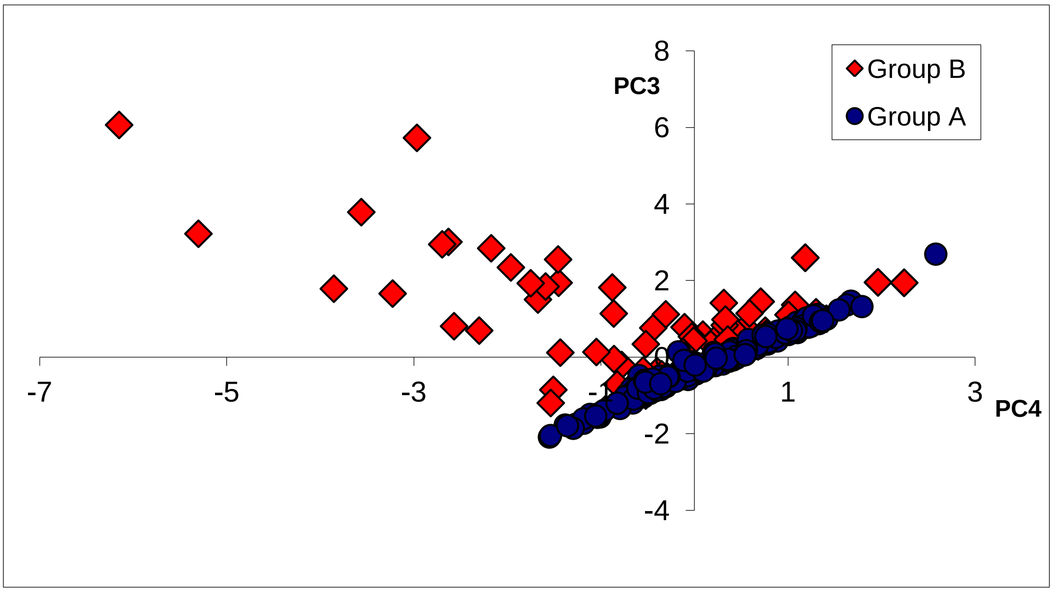


Figure 3.7 Scores from PCA plotted by chemical group for soil properties, simplified into two groups

By plotting the data as just two groups of compounds it shows that there might be another mixing trend in the dataset. Some of the data points for Group B appear to plot within the Group A trend and the chemical groups are more spread out along the Group B trend than they were in the connectivity graph. For example, there are two compounds both within the avermectin chemical group, but one is found at both ends of the Group B line and the other is more mixed in the Group A trend.

The soil types at the extremes of the lines don't seem to show any particular trend relating to the Group A and Group B compounds. The end points of the lines are a mixture of sands, silts, clays and loams. This may explain why some of the compounds are so spread out along the trends. The influence of soil type is enough to create a mixing type trend, but the influence is not strong enough to define the trends, and therefore the trends are still based on chemical group type. So although Figure 3.7 shows the interaction of pH and K_{oc}, from the PCA it would suggest that any difference in adsorption behaviour is more likely to be due to a structural feature in the compound as opposed to a soil property alone.

3.4.3 Molecular Properties

Figure 3.8 shows that like the previous graphs of the scores plotted by chemical group type, there are again two trends. The main trend in the blue oval is the majority of chemical groups and the secondary trend within the red oval are again the pyrethoid, aminophosphonic, bipyridylium, and avermectin chemical groups. As these four particular chemical groups appear in the secondary trends of all three categories of analysis suggests that it is more likely that any difference in adsorption behaviour between these four chemical groups and the rest of the data set is due to molecular properties as opposed to just an effect of soil type.

Figure 3.8 also shows that there is a clear difference in adsorption behaviour within the secondary trend as well as a difference from the main trend. To investigate this further the scatter plot has been simplified (Figure 3.9).

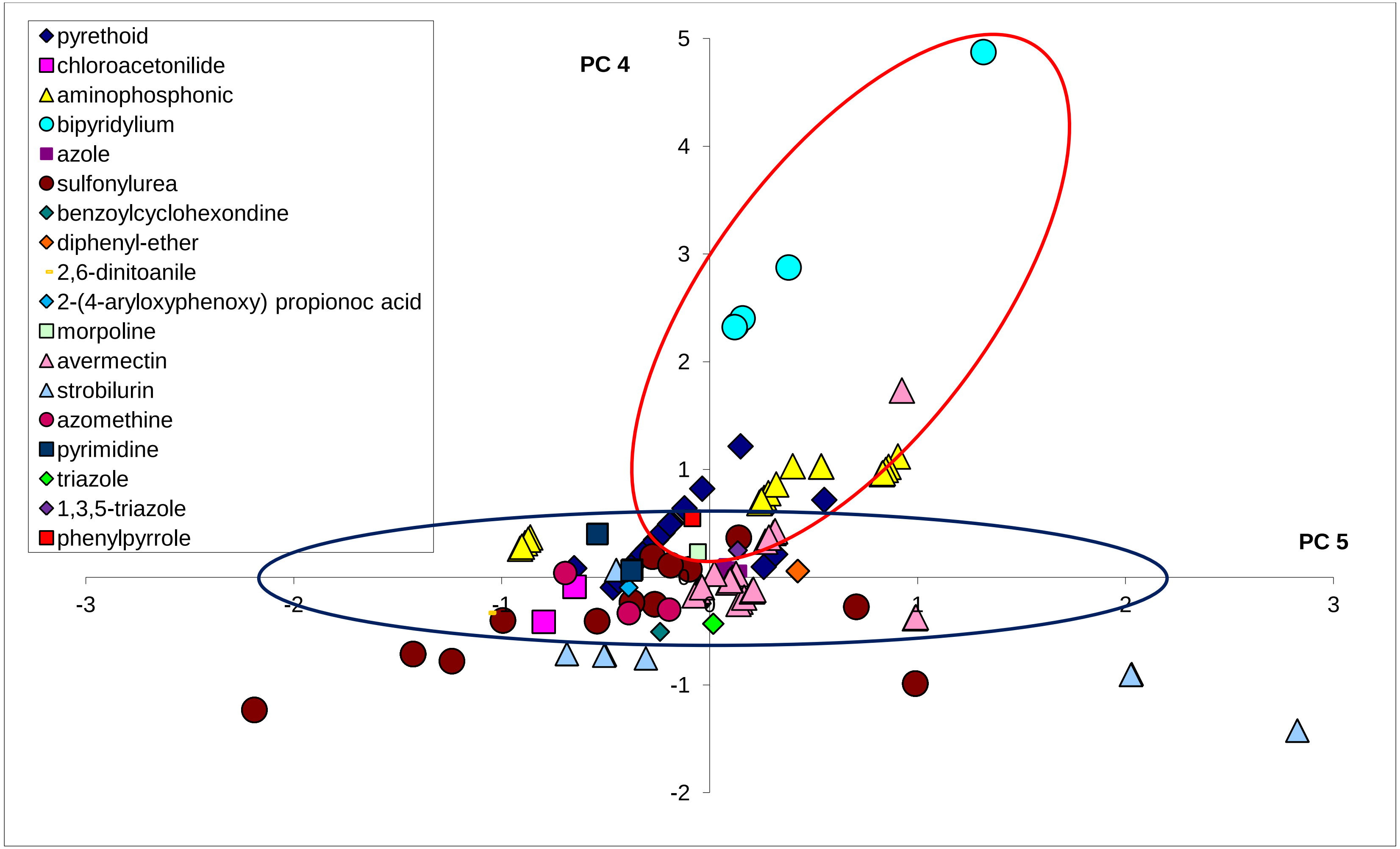


Figure 3.8 Scores from PCA plotted by chemical group for molecular properties, with red and blue groupings illustrated in Figure 3.9.

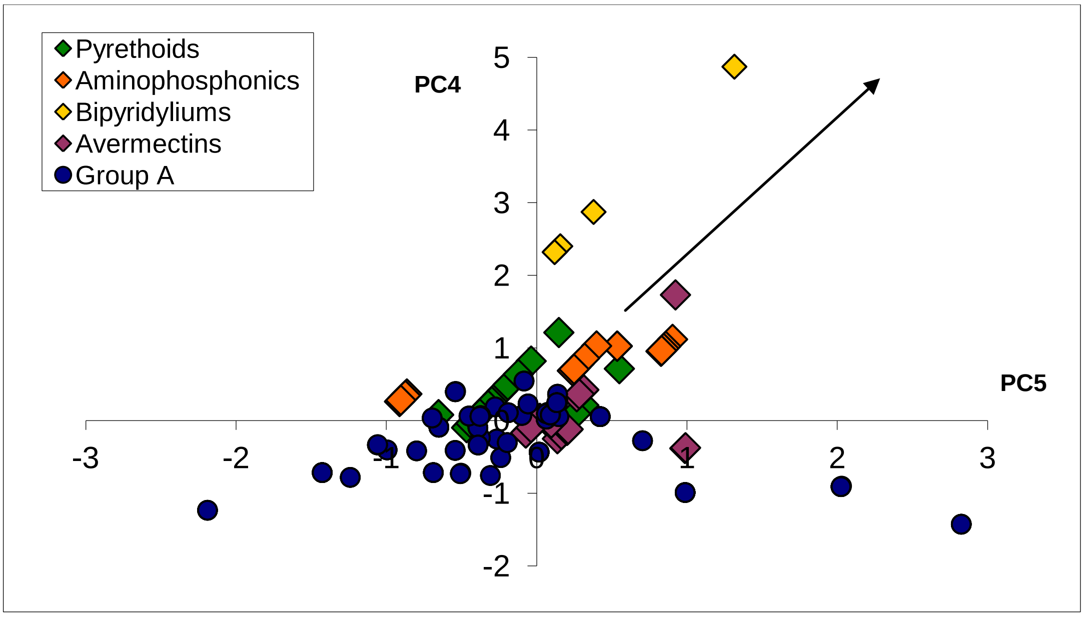


Figure 3.9 Scores from PCA plotted by chemical group for molecular properties, simplified into two groups. Group B chemical groups are separated into their individual groups. Arrow represents the direction of increasing K_{oc} values, as indicated by PCA loadings.

By plotting the molecular properties by chemical group type it again demonstrated the two trends in the data. Like the previous graphs, the composition of Group A and Group B remained the same. Figure 3.9 shows these two trends; however the Group B compounds have now been plotted as their individual chemical groups, represented by the different coloured diamonds.

What Figure 3.9 demonstrates is the complexity of the adsorption behaviours within the full data set. There is the difference in adsorption behaviour between the main group and the subgroup, as already shown by the results for the connectivity parameters, molecular fragments and soil properties. What is clearer in Figure 3.9 is the Group B compounds are showing adsorption behaviour that is different not only to the compounds in Group A, but also different to each other. Within Group B the four chemical groups are showing a linear separation. It would suggest that there is a feature in the structure of the Group B compounds that is causing the difference from the main trend, but also a feature that is causing the difference from each other. From looking at the results of the PCA it would suggest that in this case the dipole moment of each compound is important as the dipole moment has the highest loadings. The dipole moment relates to the solubility, which would have an effect on adsorption behaviour.

Worrall and Thomsen, (2004) have indicated that the dipole moment is normally proportional to the solubility. An increase in solubility will lead to lower adsorption. From the PCA results, the dipole moment has a negative loading for both principal component 4 and principal component 5. Therefore, the compounds with the higher solubility (and lower K_{oc} values) are clustered around the origin and towards the left of the graph in Figure 3.9. What Figure 3.9 is suggesting is that the Group B compounds have the higher K_{oc} values compared to the Group A compounds, because they are less soluble. When comparing the original data, the average dipole moments for both the Group A and Group B compounds are similar, but the average changes in hydration energies are different. So the original data indicates that the Group B compounds are more insoluble than Group A due to the hydration energy and in this case solubility is controlling adsorption.

3.5 Group A Compounds

The aim of analysing the Group A compounds (Dataset 2) was to create a predictive K_{oc} model for a range of chemical classes. After discussions with environmental fate scientists at Syngenta's Jealott's Hill site, they identified Group A as being of interest as these compounds represent what is more 'typical' adsorption behaviour, compared to the Group B compounds which demonstrate very high adsorption – some of the compounds in Group B had K_{oc} values >100,000 ml/g. As the majority of compounds in the data fall within the trend for typical adsorption behaviour it was important to them to be able to model and predict the environmental fate of these types of compounds.

The Group A compounds (Dataset 2) were analysed only for connectivity parameters and molecular fragments and molecular properties. The results from the soil analysis for the full dataset suggested that soil parameters were not that important for K_{oc} in this situation. Analysis of Dataset 2 was by PCA combined with analysis of compound type, as described in Chapter 2 (Section 2.2.1). This helped to gain a clearer picture of what was happening in the main trend, as when the Group A trend was included with the full dataset the trend was very noisy and non-linear compared to Group B. Whereas Dataset 1 was analysed by chemical group type, the Group A compounds in Dataset 2 have been analysed by compound type to get a more in depth view of the dataset. Multiple regression and logistic regression were also tried, as described in Chapter 2 (Section 2.2.2 and 2.2.3). The logistic regression model was created for the parent compounds only, as this was the best way of making the two categories of similar size in terms of the number of compounds, which was important for model development.

3.5.1 PCA: Connectivity Parameters and Molecular Fragments

Table 3.4 The first five principal components for Group A compounds, for connectivity parameters and molecular fragments. Eigenvalues: PC1: 21.64, PC2, 6.86, PC3: 1.99, PC4: 1.18, PC5: 0.89.

	PC1	PC2	PC3	PC4	PC5
K_{oc}	0.056	0.022	0.466	-0.316	0.237
Ki0	0.182	0.175	0.113	-0.002	-0.041
Ki1	0.186	0.177	0.062	0.014	-0.030
Ki2	0.203	0.113	0.025	0.040	-0.028
Ki3	0.194	0.158	0.018	0.040	0.008
Ki4	0.192	0.162	0.015	0.071	-0.056
Ki5	0.194	0.151	0.039	0.012	-0.024
Ki6	0.198	0.122	-0.009	0.074	-0.109
Ki7	0.199	0.123	-0.018	0.011	-0.050
Ki8	0.201	0.104	0.005	0.049	-0.031
KiCP30	0.200	0.076	-0.024	-0.118	0.010
KiCP31	0.206	0.034	-0.031	0.105	0.014
KiCP32	0.198	0.043	-0.058	0.235	-0.117
KiCP33	0.208	-0.017	0.001	0.108	0.002
KiCP34	0.202	-0.010	-0.038	0.189	-0.119
KiCP35	0.209	-0.042	-0.062	0.047	0.018
KiCP36	0.203	-0.076	-0.014	0.114	0.005
KiCP37	0.190	-0.156	0.021	-0.036	0.109
KiCP38	0.193	0.070	-0.129	-0.010	0.034
KiCP39	0.153	0.194	-0.218	-0.037	-0.019
KiCP40	0.172	-0.082	-0.123	-0.323	-0.077
KiCP41	0.159	-0.251	0.039	0.011	0.055
KiCP42	0.163	-0.238	0.029	0.059	0.017
KiCP43	0.159	-0.253	0.044	0.000	0.042
KiCP44	0.155	-0.261	0.052	-0.032	0.002
KiCP45	0.153	-0.261	0.038	-0.082	0.026
KiCP46	0.143	-0.267	0.042	-0.073	-0.007
KiCP47	0.129	-0.277	0.091	-0.109	0.116
KiCP48	0.150	-0.249	0.031	-0.138	0.054
KiCP49	0.107	0.185	-0.327	-0.354	-0.141
C1C1C	0.112	0.195	0.280	-0.215	0.339
C1C	0.101	-0.180	-0.059	0.300	-0.174

O1C	0.019	0.086	0.505	0.027	-0.510
C1O1C1C	0.063	0.204	0.395	0.077	0.075
C1C1C1C	0.074	0.163	-0.127	0.279	0.643
C2C1C	0.110	0.096	-0.194	-0.490	-0.083
Variance Explained (%)	60	79	85	88	90

The eigenvalues suggest there are five principal components worth exploring. The first five principal components explain 90% of the variance in the data (Table 3.4). The first principal component has a positive loading on KiCP35: *0.209* (the 3rd order cluster, 5th order path connectivity). The second principal component has a negative loading on KiCP47: -*0.277* (the 4th order cluster, 7th order path connectivity). The third principal component shows positive loadings for K_{oc}: *0.466* and the molecular fragment O1C: *0.505*. The fourth principal component shows negative loadings for K_{oc}: -*0.316* and the molecular fragment C2C1C: -*0490*. The fifth principal component has a high positive loading for the molecular fragment C1C1C1C: *0.643*.

The connectivity parameters can give an indication to the size and branching of the structure. The path and cluster connectivity parameters represent the structural complexity of the molecule, like the degree of branching, which influences the changes in enthalpy and entropy upon aqueous dissolution (Worrall and Thomsen 2004). An increased degree of branching in a molecule will also restrict microbial degradation (Worrall, 2001). Molecular fragments have been identified in other studies as being important to sorption (Lohninger, 1994). The molecular fragments analysed in this study can be related to the type of bonding and solubility of the compounds. In a QSAR model the hydroxyl group O1C, was found to decrease the sorption coefficient (Lohninger, 1994), which would imply that the presence of the O1C fragment in a compound decreases adsorption. As the O1C fragment is polar, the compound may be quite soluble in water, decreasing its ability to adsorb to the soil. As the aim of the study was to understand controls on K_{oc}, the two principal components with the highest loadings on K_{oc} were selected to make a scatter plot so any trends could be visualised, producing a scatter plot of PC3 vs. PC4 (Figure 3.10)

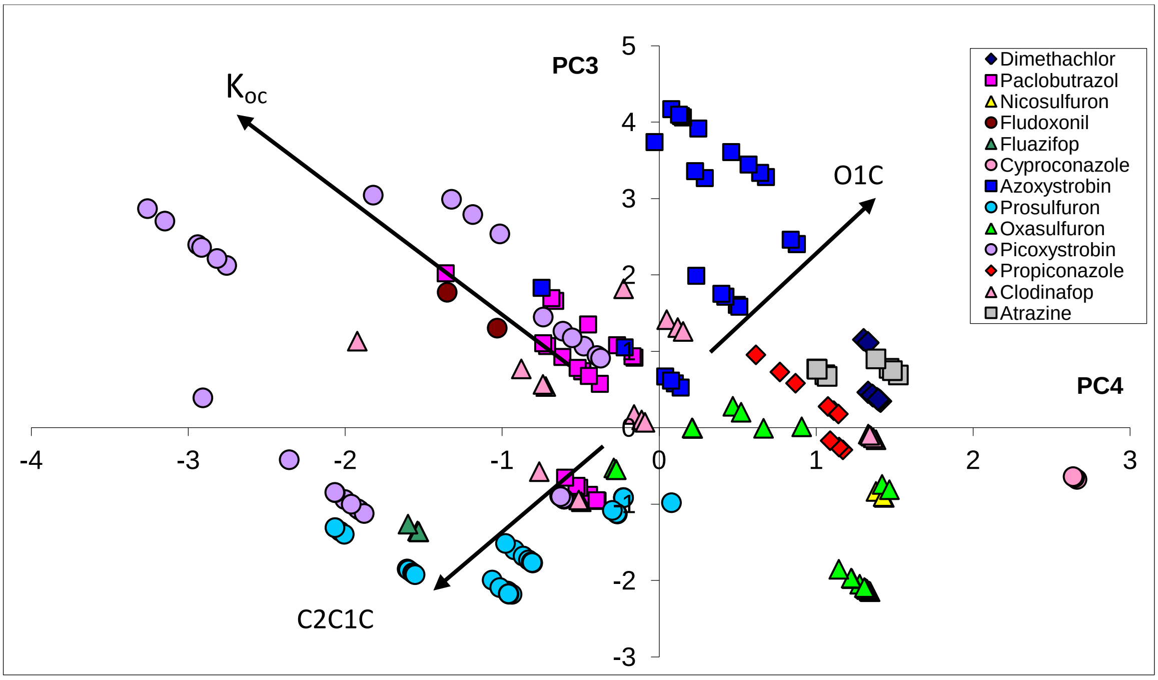


Figure 3.10 Scores from PCA for Group A plotted by compound, for connectivity parameters and molecular fragments. Arrows indicate direction of loadings on variables.

Figure 3.10 shows that when Group A was analysed separately from the rest of the dataset, the Group A compounds also showed linear separation into individual compound types. Figure 3.10 is a visual representation of the interactions of the different variables with K_{oc}. The two variables that had the highest, but opposite loading for principal component 3 and principal component 4 were the molecular fragments O1C and C2C1C.

The positive and negative loadings on the two molecular fragments will have influenced the difference in the compounds, as shown by the two arrows (Figure 3.10). Example molecular structures for some of the compounds discussed here can be found in Table 2.1 in Chapter 2. The compounds in the upper portion of the graph are those that are more likely to have the hydroxyl functional group O1C, while the compounds in the lower portion of the graph are more likely to have the alkene functional group C2C1C. For example, by focusing on Azoxystrobin, Propiconazole and Atrazine in the upper portion of the Figure 3.10, the original data shows these compounds do indeed have a higher O1C values compared to their C2C1C values. For Azoxystrobin and its metabolites have an O1C value of between 1-4, compared to a C2C1C value of 0. In the lower portion of Figure 3.10, Prosulfuron, Oxasulfuron, and Picoxystrobin have higher C2C1C values compared to their O1C values. Picoxystrobin and its metabolites, have a C2C1C value of 3 compared to its O1C value of 0. This suggests that the linear separation of the individual compound types is caused by a difference in the molecular structure of the compounds, and relates strongly to the solubility. This means that the compounds towards the left half of the graph will have higher K_{oc} values and may be more strongly adsorbing than the compounds in the right half of the graph.

3.5.2 PCA: Molecular Properties

Table 3.5 The first four principal components for Group A compounds for molecular properties. Eigenvalues are: PC1: 9.20, PC2: 1.43, PC3: 1.22, PC4: 0.88.

	PC1	PC2	PC3	PC4
K_{oc}	0.136	-0.295	-0.264	0.118
dipolm	0.101	0.209	-0.502	-0.727
NHOMO	0.183	0.043	-0.102	0.294
HOMO	0.211	-0.143	-0.348	0.470
LUMO	-0.141	-0.595	0.389	-0.210
NLUMO	-0.263	-0.395	0.067	-0.174
V_{sav}	0.322	-0.058	0.158	-0.065
A_{sas}	0.308	-0.018	0.169	-0.218
V_{vdW}	0.324	-0.071	0.120	-0.056
A_{vdW}	0.321	-0.040	0.148	-0.119
ΔHhyd	0.055	-0.549	-0.492	-0.057
logP	0.296	-0.077	-0.104	-0.031
refractivity	0.321	-0.082	0.152	-0.024
polarisability	0.322	-0.089	0.086	0.017
mass	0.318	0.067	0.135	-0.037
Variance Explained (%)	61	71	79	85

The eigenvalues suggest there are five principal components to study. The first five principal components explain 85% of the variance in the data (Table 3.5). The first principal component has positive loadings for V_{sav}: *0.322* and V_{vdW}: *0.324*. The second principal component has high negative loadings for LUMO: *-0.595* and ΔHhyd: *-0.549*. The second principal component also has a negative loading for K_{oc}: *-0.295*. The third principal component has a high negative loading for the dipole moment: *-0.502*. The third principal component also has a negative loading on K_{oc}: *-0.264*. The fourth principal component has a high positive loading for the dipole moment: *-0.727*.

The V_{sav} is a size descriptor but in hydrophobic compounds it can be related to the leaching ability of the compound, where it is inversely related to the aqueous solubility (Worrall and Thomsen, 2004). The ΔHhyd is also inversely related to solubility, meaning an increase in the change in hydration energy may lead to an increase in adsorption, whereas the dipole moment is proportional to the solubility (Worrall and Thomsen, 2004). In this case an increase in the dipole moment may lead to a decrease in adsorption. The LUMO is a parameter that can be related to a hydrogen bonding term and the basicity of the compound (Kamlet *et al*., 1987).

As the aim of the study was to understand controls on K_{oc}, the two principal components with the highest loadings on K_{oc} were selected to make a scatter plot so any trends could be visualised. This produced a scatter plot of PC2 vs. PC3 (Figure 3.11)

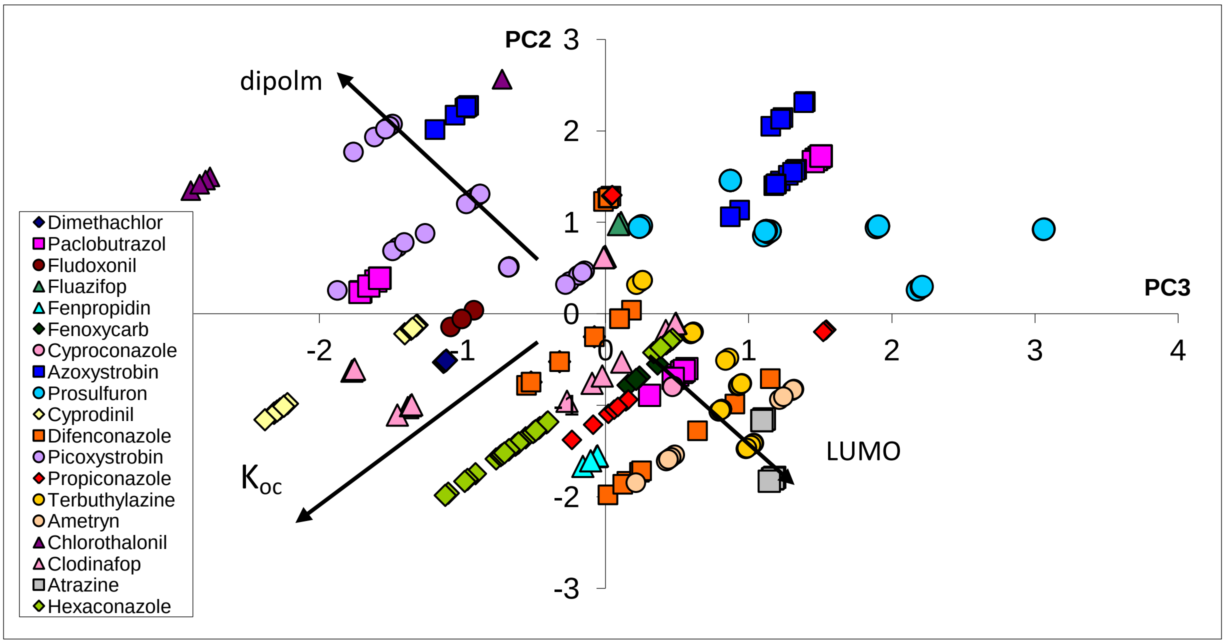


Figure 3.11 Scores from PCA plotted by compound for molecular properties. Note: The reason for the different types of compounds in each graph is due to the data being analysed for different properties so some will be have been removed as outliers by visual inspection during PCA.

Figure 3.11 shows that when the Group A compounds are analysed for molecular properties they also demonstrate a linear separation into individual compounds, example molecular structures for some of the compounds discussed in this section can be found in Table 2.1. Figure 3.11 is a visual representation of the interactions of the different variables with K_{oc}. The two variables that had the highest loadings for principal component 2 and principal component 3 were LUMO and the dipole moment. The negative loadings on the variables K_{oc}, LUMO and the dipole moment have influenced the linear separation of the compounds. The dipole moment is linked to the solubility of compounds, with the dipole moment being inversely proportional to adsorption. This would suggest that the compounds that would be the most soluble would be towards the upper left region of the graph, such as Azoxystrobin and Picoxystrobin, and therefore would have lower adsorption. Azoxystrobin and Picoxystrobin have lower K_{oc} values of around 300-900, in comparison Difenconazole that has K_{oc} values of between 2000-5000.

From using PCA to analyse the Group A compounds separately from the main data set, it showed that the adsorption behaviour of these compounds was more complex than first anticipated. Figures 3.10 and 3.11 would suggest that the presence of alkane and alkene molecular fragments and also the solubility of the compounds are important.

3.5.3 Multiple Regression

If there is a range of K_{oc} values across the data set that are related to specific structural properties in particular compounds, then it would imply that the adsorption behaviour is quite compound specific. This would suggest that a general model to predict K_{oc} that covers a variety of compounds would be hard to produce. To see if any model could be produced, multiple regression was attempted on the Dataset 2, for the categories of connectivity and molecular fragment data and molecular data.

The first runs produced very poor models, in terms of accountable variation in the data. The R-sq value for the connectivity and molecular fragment model was 28.2% and for the molecular properties model was just 24.2%. This meant that around three quarters of the variation in the dataset was unaccounted for. To improve this value, soil properties were included in the model. The PCA from the preliminary results had shown that soil pH was important to K_{oc} for both Group A and Group B compounds so this was added to the model. The addition of pH did improve the model slightly, but still created a very poor fitting model. When the connectivity and molecular properties were combined the best model produced gave an R-sq value of 35.6% (Table 3.6).

Table 3.6 The significant parameters of MLR model for predicting K_{oc} of Group A compounds, the regression coefficients and R^2 value for the model.

Parameter	**Coefficient**	**Standard Error (±)**			
Constant	372	578.7			
pH	-166.77	65.59			
Asas	10.328	2.999			
AvdW	-32.819	5.709			
polarisability	298.76	30.79			
Ki3	-1524.8	420.7			
Ki8	5183	1814			
KiCP31	2791	525.7			
KiCP41	-14185	3404			
KiCP42	62777	14509			
KiCP43	-125230	27814			
KiCP47	155899	39019			
O1C	-270.77	60.94	S	**R^2**	Residual Error
			1320.42	**35.6**	472

3.5.4 Logistic Regression

From the models produced (Table 3.6) it was clear that using multiple regression to create predictive models for K_{oc} wasn't successful in this scenario. This meant that an alternative method to analyse the adsorption behaviour of the compounds was needed. Binary logistic regression can be used to try and predict which of the two trends the compounds in Dataset 1 would fall on (Group A trend or Group B trend). This is a way of predicting general adsorption behaviour of compounds based on their chemical structures and properties. Logistic regression is the best way of modelling a direct comparison between the two groups. The best combined model created was 93% concordant with the data and therefore good at identifying the Group B compounds (Equation 6).

$$\ln\left(\frac{\theta}{1-\theta}\right) = -4.99 + 0.414\mathrm{dipolm} + 0.02\Delta\mathrm{Hhyd} + 0.237\mathrm{C1C1C} - 1.14\mathrm{C2C1C} + 15.99\,{}_{1}^{4}\chi_{p}^{c}$$

(Equation 6)

Table 3.7 The significant parameters for Group A versus Group B model. Table also shows regression coefficients and concordance of the model.

Predictor	Coefficient	Standard Error (±)	Odds Ratio	
Constant	-4.99762	0.646875		
dipolm	0.413999	0.104137	1.51	
Δhhyd	0.0243379	0.005156	1.02	
C1C1C	0.236919	0.084667	1.27	
C2C1C	-1.13601	0.428668	0.32	
KiCP41	15.9998	2.22863	8884258	**Concordance**
				93.4%

Equation 6 shows the variables that are significant in determining potential adsorption behaviour for a group of compounds. As this model identifies the Group B compounds, it suggests that these are some of the variables that are significant for compounds that have a high adsorption potential.

The dipole moment and hydration energy parameters are related to solubility. An increase in solubility will reduce adsorption and lower K_{oc} values. The model is suggesting that Group B compounds are more insoluble compared to Group A compounds. When comparing the original data of the Group A and B compounds, although on average Groups A and B have similar dipole moments, the Group B compounds have a larger change in hydration energy which makes them more insoluble. The scatter plots from the PCA support the model as the more soluble Group A compounds are not plotting in the direction of increasing K_{oc} like the Group B compounds (Figure3. 4).

The cluster and path connectivity parameters are related to how branched the chemical structures are. The model is suggesting that the Group B compounds are more branched than Group A which is confirmed when checking with the original data that the Group B compounds that on average have the largest connectivity parameters.

The model suggests the C2C1C molecular fragment is either absent or less prominent in the Group B compounds compared to Group A, which does disagree slightly with the results of the PCA that was shown in Figure 3.5. Although there are some Group A compounds where the C2C1C fragment is more prominent, they are at the extreme of the trend and do not appear to be representative of the majority of Group A compounds. The C2C1C fragment is the alkenes functional group and also relates to solubility (Lloyd, 1989). The alkenes are insoluble in water, so with an absence or reduction of alkenes in the majority of the Group A structures, they will have higher solubility and lower K_{oc} values compared to Group B compounds. When checking the original data the Group A and B compounds have the same average number of alkane and alkene functional groups.

3.6 Summary

Principal Component Analysis was successful in analysing the dataset. The PCA worked in two ways: 1) in identifying that there were two main trends in the data set, a trend that comprised compounds with "typical" adsorption behaviour (Group A) and a sub group of compounds made from only the pyrethoid, avermectin, bipyridylium and aminophosphonic chemical groups with much higher K_{oc} values (Group B). 2) PCA also identified variables that could be deemed important for understanding controls on adsorption.

As there were two distinct trends within the dataset a decision was made to split the data into Group A and Group B compounds to be analysed separately. PCA showed that there was a linear separation between each chemical group type. This suggested that there was a difference in the chemical structure of the types of chemical groups causing the difference in K_{oc} between the compounds. This indicates that for this dataset, compound properties rather than soil properties has the biggest control on adsorption.

Multiple regression has showed that in this situation it not possible to create accurate models for predicting K_{oc}. Attempting to have one model that can cover a wide range of compounds is suggesting a simple solution that is not feasible. The preliminary analysis has identified that understanding the controls on adsorption is more complex than was first anticipated. The best model from this set of data can predict the general adsorption potential of a particular group of compounds compared to another. The Logistic regression model is not good for predicting K_{oc} values, but is successful in identifying the type of properties that might influence adsorption, in this case the solubility of the compounds have been identified as being important. This type of model may be useful for the early synthesis stages of new active ingredients.

The preliminary analysis meant that a general over view of the data could be gained. By understanding some of the trends that were in the data it became possible to target specific chemical groups in the dataset for more detailed analysis. It was decided that the areas for further research were 1) Parent and metabolite compounds, including a lab based study for Glyphosate and a metabolite, AMPA, one of the aminophosponic compounds. This parent/metabolite study also incorporated some work with the Group B compounds as extra data was created. 2) A lab based study on soil properties using a compound from Group A. 3) A predictive modelling study for the benzazole chemical group.

4. PARENT/METABOLITE STUDY

4.1 Introduction

The preliminary analysis of Dataset 1 (Chapter 3) had identified that the relationship between a compound's structural properties and K_{oc} was more complex than first anticipated, and that specific compound groups (as defined by their chemical class in The Pesticide Manual (Tomlin, 1997)) should be targeted and analysed separately. Unfortunately, there were problems with the quantity of data for individual chemical classes needed to be able to develop successful models. However, it was also noted during the preliminary analysis (Chapter 3) that a large proportion of the data was comprised of parent and metabolite compounds. It was thought that the parent and metabolite compounds may be able to produce successful K_{oc} models. Therefore, the parent and metabolite compounds were singled out for a more detailed study into adsorption behaviour. These parent and metabolite compounds will be collectively referred to as Dataset 3. Although, all of the compounds in the Dataset 1 are either a parent or a metabolite, only the parent compounds that had data for their metabolites were selected for use in Dataset 3. So, parents without data for their metabolites and metabolites without parent data were excluded.

The aim of this study is to understand adsorption behaviour of parent and metabolite compounds and to develop predictive K_{oc} models for these compounds. There are two reasons why the parents and their metabolites have been selected for further study. The first reason is to understand the relationship these compounds have with each other, in terms of their structural similarities and their potential influence on K_{oc}. There has been work in the literature on the structural properties of metabolites and their adsorption, for example, work by Gooddy *et al*. (2007) has modelled the transport of some pesticides and metabolites through soil to understand their binding mechanisms, but the model is not specifically related to K_{oc}. It is important that K_{oc} is considered in any model developed as it is a parameter that is used in the pesticide registration process.

The second reason is the importance of the environmental fate of the parent and their metabolite compounds. Pesticides and their metabolites have been detected in groundwater in the US (Koplin *et al*., 2004) and the UK (Johnson *et al*., 2000). Therefore it is important to understand the adsorption behaviour of metabolites as there is the potential for these compounds to leach into groundwater. As of December 2003, the European Union Drinking Water Directive (Council Directive 98/83/EC), has set limits of maximum allowable concentrations of pesticides in water for human consumption. However, there is no requirement for water suppliers to analyse for metabolites

(Gooddy *et al*., 2002). One of the concerns is that some metabolites can show a higher toxicity than their parents (e.g.Tixier *et al*., 2000).

4.2 Study Approach

The dataset used for this study has been a combination of K_{oc} values taken from Syngenta reports, selected from the original full Dataset 1 used in Chapter 3, and a lab based study of a parent compound and a metabolite (Glyphosate and AMPA). This study first focused on the experimental study for Glyphosate and AMPA, detailing the results from this experiment. Finally the results of the analysis are presented and discussed.

4.3 Glyphosate and AMPA Adsorption Study

The experimental work was performed according to OECD guidelines (OECD 2000) and is outlined in Chapter 2, Section 2.3.2. Glyphosate and AMPA were chosen for this study because:

- They are part of the aminophosphonic chemical group and were highlighted in the preliminary analysis (Chapter 3) as showing different adsorption behaviour when compared to other groups of compounds.
- The aminophosphonic chemical group was under-represented in the full dataset, taken from the original Syngenta adsorption reports. By testing these two compounds on different soils meant that a more varied dataset (in terms of soil types) and therefore a wider range of K_{oc} values were obtained.
- AMPA is a metabolite of Glyphosate and so the results provided an opportunity for researching the adsorption properties of a metabolite compared to its parent.

The molecular structures for Glyphosate and AMPA can be found in Figure 2.1

4.3.1 Soil Choice

For this study 17 soils were selected for experimentation (Table 4.1). These soils vary in organic matter content, textural class and global location. Glyphosate and AMPA have not been tested on these soils in the Syngenta database; therefore this will provide new results for analysis of each soil type.

Soil	USDA Textural Classification	Country	%OM	%OC	%Sand	%Silt	%Clay	pH ($0.01M CaCl_2$)	CEC (meq/100g)
Leyland	Loam	USA	1.1	0.6	42.0	45.0	13.0	5.6	8.2
North Carolina	Loamy Sand	USA	2.1	1.2	84.0	10.0	6.0	5.7	7.0
Iowa	Sandy Loam	USA	7.6	4.4	60.0	29.0	11.0	6.6	30.4
Illinois	Silty Clay Loam	USA	4.1	2.4	20.0	52.0	28.0	5.9	28.9
Minnesota	Clay Loam	USA	7.1	4.1	26.0	45.0	29.0	7.3	44.1
Ohio	Loam	USA	5.7	3.3	35.0	39.0	26.0	5.5	63.4
Washington	Sand	USA	0.5	0.3	89.0	7.0	4.0	7.0	3.5
Ushiku	Clay Loam/Sandy Silt Loam	Japan	8.8	5.1	27.0	55.0	18.0	6.4	50.2
Kumamoto	Clay Loam	Japan	9.6	5.6	39.0	41.0	20.0	5.7	51.1
Kagoshima	Sandy Loam	Japan	2.9	1.7	57.0	27.0	16.0	5.8	11.9
Gunma	Sandy Loam	Japan	6.5	3.8	66.0	16.0	18.0	5.4	13.7
Marsillargues	Loam	France	1.0	0.6	31.0	44.0	25.0	7.8	10.7
Gartenacker	loam/silt loam	Switzerland	3.7	2.1	35.9	52.3	11.8	7.3	13.9
Pappelacker	loamy sand	Switzerland	2.1	1.2	71.7	21.9	6.4	7.5	7.3
Borstel	loamy sand	Germany	1.7	1.0	74.9	17.9	7.3	5.1	7.2
18 Acres	sandy clay loam	UK	4.0	2.3	48.0	21.0	31.0	6.8	16.9
Kenny Hill	sandy loam	UK	6.2	3.6	77.0	9.0	14.0	7.4	16.2

Table 4.1. List of soils tested for the Glyphosate and Ampa adsorption experiments and their characterisation.

4.3.2 Experimental Study Results

Table 4.2 Results for 24 hour adsorption step, by soil type for Glyphosate. The complete calculations in the adsorption spreadsheets and LSC Bq counts are in table Appendix B- Folder 1_Glyphosate

Soil	**Classification**	**K_d (ml/g)**	**K_{oc} (ml/g)**	**% Adsorption**
Leyland	Loam	1076	**168573**	98
North Carolina	Loamy Sand	188	**15681**	90
Iowa	Sandy Loam	105	**2394**	82
Illinois	Silty Clay Loam	899	**37424**	98
Minnesota	Clay Loam	84	**2043**	80
Ohio	Loam	486	**222071**	97
Washington	Sand	74	**25439**	79
Ushiku	Clay Loam	4104	**80403**	99
Kumamoto	Clay Loam	18583	**333710**	99
Kagoshima	Sandy Loam	1321	**117832**	99
Gunma	Sandy Loam	10231	**271361**	99
Marsillargues	Loam	252	**43390**	92
Gartenacker	Silt Loam	50	**2341**	67
Pappelacker	Loamy Sand	49	**4051**	69
Borstel	Loamy Sand	210	**21248**	91
18 Acres	Sandy Clay Loam	1276	**55001**	98
Kenny Hill	Sandy Loam	65	**1812**	75

Table 4.3 Results for 24 hour adsorption step, by soil type for AMPA. The complete calculations in the adsorption spreadsheets and LSC Bq counts are in table Appendix B- Folder 1_AMPA

Soil	Classification	K_d (ml/g)	K_{oc} (ml/g)	% Adsorption
Leyland	Loam	213	**33406**	95
North Carolina	Loamy Sand	163	**13597**	94
Iowa	Sandy Loam	69	**5761**	86
Illinois	Silty Clay Loam	415	**17270**	97
Minnesota	Clay Loam	27	**647**	71
Ohio	Loam	305	**13845**	98
Washington	Sand	30	**10459**	75
Ushiku	Clay Loam	1143	**22383**	99
Kumamoto	Clay Loam	1554	**27909**	99
Kagoshima	Sandy Loam	443	**39518**	99
Gunma	Sandy Loam	1038	**157535**	99
Marsillargues	Loam	58	**9935**	85
Gartenacker	Silt Loam	30	**1381**	71
Pappelacker	Loamy Sand	27	**2216**	70
Borstel	Loamy Sand	173	**17581**	95
18 Acres	Sandy Clay Loam	427	**18414**	98
Kenny Hill	Sandy Loam	41	**1143**	79

4.4 Parent and Metabolite Analysis: Methodology

The K_{oc} values from Tables 4.2 and 4.3 were added to Database 3 (created from the Syngenta reports). The reports that had parents with their metabolites were selected and removed from the full dataset to create a database of only parent and metabolite data. By following the same methodology as that used in the Syngenta reports then the K_{oc} values in the Glyphosate and AMPA studies should be comparable to the K_{oc} values already in the dataset and any experimental error between results due to methodological changes should be minimal. In order to understand the variation in K_{oc} between parent and metabolite compounds, this study used a range of multivariate statistical techniques including principal component analysis, multiple regression and analysis of variance. These techniques were explained in Chapter 2, Section 2.2.

4.5 Parent and Metabolite Analysis: Results

4.5.1. PCA: Connectivity Parameters and Molecular Fragments

Table 4.4 The first five principal components for Parent and Metabolite compounds, analysed for connectivity parameters and molecular fragments. Eigenvalues: PC1: 26.86, PC2: 2.77, PC3: 1.79, PC4: 1.1, PC5: 0.86

	PC1	**PC2**	**PC3**	**PC4**	**PC5**
K_{oc}	0.012	-0.145	-0.366	0.360	0.767
Ki0	0.180	0.132	0.180	0.021	0.094
Ki1	0.183	0.109	0.135	0.052	0.107
Ki2	0.188	0.406	0.093	0.053	0.079
Ki3	0.189	0.057	0.107	0.022	0.033
Ki4	0.188	0.060	0.105	-0.001	0.013
Ki5	0.190	0.070	0.076	0.007	0.028
Ki6	0.191	0.047	0.042	0.003	0.002
Ki7	0.190	0.072	-0.008	0.005	-0.002
Ki8	0.189	0.089	-0.058	0.005	0.004
KiCP30	0.186	-0.031	0.083	0.092	0.034
KiCP31	0.183	-0.085	0.069	-0.022	-0.037
KiCP32	0.185	-0.048	0.067	-0.066	-0.052
KiCP33	0.189	-0.042	0.022	-0.037	-0.028
KiCP34	0.190	-0.042	0.005	-0.041	-0.035
KiCP35	0.190	0.038	-0.091	-0.011	-0.037
KiCP36	0.187	0.073	-0.126	-0.019	-0.023
KiCP37	0.184	0.080	-0.177	0.002	-0.024
KiCP38	0.182	0.101	-0.193	0.019	-0.026
KiCP39	0.181	0.097	-0.203	0.012	-0.022
KiCP40	0.113	-0.395	0.057	0.258	0.055
KiCP41	0.100	-0.443	0.084	-0.096	0.021
KiCP42	0.140	-0.366	0.133	-0.122	0.036
KiCP43	0.164	-0.296	0.025	-0.105	-0.013
KiCP44	0.156	-0.317	0.042	-0.045	-0.020
KiCP45	0.179	-0.113	0.047	0.024	-0.125
KiCP46	0.183	-0.017	-0.043	0.037	-0.085
KiCP47	0.181	0.084	-0.189	0.052	-0.078
KiCP48	0.180	0.095	-0.192	0.069	-0.073
KiCP49	0.180	0.091	-0.196	0.049	-0.043
C1C1C	0.081	0.291	0.427	0.096	0.332
C1C	0.155	0.023	-0.165	-0.242	-0.116
O1C	0.136	0.233	-0.140	-0.130	0.048

C1O1C1C	0.091	0.110	0.363	-0.446	0.340
C1C1C1C	0.148	-0.071	-0.014	0.078	0.040
C2C1C	0.053	0.077	0.341	0.664	-0.297
Variance Explained (%)	75	82	87	90	93

The eigenvalues suggest there are five principal components. The first five principal components explain 93% of the variance in the data (Table 4.4). The first principal component has positive loadings for the 6th order connectivity: *0.191* (Ki6). The second principal component has a positive loading for the 4tH order cluster, 1st order path connectivity: *-0.443* (KiCP41). The third principal component has a positive loading for the molecular fragment C1C1C: *0.427* and a negative loading for K_{oc}: *-0.366*. The fourth principal component shows a high positive loading on the molecular fragment C2C1C: *0.664*. Principal component 5 has a high positive loading for K_{oc}: *0.767* and also positive loadings for the molecular fragments C1C1C: *0.332* and C1O1C1C: *0.340*. The values shown in Table 4.4 suggest that it is a combination of molecular fragments and connectivity parameters that are important in influencing K_{oc} in parents and metabolites.

The connectivity parameters can give an indication to the size and branching of the structure while the lower order connectivity parameters, like the Ki6 parameter are usually associated with molecular size (Worrall and Thomsen 2004). Lohninger. (1994) showed that molecular volume is important for sorption. Gramatica *et al*. (2000) indicated that an increase in size of a compound leads to increased hydrophobic effects with a compound tending to bind with the soil organic matter. The path and cluster connectivity parameters represent the structural complexity of the molecule, like the degree of branching, which influences the changes in enthalpy and entropy upon aqueous dissolution (Worrall and Thomsen 2004). An increased degree of branching in a molecule will also restrict microbial degradation (Worrall, 2001). Molecular fragments have been identified in other studies as being important to sorption (Lohninger, 1994). The molecular fragments analysed in this study can be related to the type of bonding and solubility of the compounds.

As the aim of this study was to investigate how K_{oc} varies between parent and metabolite compounds, therefore the two principal components with the highest loadings for K_{oc} were selected to make a scatter plot so any trends could be visualised (Figure 4.1).

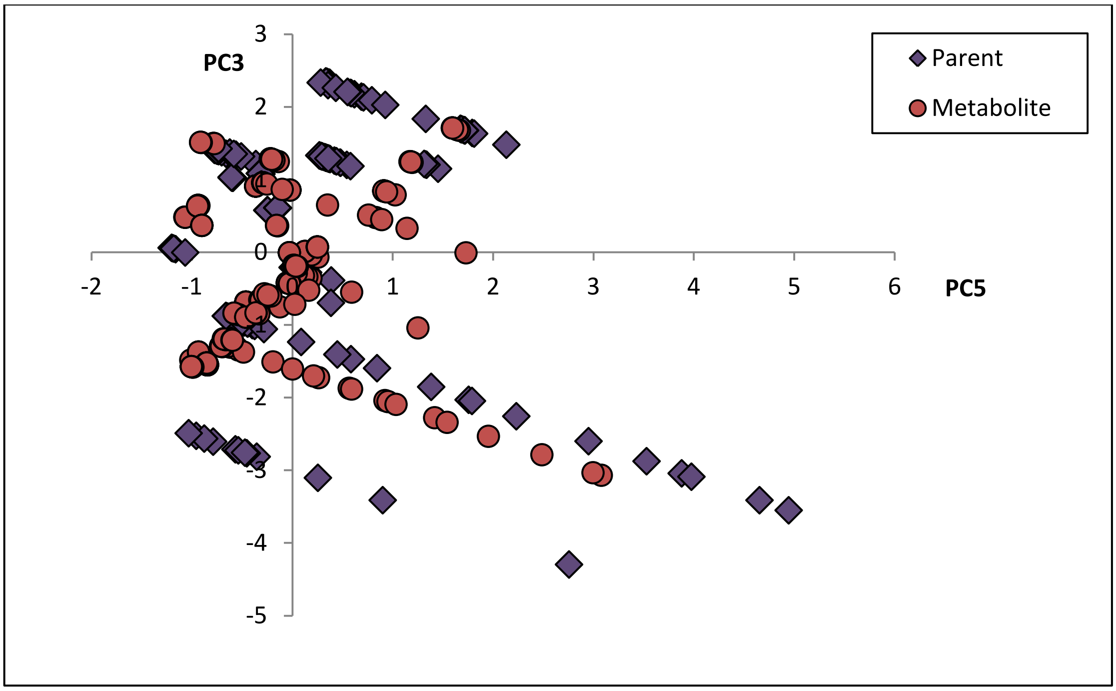


Figure 4.1 Scatter plot of principal component 3 versus principal component 5 for all parent and metabolite data, analysed for connectivity parameters and molecular fragments.

Figure 4.1 shows that the K_{oc} of metabolites differs from their parents. Some of the parents and metabolites show segregation along the y axis while others do not. To identify how the compounds are behaving relative to each other and which metabolites show segregation, Figure 4.1 was redrawn by compound type and separated into parent and metabolite (Figure 4.2).

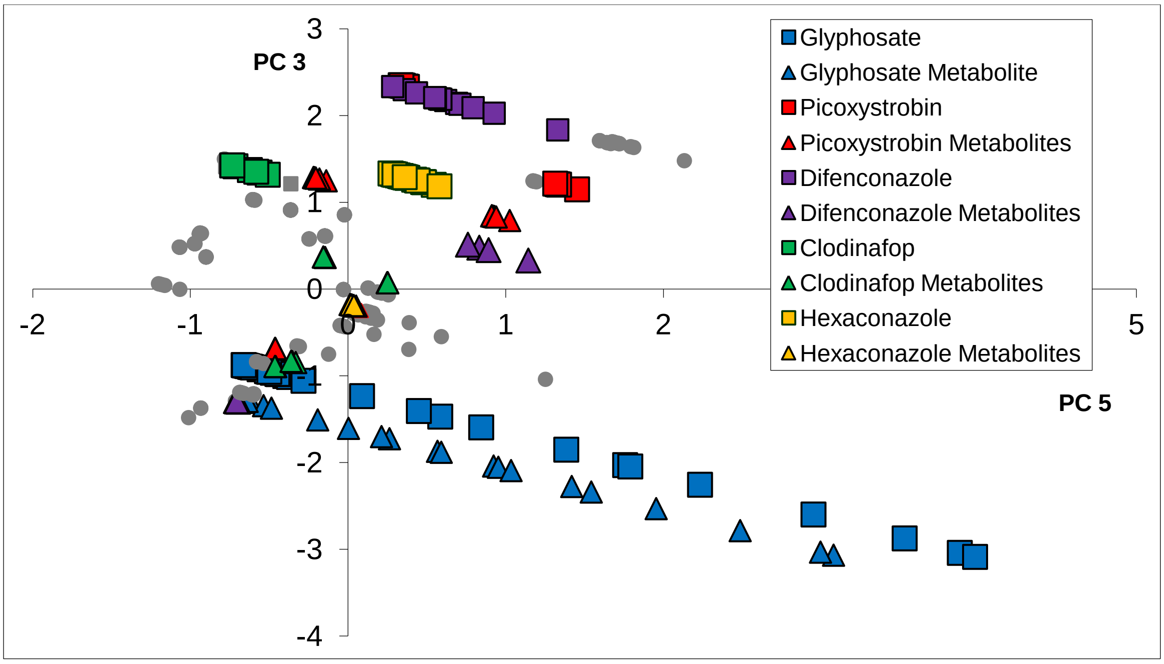


Figure 4.2 Scores from PCA analysed for connectivity parameters and molecular fragments plotted by compound and by parent and metabolite. For labelled compounds square = parent, triangle = metabolites. Grey circles = compounds not of interest.

Figure 4.2 shows that not all of the compounds in the dataset displayed the parallel relationship that was identified in Figure 4.1. The five types of compounds that did show a linear trend have been marked as a square to show the parent compound and its metabolite has been marked as a triangle. The compounds not showing this trend are marked as grey circles.

Based on Figure 4.2 most of the variation between parent and metabolite is along the y axis, which relates to principal component 3. The exception being Glyphosate and metabolite, which shows a wide spread along principal component 5. The spread in data is likely to be due to the range in K_{oc} values created by the range of soil properties in the batch equilibrium study (Table 4.2). Principal component 3 has a positive loading on the molecular fragment C1C1C, which is a propyl chain. This figure suggests that the parent compounds either have the alkane fragment present or have more alkane fragments than their metabolites. When checking the original data, the parent compounds on average do have a larger value for the C1C1C fragment compared to the metabolites (a value of 4 for the parents versus 3 for the metabolites). Alkanes are relatively insoluble in water (Lloyd, 1989) so this would suggest that for these particular five types of compounds the metabolites would be more soluble than their parent.

4.5.2 PCA: Molecular Properties

Table 4.5 The first five principal components for Parent and Metabolite compounds, analysed for molecular properties .Eigenvalues: PC1: 10.00, PC2: 3.10, PC3: 1.64, PC4: 1.10, PC5: 0.77.

	PC1	PC2	PC3	PC4	PC5
K_{oc}	0.011	0.366	-0.141	-0.014	-0.070
dipolm	0.118	0.046	-0.220	-0.552	-0.749
rmsgrad	0.064	-0.088	0.151	0.760	-0.596
totalE	-0.302	-0.153	-0.048	-0.002	-0.022
NHOMO	0.226	-0.303	-0.111	-0.054	0.169
HOMO	0.234	-0.244	0.004	-0.151	-0.017
LUMO	-0.166	0.296	0.483	-0.116	-0.001
NLUMO	-0.182	0.422	-0.029	-0.036	-0.061
Diff	0.199	-0.312	-0.411	0.062	-0.003
V_{sav}	0.301	0.169	0.010	0.035	0.038
A_{sas}	0.282	0.220	-0.087	0.097	-0.012
V_{vdW}	0.301	0.162	0.045	0.006	0.061
A_{vdW}	0.298	0.184	0.029	0.014	0.039
ΔHhyd	-0.072	0.279	-0.580	0.246	0.170
logP	0.229	-0.225	0.361	-0.042	0.038
refractivity	0.306	0.120	0.079	0.010	0.052
polarisability	0.305	0.119	0.079	0.008	0.071
mass	0.302	0.153	0.035	0.011	0.012
Variance Explained (%)	56	73	82	88	92

The eigenvalues suggest there are five principal components. The first five principal components explain 92% of the variance in the data (Table 4.5). The first principal component has positive loadings for refractivity: *0.306* and polarisability: *0.305*. The second principal component has a positive loading on K_{oc}: *0.366* and NLUMO: *0.422*. The third principal component has a high negative loading for change in hydration energy: *-0.580* (ΔHhyd) and a positive loading for LUMO: *0.483*. The third principal component also has a negative loading on K_{oc}: *-0.141*. The fourth principal component shows a high negative loading for the dipole moment: *-0.552*. The fifth principal component also shows a high negative loading for the dipole moment: *-0.749*.

The polarisability, ΔHhyd and dipole moment are related to aqueous solubility. The ΔHhyd is usually inversely related to the aqueous solubility and the polarisability and dipole moment are normally proportional to the solubility (Worrall and Thomsen, 2004). Reddy and Locke, (1994a) found the LUMO to be significant in establishing a relationship with K_{oc}.

As the aim of this study was to investigate how K_{oc} varies between parent and metabolite compounds, therefore the two principal components with the highest loadings for K_{oc} were selected to make a scatter plot so any trends could be visualised (Figure 4.3).

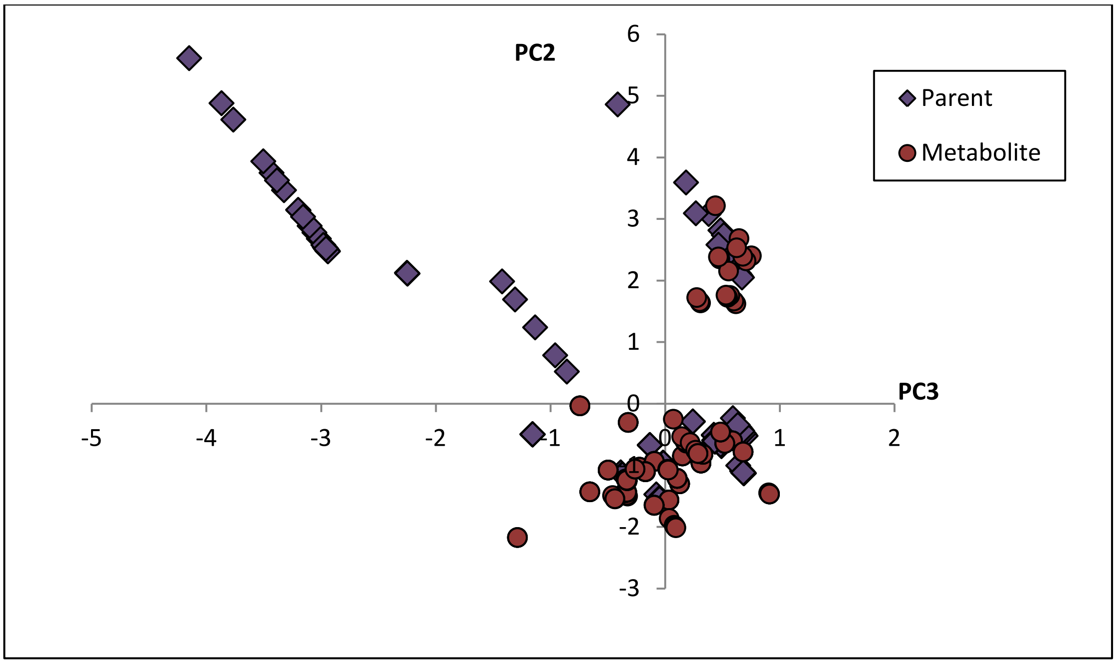


Figure 4.3 Scatter plot of principal component 2 versus principal component 3 for all parent and metabolite data, analysed for molecular properties.

Figure 4.3 shows that there is some linear segregation in K_{oc} between parents and metabolites. Parts of the data are noisy but some of the data points do show a linear trend. To study the difference in parent and metabolite K_{oc} in more detail, this scatter plot was coloured by compound type and by parent and metabolite (Figure 4.4).

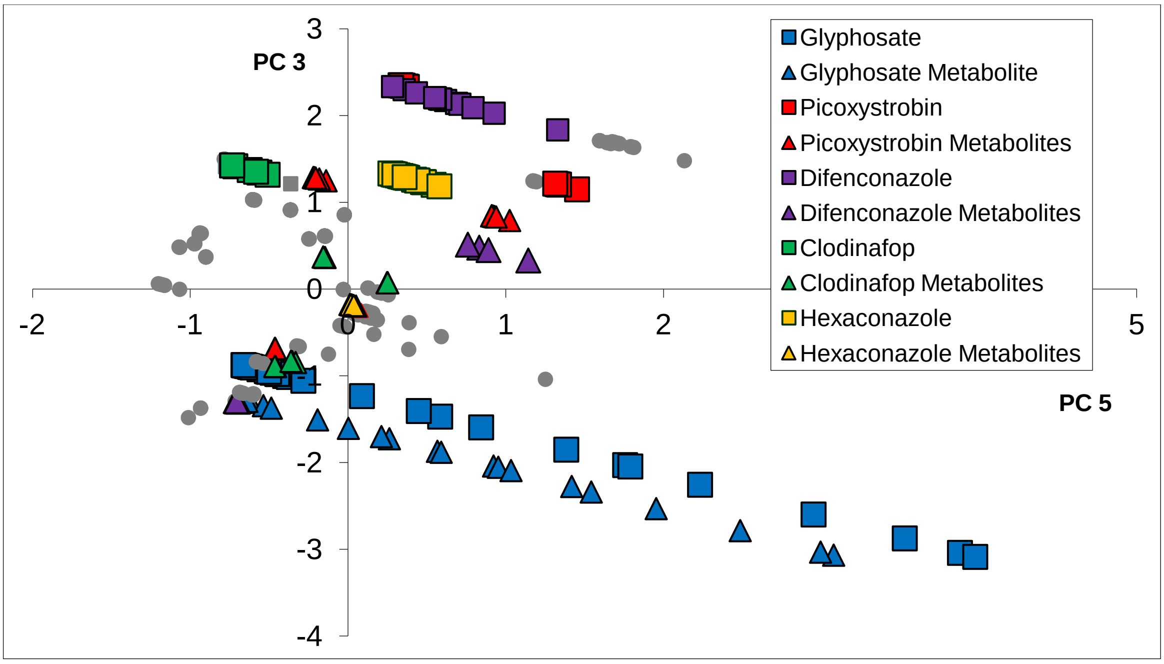


Figure 4.4 Scores from PCA analysed for molecular properties, plotted by compound and by parent and metabolite. For labelled compounds square = parent, triangle = metabolites. Grey circles = compounds not of interest.

Figure 4.4 shows that not all of the compounds in the dataset displayed the parallel relationship that was identified in Figure 4.3. The five types of compounds that did show a linear trend have been marked as square to show the parent compound and its metabolite has been marked as a triangle. The compounds not showing this trend are marked as grey circles.

Based on Figures 4.3 and 4.4, most of the variation between parents and metabolites is along the x axis for principal component 3, which relates to the change in hydration energy. The variation along the y axis is related to K_{oc}. As there was a positive loading on K_{oc} for principal component 2, this means that the compounds that have been coloured and marked in Figure 4.2 are generally the compounds with low K_{oc} values. A possible reason why the compounds that have been marked as displaying the linear trend are also the compounds that have lower K_{oc} values may be due to the interaction with the change in hydration energy. The hydration energy is linked to the solubility of compounds (Worrall, 2001). If the compounds are more soluble then they would have a lower K_{oc} value. Figure 4.2 also suggests that the metabolites may be more soluble than their parent compound, this result is supported by work in the literature *e.g.* van der Linden *et al.*, (2009). It is worth noting that the compounds that have been marked and coloured in figures 4.2 and 4.4 are not the same compounds in both graphs. This would imply that the influences on adsorption behaviour may be specific to particular compound types.

4.5.3 PCA: Soil Properties

Table 4.6 The first four principal components for Parent and Metabolite compounds, analysed for soil properties. Eigenvalues: PC1: 3.75, PC2: 1.56, PC3: 1.19, PC4: 0.822.

	PC1	**PC2**	**PC3**	**PC4**
K_{oc}	-0.020	0.016	0.694	0.718
% OM	0.420	0.448	0.029	-0.043
%OC	0.480	0.452	0.011	-0.040
%Sand	-0.421	0.451	0.061	-0.076
%Silt	0.377	-0.397	-0.133	0.147
%Clay	0.356	-0.396	0.078	-0.070
pH	0.002	0.147	-0.693	0.669
CEC	0.459	0.236	0.101	-0.032
Variance Explained (%)	47	67	81	92

The eigenvalues suggest that there are four principal components. The first four principal components explain 92% of variation in the data (Table 4.6). The first principal component has positive loadings for % organic carbon content: *0.480* and Cation Exchange Capacity: *0.459*. The second principal component has positive loadings for % Sand: 0.451 and % organic carbon content: *0.452*. The third principal component has a high positive loading for K_{oc}: *0.694* and a high negative loading for pH: *-0.693*. The fourth principal component shows a high positive loading K_{oc}: *0.718* and pH: *0.669*. By studying the scores in Table 4.6, it suggests that the most important factor in influencing K_{oc} is the soil pH.

As the aim of this study was to investigate how K_{oc} varies between parent and metabolite compounds, therefore the two principal components with the highest loadings for K_{oc} were selected to make a scatter plot so any trends could be visualised. This produced a scatter plot of PC3 v PC4 (Figure 4.5).

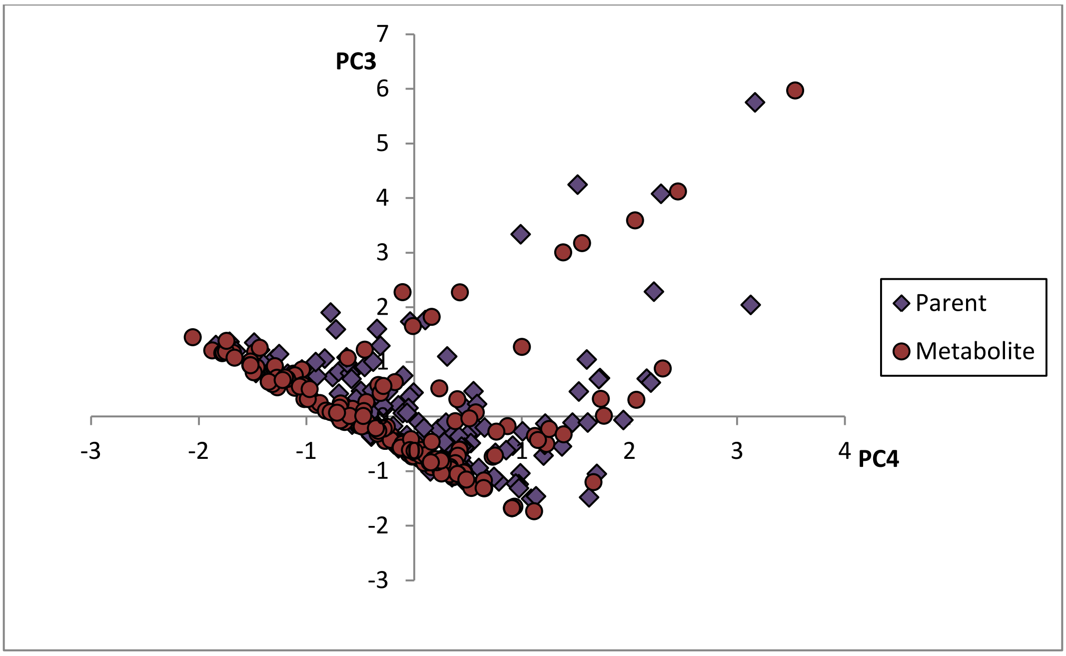


Figure 4.5 Scatter plot of principal component 3 versus principal component 4 for all parent and metabolite data, analysed for soil properties.

Unlike the Figure 4.1 and Figure 4.3, the scatter plot of soil properties (Figure 4.5) doesn't show much variation between parent and metabolite. The parents and metabolites seem to follow the same trend and cluster together. As Figure 4.5 is a plot of pH vs. K_{oc}, it seems to suggest that although the scores show pH as being important (compared to the other soil properties), the actual influence on K_{oc} is minimal. This graph would imply that soil properties are not as important as other molecular properties in understanding K_{oc} of parent and metabolite compounds.

4.6 Modelling Parent and Metabolite Compounds

4.6.1 Multiple Regression Model

The first model attempted on the parent and metabolite compounds was a multiple regression model to predict K_{oc} in both parents and metabolites. The preliminary results (Chapter 3) showed that first attempts at a predictive model for K_{oc} was unsuccessful, one reason for this could have been the range and quantity of compounds. Therefore, it was hoped that by taking a select group of compounds from the original dataset that have similar, but not identical structures (i.e. the parent and metabolite compounds) that the predictive models could be improved.

Table 4.7 The significant parameters of MLR model for predicting K_{oc} of parent and metabolite compounds, the regression coefficients and R^2 value for the model

Parameter	**Coefficient**	**Standard Error (±)**			
Constant	-7.421	8883			
Ki3	54416	13350			
Ki5	-159194	29983			
KiCP32	-79469	23894			
KiCP34	527284	91107			
KiCP36	-1226261	207018			
KiCP38	1956433	255739			
KiCP41	-205018	60350			
KiCP48	18582933	3666392			
C1C	-9464	3374			
C1O1C1C	9274	2852			
C2C1C	-8054	3675	S	**R^2**	Residual Error
			58882.1	**21.1**	477

Table 4.7 shows significant parameters for the best model for predicting K_{oc} of all the parent and metabolite compounds in the database. However it only explains 21% of the variation in the dataset which means nearly 80% of the variation is unaccounted for.

4.6.2 Logistic Regression

Classification Model

It seemed that multiple regression was not the most suitable method for analysing the parent and metabolite dataset. As there were two classes of data, the parents and the metabolites, binary logistic regression was suitable. The first model attempted was to simply classify the compounds based on their molecular properties. The parameters of the model are displayed in Table 4.8.

Table 4.8 The significant parameters of model classifying parent and metabolite compounds. Table also shows regression coefficients and concordance of the model.

Parameter	Coefficient	Standard Error (±)	Odds Ratio	
Constant	0.564101	0.508259		
C2C1C	0.749073	0.131965	2.12	
C1C1C1C	0.395536	0.108986	1.49	
C1O1C1C	0.334752	0.102773	1.4	
NLUMO	0.191507	0.0520043	1.21	
ΔHhyd	0.0525179	0.0188865	1.05	**Concordance**
				69.9

With further calculations, it was found that the classification model had 69.9% concordance with the data, which is good, and could predict the parent compounds based on their molecular properties. The parameters suggest that much of the difference between the parents and metabolites was due to the presence of the C2C1C, C1C1C1C and C1O1C1C molecular fragments in the parent compounds. The presence of these fragments in the parents is confirmed by the original data and on average the parents have higher fragment values for C2C1C, C1C1C1C and C1O1C1C compared to the metabolites. The odds ratio show that the most important parameter in the model is due to the alkene chain, molecular fragment C2C1C. Two thirds of the parent compounds in the dataset possess an alkene chain compared to 30% of the metabolites.

Adsorption Potential Model

The classification model was able to successfully identify the difference between parents and metabolites based on molecular properties but could not explain anything about their adsorption. The simplest adsorption question to consider would be 'is a metabolite more or less adsorbing than its parent?' This type of model would allow any results to be considered relative to the parent compound. The percentage changes in metabolite K_{oc} and metabolite molecular properties were calculated using the formula: *(parent property – metabolite property) / parent property* and based on the K_{oc} values marked as 'more' or 'less' adsorbing than the parent. The new calculated relative values were used in the logistic regression model. The model for predicting adsorption potential relative to a parent is shown in Equation 7.

$$\ln\left(\frac{\theta}{1-\theta}\right) = 1.84 - 27.3\,{}^{3}_{2}\chi^{c}_{p} + 43\,{}^{0}\chi - 9.13\,{}^{3}_{8}\chi^{c}_{p} - 3.32\mathrm{C1C}$$

(Equation 7)

Table 4.9 The significant parameters of adsorption potential model. Table also shows regression coefficients and concordance of the model.

Predictor	Coefficient	Standard Error (±)	Odds Ratio	
Constant	1.842	1.37598		
%change KiCP32	-27.264	12.9535	0	
%change Ki0	43.441	20.4248	7.35	
%change KiCP38	-9.135	4.4334	0	
%change C1C	-3.321	1.66727	0.04	**Concordance**
				91.4

Equation 7 was 91% concordant with the data and therefore good at identifying the metabolites that are less adsorbing than their parents; these are the compounds that may be more likely to leach into groundwater. Equation 7 can be visualised in Figure 4.6

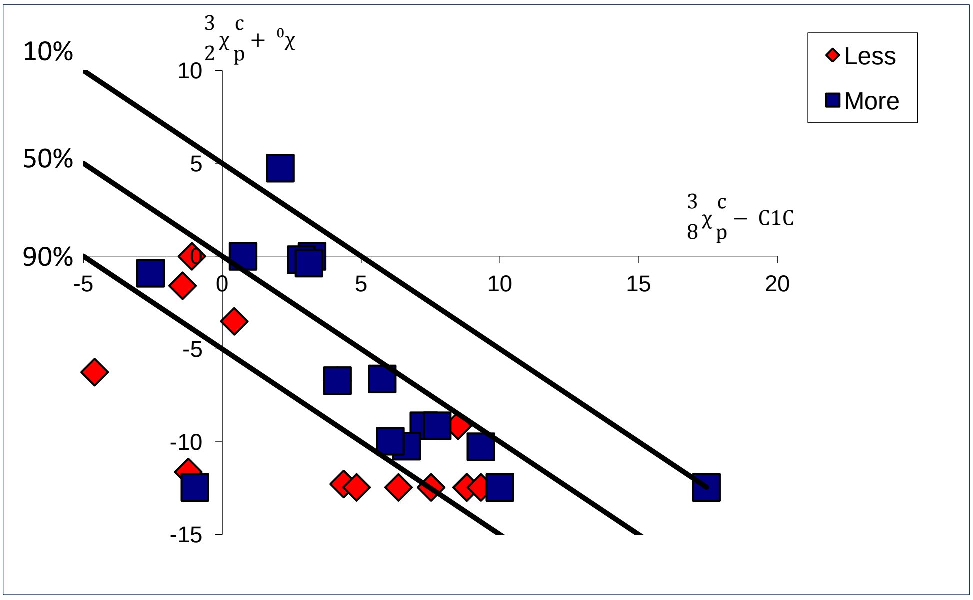


Figure 4.6 Probability plot of Equation 7, showing metabolites that are less adsorbing and more adsorbing than their parents.

The parameters in Equation 7 suggest that a cause of the difference in adsorption potential is due to the size and level of branching in the structures. The odds ratio shows the most important parameter is the percentage change in the connectivity parameter Ki0, zero order connectivity. The Ki0 parameter is usually an indication of the volume of the compound, but can also usually be strongly correlated with molecular mass (Lohninger; 1994). Therefore Equation 7 implies that the metabolites that are less adsorbing will have a larger mass than those that are more adsorbing than their parents. Two of the other parameters in the model are the percentage changes in the connectivity parameters KiCP32 and KiCP38. These parameters can be linked to the complexity of branching in the molecular structure and suggests that the more mobile metabolite have a structure with little branching. What Equation 7 is suggesting is that the metabolites that are less adsorbing (than their parents) are compounds that have a simple non-branched structure, but have a relatively large mass – when compared to those metabolites that are more adsorbing, which would have a smaller mass but a more branched structure.

4.7 Summary

Analysis of parent and metabolite compounds show that there is a difference in K_{oc} between parent compounds and their metabolites that can be identified using principal component analysis. Principal component analysis also showed that soil properties aren't as important as molecular properties in trying to understand the controls on adsorption of parent and metabolite compounds.
It has been possible to model the structural differences between parent and metabolite compounds using logistic regression. This classification model identified the properties most present in the parent compounds. The significant properties in the model are those related to solubility. These types of molecular properties would suggest that the parent compounds are more insoluble than their metabolites, meaning the metabolites are the compounds that would be more at risk of leaching to groundwater. However it is important to note that although the experimental side of this study did focus on Glyphosate and AMPA, the model in Equation 7 covered a range of parent and metabolites. As Figure 2.2 demonstrates both the structures and reported solubilities for Glyphosate and AMPA with both having relatively simple structures it may be that Equation 7 does not apply as strongly to them as it does to other metabolites.

The metabolite adsorption potential model was successful in understanding adsorption relative to the parent compound. The model identified the metabolites that were less adsorbing than their parents, which are the metabolites that are 'higher risk' in terms of groundwater pollution. The model identified the molecular properties that may be influencing the compounds to be more mobile, as being related to the size and the branching complexity of the metabolite.

Understanding controls on metabolite adsorption is important as metabolites have the potential to leach into groundwater. These types of logistic regression model could be used in the pesticide development process to minimise the potential for leaching. The significant parameters in the models have identified the types of properties that influence adsorption and can then be used to create suitable parent compounds. The relative model would be more suitable as part of the risk assessments, in particular the environmental safety study package, as the model provides an indication of the environmental fate of metabolites.

5. SOIL STUDY

5.1 Introduction

As well as considering the variation in K_{oc} due to the molecular properties of the compounds, the variation in K_{oc} that may be due to soil properties also needs to be studied. It is important to consider the variation from soil properties as K_{oc} has been shown to vary across different types of soils (Hornsby *et al*., 1996). Soil organic matter is generally an important adsorption site for pesticides (Farenhorst, 2006), but studies have shown that pesticides also bind to different materials in the soil *e.g.* iron oxides (*e.g.* Clausen and Fabricus, 2001), aggregates (*e.g.* Van Beinium *et al*., 2005) and Fe/Al oxides and clay (*e.g*. Albers *et al*., 2009).

To test the importance of soil properties in understanding and predicting pesticide adsorption, one compound will be tested on a range of soils in a batch equilibrium experiment. The fungicide Hexaconazole is being used for this experiment. The reason for choosing this compound is that it was one of the "Group A" compounds (Section 3.4) and therefore had more 'typical' adsorption behaviour than, for example the Glyphosate and AMPA compounds that were tested in Chapter 4. This means the results of this experiment would be more representative of the majority of compounds in the database. Obviously it cannot be fully representative of the whole of the database as some individual compounds may adsorb by specific mechanisms but it still provides an insight into the effects of soil properties on adsorption.

5.2 Study Approach

This study will first outline the details of the fieldwork for collecting additional soils. Then the experimental methodology will be presented, detailing the additional analysis of the soil phase and the aqueous phase; and results from the lab study. The analysis of the results will be presented and discussed.

5.3 Fieldwork: Collecting Soil Samples

In addition to the soils that were available from Syngenta that are used in their adsorption experiments, a range of soils were collected from various parts of the UK. The locations that were selected for soil sampling were chosen to cover a range of underlying geology. By selecting from differing geologies it would be expected that the soil properties would also vary. The aim of collecting the extra soils was to include some samples that covered a range of soil properties than the Syngenta soils, for example very high organic matter or a very high sand content. Therefore the study could consider the widest possible range of behaviour.

The locations that were selected for sampling are shown on the map in Figure 5.1.

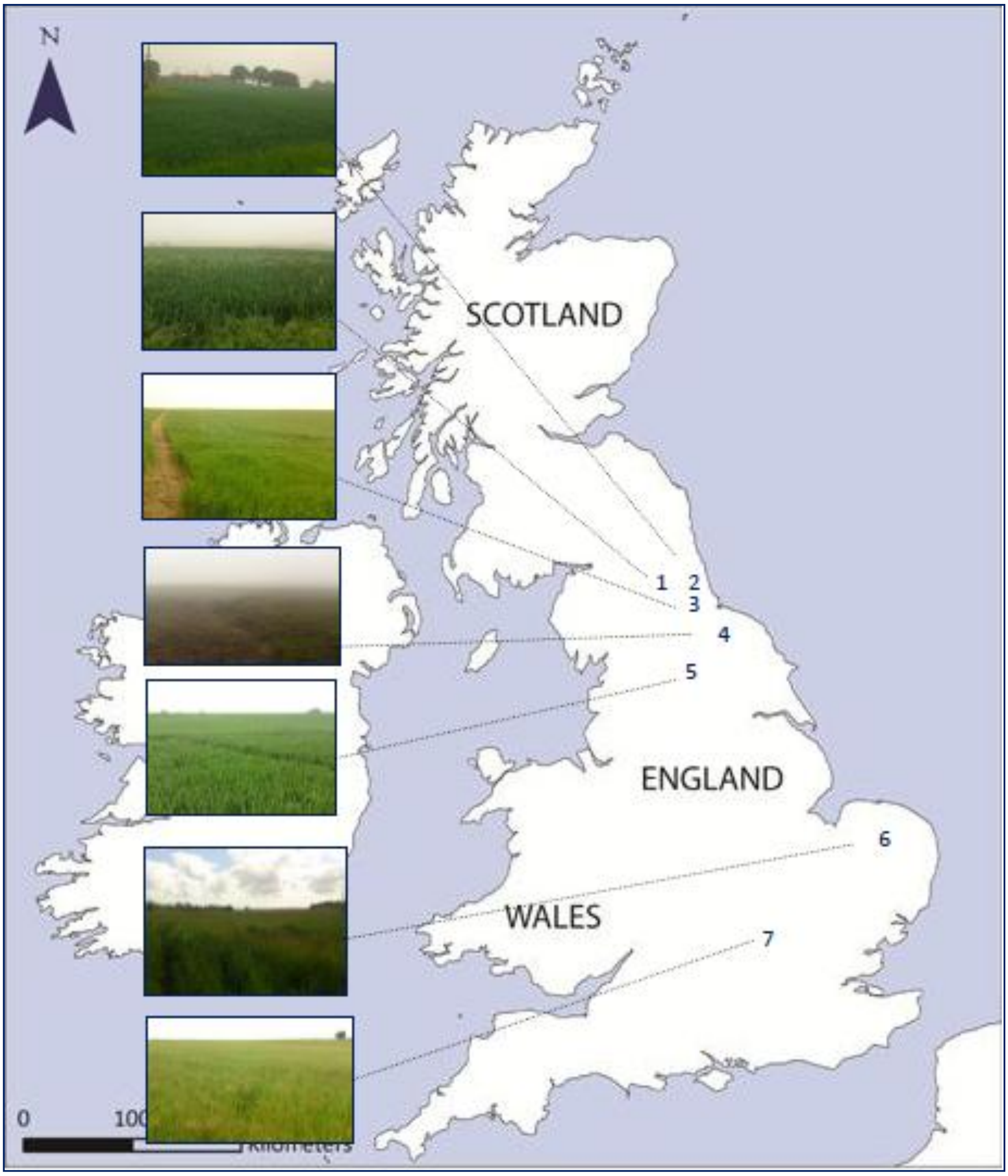


Figure 5.1 Locations of sites for new soil samples. Numbers correspond to site information in Table 5.1

The samples were collected in June 2010 and the 7 samples were all collected from the top 10cm of the soil profile. The geology and land use of the locations intended for sampling are listed in Table 5.1. The information in Table 5.1 has been obtained from the online British Geological Survey Geology maps (http://www.bgs.ac.uk/discoveringGeology/geologyOfBritain/viewer.html) and from the soil maps of England (Soil Survey England and Wales, 1983).

Table 5.1 Geology, soil type, and land use of locations taken for soil sampling

Sample Number	Location	Geology	Soil Type	Land Use
1	Brancepeth, County Durham	Pennine Lower Coal Measures Formation	Acid loams and clays	Winter cereals
2	High Hesleden, County Durham	Roker Formation, Dolostone (upper magnesian limestone)	Fine loamy soils	Cereals
3	Elwick, Hartlepool	Sherwood Sandstone	Loamy and clayey soils	Winter cereals, potatoes, cereals, field vegetables
4	Glaisdale Moor, North Yorkshire	Blanket Peat	Acid raw peat soils	Wet moorland Sheep grazing
5	Tholthorpe, North Yorkshire	Sherwood Sandstone (Aeolian sandstone)	Fine sandy soil	Cereals, potatoes, sugar beet
6	Rockland St Mary, Norfolk	Fenland Peat	Peat soils in part very acid	Cereals, sugar beet, field vegetables
7	St Albans, Hertfordshire	Lewes Nodular Chalk Formation/Seaford Formation (White chalk formation)	Fine loam over clayey	Cereals and other arable crops

5.4 Soil Characteristics

The 7 soils collected from the fieldwork were air dried then ground and sieved through a 2mm mesh, as specified in the available soil reports from Syngenta (OECD 2000). This meant that the way the soil samples were prepared was comparable to the samples used in Syngenta experiments.

The samples were taken to Syngenta's laboratories at their Jealott's Hill research site, which is where the batch study took place. The 7 extra soil samples were sent out for characterisation at the NRM Laboratories in Bracknell, the characteristics are shown in Table 5.2.

Table 5.2 Soil Characterisation as determined by NRM Laboratories Ltd, Coopers Bridge, Braziers Lane, Bracknell, Berkshire RG42 6NS

Sample	**Soil Name**	**USDA Textural Classification**	**Organic Matter**	**% Sand**	**%Silt**	**%Clay**	**pH** (0.01M$CaCl_2$)	**CEC** (meq/100g)
1	**Brancepeth**	Sandy Loam	6.2	61	24	15	5.6	15.1
2	**High Hesleden**	Sandy Clay Loam	5	52	24	24	7.1	14.1
3	**Elwick**	Sandy Clay Loam	6.7	56	21	23	6.7	20
4	**Glaisdale**	Sand/Loamy Sand	11.1	88	6	6	3.6	13.8
5	**Tholthorpe**	Clay	4.4	43	21	36	7.3	21.8
6	**Rockland**	Sandy Loam	11.9	81	7	12	7.4	29.6
7	**Sandridge**	Clay Loam	6	38	40	22	6	18.7

For this study 24 soils had been tested. The remaining 17 soils were Syngenta standard soils. These 17 soils varied in organic matter and textural class and varied globally. With the exception of 18 Acres and Kenny Hill soils, Hexaconazole had not been tested on these soils in the Syngenta database of batch equilibrium studies. Hexaconazole had not been tested on the 7 UK soils. The 17 Syngetna soils are the same soils that were used in the Glyphosate and AMPA study (Chapter 4, Table 4.1).

5.5 Hexaconazole Study

Hexaconazole was chosen as the compound of interest for this study because:

- Hexaconazole represents more "typical" adsorption behaviour compared to compounds previously studied (Glyphosate and AMPA, Chapter 4).
- The aim of the study was to investigate the effects of soil properties on adsorption – using a single compound would be best at highlighting any soil effects

The molecular structure of Hexaconazole can be found in Figure 2.1

5.5.1 Adsorption Study Methodology

The experimental work was performed according to OECD guidelines (OECD 2000) and is outlined in Chapter 2, Section 2.3.2. However, due to unforeseen circumstances, the tubes containing the Hexaconazole and soil slurries could only be sampled after 72hrs. The study supervisor decided that this delay meant that there was a risk that the compound had degraded. Therefore, the tubes were sampled as normal and the supernatants were stored in labelled glass vials to be analysed further.

Analysis of the Aqueous Solution

To test if Hexaconazole had degraded during the longer than planned adsorption step, the aqueous solution was analysed by High Phase Liquid Chromatography (HPLC) with radio-detection. For each soil type a 1.5ml aliqout was taken from the vials containing the supernatants and dispensed into separate HPLC vials, giving a total of 24 vials. The samples were analysed using the HPLC conditions outlined in Table 5.3, which are the conditions that have been used in a previous Syngenta study of Hexaconazole (Oliver and Kuet 1999).

Table 5.3 HPLC Conditions used for analysis of Hexaconazole aqueous phase

HP1100 System:	HP1100 Vacuum Degasser HP1100 Quaternary Pump HP1100 ALS autosampler HP1100 Column compartment HP100 Photodiode Array Detector (DAD) HP Colour Laserjet 4650dn HP Compaq d530 SFF computer (Lablogic Laura 4.0.4.101 SP1)		
HPLC Conditions:	Zorbax CN 5 µm (25cm x 4.6mm id) A: Acetonitrile B: 0.1%Formic Acid (aq)		
Injection Times:	Time (min)	%A	%B
	0	10	90
	5	10	90
	12	42	58
	15	100	0
	17	100	0
	20	10	90
	25	10	90
Flow Rate:	1ml/min		
Temp:	30°C		
UV Detector Wavelength :	205nm		
Injection Volume:	500µl		
Radio Detection System:	Packard Flow Scintillation Analyser 500TR Series with Ultima-Flo M (Packard) Scintillation Cocktail (Flow rate of 4.5ml/min) using Lablogic Laura 4.0.4.101.SP1 software		

The results of the HPLC showed that there was no sign of degradation within the Syngenta soils (Appendix B – Folder 2). However in the 7 UK soils, high radio counts showed that there could be signs of Hexaconazole degradation due to them still being 'live', i.e. not sterilised. When the soils had been characterised, they had not been prepared as thought, and were therefore not sterilised. To check for Hexaconazole degradation, the soil phase was tested for the live UK soils.

Analysis of Soil Phase

The soil phase was analysed by Thin Layer Chromatography (TLC) to check the stability of Hexaconazole. Based on the size of the TLC plate, six soils could be tested. The Glaisdale and High Hesleden soils were selected as they showed high radio counts from the HPLC, indicating there may be metabolites present (Appendix B – Folder 2). The Gunma soil from the Syngenta soils was selected as a comparison as this soil showed no signs of degradation. Three other soils, two from the new soil set and one from Syngenta's soils, were selected at random to use as extra confirmation of any degradation. The three extra soils were Brancepeth and Sandridge from the new soils and North Carolina from the original soil set. To confirm the HPLC results, two aqueous phase samples of the new soils were also tested by TLC. The soil types and the phases tested are listed in Table 5.4

Table 5.4 Soils selected for TLC analysis

Soil Sample	Phase Tested
Glaisdale	Soil
High Hesleden	Soil
Gunma	Soil
North Carolina	Soil
Brancepeth	Aqueous
Sandridge	Aqueous

The Brancepeth and Sandridge aliquots could be sampled directly, whereas the Glaisdale, High Hesleden, Gunma and North Carolina samples had to be extracted. To extract the sample from the soil phase, 25ml of acetone was added to each Teflon tube, containing the compound and soil slurry, and shaken for 10 minutes at 300rpm. The shaking was to make sure the soil was agitated and mixed with the acetone. The tubes were then centrifuged for 10 minutes at 3000rpm so the supernatant could be separated. The supernatant was removed using a glass pipette transferred into a 20ml volumetric flask for storage. From each volumetric flask, a 1ml aliquot was taken from the Glaisdale, High Hesleden, Gunma, and North Carolina samples and dispensed into 1.5ml vials for analysis. From the Brancepeth and Sandridge samples, a 1ml aliquot was taken from the supernatants and dispensed directly into the 1.5ml vials. All the samples were analysed directly from the 1.5ml vials using the autosampling machine.

To analyse the samples, a 20cm x 20cm Merck 60F Silica UV_{254} Plate was used. The samples were applied using the Automatic TLC Sampling ATS4 machine. The samples were sprayed at 2cm from the bottom of the plate. The samples were sprayed in bands of 1cm with a 2cm space between bands. The dosing speed for the soil samples in acetone was 200nl/s .The dosing speed for the aqueous samples in $0.01MCaCl_2$ was 50nl/s.

While the plate was being sprayed the tank and solvents were prepared. For a normal phase TLC the solvents were Hexane: ethanol (80: 20 v/v). The solvents were placed in the tank in between saturation paper and left to equilibrate. The solvents used were determined from previous Syngenta adsorption reports for Hexaconazole (Oliver and Kuet 1999). The plate was placed in the solvent, in between the saturation paper and left until the solvent had reached 4cm from the top of the plate. After it reached 4cm it was removed and left to dry. When the plate was completely dry, it was marked with a sticker for identification and using radioactive ink the solvent fronts were marked. The TLC plate was placed in a cassette and a phosphor imaging plate was placed on top and the cassette was secured. The cassette was placed in a lead box to protect it from atmospheric radiation and left the plate to be exposed for one week. A one week exposure time was chosen due to the levels of radiation in the samples. After one week the phosphor imaging plate was removed from the cassette and loaded into the image reader and scanned (Figure 5.2).

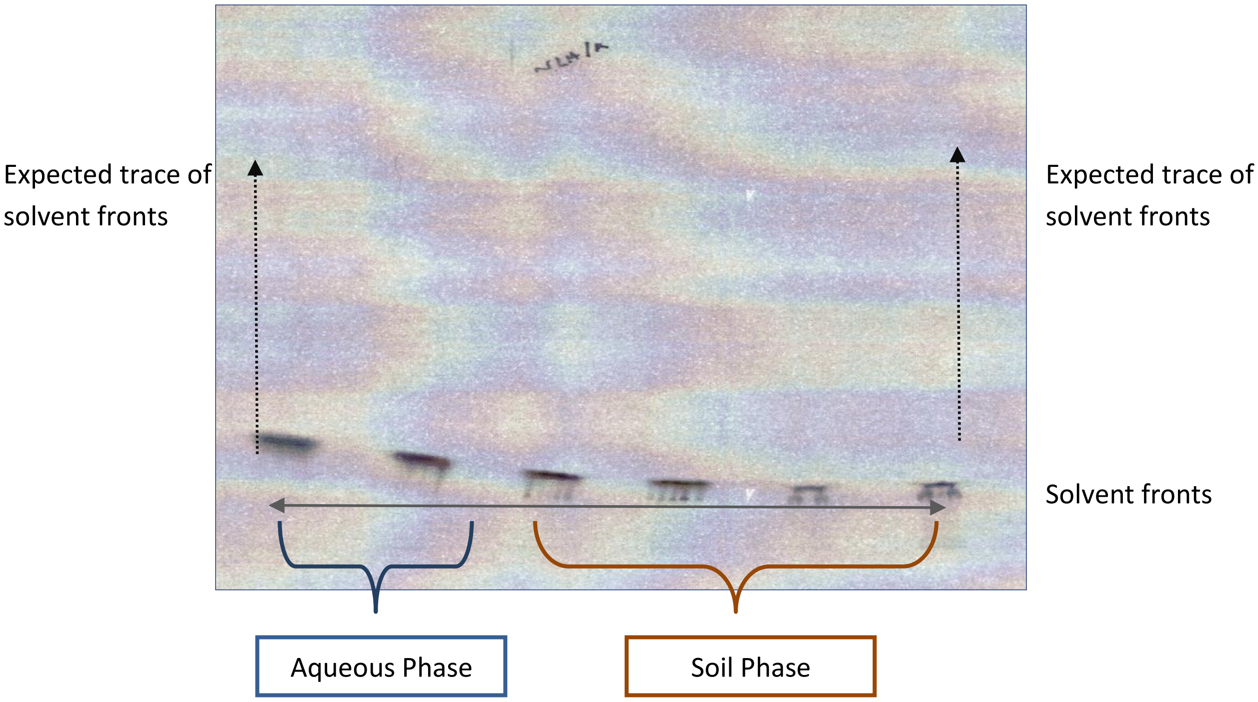


Figure 5.2 Scan of developed TLC plate after one week exposure, showing the position of the aqueous and soil phase solvent fronts. Solvent fronts correspond to samples given in Table 5.4. Dashed arrows give direction of expected trace of solvent fronts, i.e. in a vertical line. If metabolites were present then marks behind the original solvent fronts would be seen on the image where the scan had identified radioactive material.

The scanned image (Figure 5.2) shows that there is an unexpected curve in the trace of the samples. This may be due to the plate having touched the side of the tank or the solvents not being appropriate for the samples. However, the solvents and the methods used were the same as previous Hexaconazole studies so it was assumed that they would be suitable for this study. The image also shows the samples hadn't traced very far up the TLC plate which was also unusual. However, the study supervisor confirmed that the material identified was likely to be the parent material and no traces of metabolite were identified on the plate. Therefore, the K_{oc} values for Hexaconazole could be calculated (Table 5.5).

5.5.2 Hexaconazole Experimental Study Results

Table 5.5 K_{oc} values for Hexaconazole after 72hr adsorption. The complete calculations in the adsorption spreadsheets and LSC Bq counts are in table Appendix B- Folder 1_ Hexaconazole.

Soil	Classification	K_d (ml/g)	K_{oc} (ml/g)	% Adsorption
Brancepeth	Sandy Clay Loam	70	**1944**	87
Elswick	Sandy Clay Loam	42	**1089**	81
High Hesleden	Sandy Clay Loam	30	**1031**	75
Tholthorpe	Clay	57	**2221**	85
Glaisdale	Sand/Loamy Sand	442	**6867**	97
Rockland	Sandy Loam	77	**1669**	92
Sandridge	Clay Loam	36	**1019**	78
Marsillargues	Loam	11	**1861**	51
18 Acres	Sandy Clay Loam	42	**1812**	80
Ohio	Loam	34	**1481**	82
Iowa	Sandy Loam	29	**644**	71
N.Carolina	Loamy Sand	17	**1437**	63
Ushiku	Clay Loam	53	**1038**	78
Leyland	Loam	11	**1669**	51
Illinois	Silty Clay Loam	29	**1205**	71
Minnesota	Clay Loam	49	**1185**	82
Washington	Sand	4	**1298**	27
Kummamoto	Clay Loam	31	**556**	67
Kagoshima	Sandy Loam	28	**1668**	73
Gunma	Sandy Loam	25	**650**	68
Gartenacker	Loam/Silt Loam	16	**765**	57
Pappelacker	Loamy Sand	15	**1199**	54
Borstel	Loamy Sand	35	**3546**	77
Kenny Hill	Sandy Loam	40	**1101**	78

5.6 Hexaconazole Analysis: Methodology

The results in Table 5.5 were added to the Hexaconazole K_{oc} data already available in the original Dataset 1. This created a new dataset of K_{oc} values for Hexaconazole tested on 51 different soil types (Dataset 4). By following the same methodology as that used in the Syngenta reports, it meant that the K_{oc} values in the Hexaconazole study should be comparable to the K_{oc} values already in the dataset and any experimental error between results should be minimal.

This study used a range of multivariate statistical techniques including principal component analysis, multiple regression and analysis of variance. These techniques were explained in Chapter 2, Section 2.2. As there was only a single compound used in the study then there was no variation in molecular properties or connectivity parameters and molecular fragments, therefore only soil properties were analysed.

5.7 Hexaconazole Analysis: Results

5.7.1. PCA: Soil Properties

Table 5.6 the first four principal components for Hexaconazole analysed for soil properties.

Eigenvalues: PC1: 3.88, PC2: 1.5, PC3: 1.44, PC4: 0.57, PC5: 0.33.

	PC1	PC2	PC3	PC4
K_{oc}	-0.162	-0.559	-0.387	0.505
% OM	0.416	0.279	-0.335	0.032
%OC	0.402	0.325	-0.345	0.005
%Sand	-0.418	0.342	-0.313	0.048
%Silt	0.355	-0.378	0.330	-0.425
%Clay	0.394	-0.181	0.189	0.607
pH	-0.051	0.457	0.587	0.437
CEC	0.424	0.091	-0.180	0.048
Variance Explained (%)	49	67	85	92

As outlined in the methodology, the PCA was run a number of times to remove any outliers, to create the best possible dataset for use in the multiple regression. Table 5.6 shows the principal components for the fourth run, which produced the most successful multiple regression model. For the fourth run there are K_{oc} values for the remaining 45 different soil types. The eigenvalues suggests there are four principal components worth investigating. The first four principal components explain 92% of the variance in the data (Table 5.6).

The first principal component shows high positive loadings for % organic matter*: 0.418*, and CEC: *0.424*. The second principal component has high negative loadings for K_{oc}: *-0.559*. The third principal component has a high positive loading for pH: *0.587*. The fourth principal component has a high positive loading for K_{oc}: 0.505, % clay: *0.607* and pH: *0.437*. The loadings suggest that the soil properties that are important for influencing K_{oc} are %clay content and pH. As the aim of the lab study was to investigate the effect of soil properties on K_{oc}, therefore the two principal components with the highest loadings for K_{oc} were selected to make a scatter plot so any trends within the data could be visualised. This produced a graph of PC2 vs. PC4 (Figure 5.3)

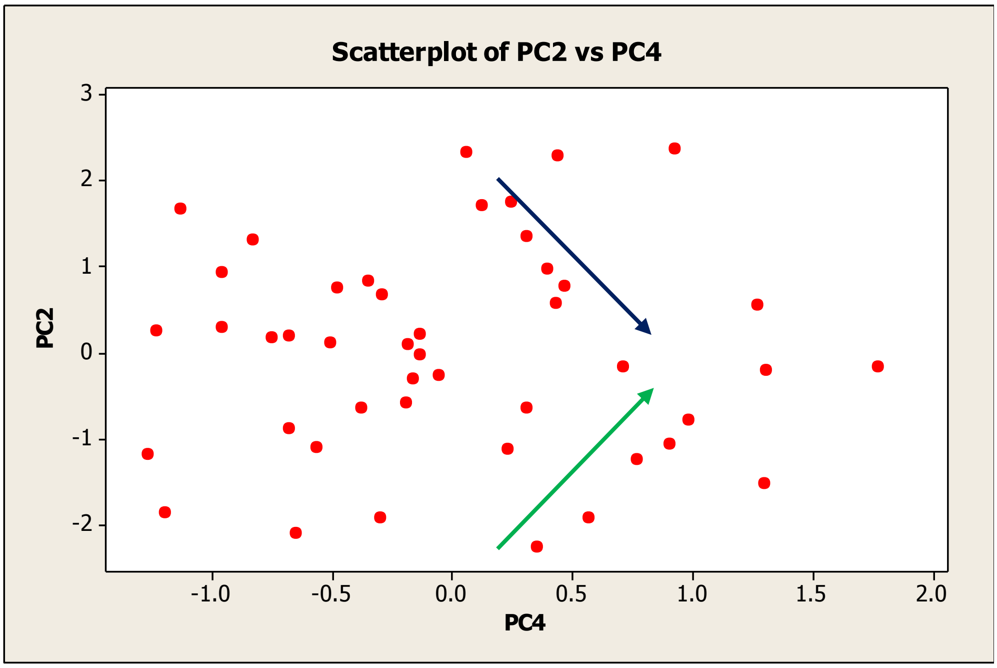


Figure 5.3 Scatter plot of principal component 2 versus principal component 4 for Hexaconazole, analysed for soil properties. Blue arrow demonstrates direction of trend influenced by K_{oc} and soil pH. Green arrow shows direction of trend influenced by K_{oc} and % clay.

Based on the scores from Table 5.6, the scatter plot in Figure 5.3 is mostly showing the interaction between K_{oc}, % clay content and soil pH. There does not appear to be any groupings within the data points. The blue arrow is showing the interaction of the variables for PC2, the interaction between K_{oc} and soil pH, (There is a negative loading on K_{oc} but a positive loading for pH). The green arrow is showing the interaction of the variables for PC4, the interaction between K_{oc} and % clay, (There is a positive loading for K_{oc} and positive loading for % clay). This is creating a chevron trend in the data points with the highest K_{oc} values meeting in the middle. Figure 5.3 is suggesting that the highest K_{oc} values for Hexaconazole will be found on a soil with a low pH and high % clay content.

The findings reported in this study, are supported by similar findings in the literature for similar studies. For example, Ertli *et al.* (2004) studied the effects of soil pH on the adsorption of isoproturon and also found that adsorption increased with decreasing soil pH, which was explained by the formation of hydrogen bonds, typical at lower pH values, between the oxygen, nitrogen, and hydrogen atoms of the isoproturon and the surface groups. When looking at the molecular structure of Hexacoanzole (Chapter 2, Figure 2.1), there are also suitable oxygen, nitrogen, and hydrogen atoms within the structure that could form hydrogen bonds with the soil surface. As discussed in previous chapters, the high clay content that has been identified in Figure 5.3 as being important for adsorption is likely to be related to cation exchange processes occurring in the soil. Paszko, (2012) also stated that at low pH values, a carbamate pesticide was adsorbed by the clay fractions in its protonated form, wheras at highe pH values, the compound was adsorbed by organic and mineral fractions involving non-ionic interactions. These results, although related to a carbamate pesticide, further strengthen the link between acidic and clay soils, and high adsorption values.

The results from Figure 5.3, seemed to match with the results of the previous Hexaconazole studies performed by Syngenta that also indicated that adsorption of Hexaconazole was influenced by increasing clay content and decreasing soil pH, with the highest K_{oc} values being found on acidic soils (Oliver and Kuet, 1999).

5.7.2 Multiple Linear Regression: Modelling Soil Properties

The PCA has identified there are some soil properties that are important in understanding the relationship between soil properties and K_{oc}. MLR was used with the aim of predicting K_{oc} based on soil properties. Unfortunately, the best model only accounted for 38% of the variation in the data (Equation 8).

$$\mathrm{K_{oc}} = 4724 - 418\ \mathrm{pH} - 216\ \%\mathrm{OC}$$

(Equation 8)

Table 5.7 *The significant parameters of MLR model for predicting K_{oc} of Hexaconazole based on soil properties, and the regression coefficients and R^2 value for the model.*

Parameter	**Coefficient**	**Standard Error (±)**			
Constant	4723.5	664.6			
pH	-418.19	94.52			
%OC	-215.77	68.34	S	**R^2**	Residual Error
			555.63	**38.3%**	42

Equation 8 showed that there were two soil parameters that could be modelled to predict K_{oc}; which were pH and %OC. Soil pH was also identified from the PCA as being important for influencing $K_{oc.}$ As there was only one compound tested, the only source of variation had to come from the soil properties. Therefore it is surprising that Equation 8 returned such a poor fitting model. Equation 8 is suggesting that over 60% of the variation in the data is unaccounted for by the model, implying that the variation in the data is coming from soil properties that haven't been measured in this experiment, which may be properties such as mineral content.

Paszko, (2012) raises some considerations regarding mineral content in soil and the effect on pesticide adsorption. The soil characteristics in Table 5.2, list only the percentage clay content in the soil, most likely calculated by its particle size, not the type of clay present. Within the silicate sheets that comprise the structure of the clay there are octahedral sheets and tetrahedral sheets. The octahedral sheet can be comprised of a either trioctahedral sheet, where the sites are occupied by divalent cations such as Mg^{2+} or $Fe^{''+}$, or a dioctahedral sheet, with the sites occupied by trivalent cations such as Al^{3+} (Nesse, 2000). Paszko, (2012) found that the adsorption of carbendazim was negatively correlated with the saturation of the soil complex with Al^{3+} cations. It was hypothesised that as the Al^{3+} cations could 'block' the negative adsorption sites on a soil particle that would otherwise be capable of adsorbing the pesticide cations, reducing the potential for cation exchange.

Equation 8 indicates that lower pH values influence K_{oc}. The influence of lower pH values on increasing K_{oc} is supported by the findings in the Syngenta reports (Oliver and Kuet 1999) and similar work in the literature has found that adsorption of other compounds is also linked to more acidic soils (Albers *et al.*, 2009, Li *et al.*, 2003). However, the percentage organic carbon parameter in the model seems contradictory to what would be expected for this study. Soil organic carbon is usually considered important in influencing K_{oc} (*e.g.* Kah and Brown 2007). For Equation 8, the percentage organic carbon content is negatively correlated with K_{oc}, suggesting that for this compound, organic carbon content limits the adsorption. Oliver *et al.*, 2005 found that for three compounds tested on temperate and tropical soils, organic carbon content was not significantly correlated with adsorption for any of the three compounds. To try to understand the interaction soil properties and compound properties have on K_{oc}, an Analysis of Variance was attempted.

5.7.3 Analysis of Variance (ANOVA)

As using a data set limited to one compound and a range of soil types proved unsuccessful in predicting K_{oc}, the original data had to be analysed by ANOVA in order to understand the causes of variation in K_{oc}. The data set for ANOVA had 45 compounds tested on six soils.

The factors for the general linear ANOVA were compound and soil and the response was K_{oc}. The compound properties and soil properties were added as covariates.

Table 5.8 ANOVA results testing for significance of soils and compounds on K_{oc}.

	K_{oc}	
Soils	0.161	
Compounds	0.000	**R^2**
		72.8%

The ANOVA model in Table 5.8 is suggesting that it is the difference between compounds that has the most influence on K_{oc}.

5.8 Summary

The MLR model has shown that when predicting the K_{oc} of a single compound, only 38% of the variation could be explained by the soil properties measured in the experiment. In the case of Hexaconazole, soil pH seems to be a controlling factor; however the majority of variation in adsorption is unaccounted for. The most likely cause of the unaccountability is that it is due to a soil property that had not been measured in the soil characterisation. Ideally the experiment should be repeated using the same soils, but tested with a wider range of soil characteristics to test this theory. Is this MLR result exclusive to Hexaconazole? Is this low fit only related to Hexaconazole because it has an unusual or missing soil parameter that is controlling its adsorption? Another compound from a different chemical class should be tested on the same 24 soils as a comparison. The ANOVA showed that when comparing a range of soils and a range of compounds for their influence on K_{oc}, it is the compound properties that are the most important for K_{oc}.

These results have implications for the pesticide registration process. At present, the batch equilibrium reports used by Syngenta feature a compound that is usually tested on five soils, although the number can vary. The results shown in 5.7.2 and 5.7.3 are suggesting that either the compounds are not being tested against the right soil properties or in the case of the ANOVA the soil properties are not as important as first considered. It is important that the questions posed in this summary are researched further as, not only do they have financial and time implications for experimental work, they also raise scientific implications for the regulatory system. These questions will be addressed further during the validation and discussion (Chapter 7).

6. STAGE 1 COMPOUND STUDY

6.1 Introduction

As seen in the Group A and B study (Chapter 3), the parent and metabolite Study (Chapter 4), and the soil study (Chapter 5) attempts to create a model for predicting K_{oc} have had limited success. A possible reason for this may have been the variety in the adsorption studies available from Syngenta's SmartDoc database. The results from Chapter 3 have already identified that adsorption behaviour appears to be specific to particular chemical groups of compounds and that attempting to have one model to cover the variety in K_{oc} between the different groups is not reasonable. It seems that there may need to be individual K_{oc} models for each group. However, when multiple linear regression (MLR) has been attempted on individual groups, that has also been unsuccessful (Chapter 3).

Assuming that it is possible to predict K_{oc} from molecular descriptors, then the problem of data quality needs to be addressed. There maybe two possible reasons for the failure of the predictive models. The first is that there aren't enough compounds in each of the individual chemical groups in order to estimate a model. An example of this poor model fit was the Aminophosphonic chemical group. Although there was a lot of K_{oc} data available for the Aminophosphonic chemical group (studies where the compounds have been tested on many soil types) there were only data available for two compounds, Glyphosate and a metabolite (AMPA). As there was then not much difference in the structures of the two compounds, when a MLR model was tried using molecular and structural properties it could not identify any variability that may affect K_{oc}. The second reason for poor model fitting could be that there was not enough data in each of the chemical groups to be able to run the model. An example of this was the Bipyridlium chemical groups. As well as only having data for two compounds, the data for this chemical group comprises only seven different K_{oc} values. Therefore there is simply not enough data for a reasonable model to be generated.

The problems associated with model building experienced so far means that a very specific type of data is needed to create predictive models. There has to be enough variability between the compounds structural properties for it to be identified by the model, but not too much variability that it is impossible to model, as was the case when trying to model the full dataset. The compounds also have to be tested on a range of soils types to create a dataset of a suitable size. Although it is an added advantage if the different compounds in the group can be tested on the same soil types so an appropriate comparison in K_{oc} values can be made.

6.2 Stage 1 Data

The best solution to this problem was to obtain extra data, as the original dataset appeared not to be suitable for this type of modelling work. The extra data came from a stage 1 study, as it fitted the above criteria. Stage 1 studies are completed in the very early stages of pesticide development and involve creating a range of potential compounds with very similar chemical structures, but each compound has a slight variation in its structure so that the compound with the greatest efficacy can be found. Potential metabolites are also created from the original structure. Each potential compound in the stage 1 study is tested on the same range of soil types, in a batch equilibrium study, to obtain a K_{oc} value.

An experimental dataset like the stage 1 dataset is ideal for the predictive modelling work that is being attempted in this project. This is because the aim of a stage 1 study is to investigate how different chemical structures of potential compounds affect K_{oc}, the most suitable compounds are then taken through to the next stage of pesticide development. The aim of this project is also to study the effect of compound properties on K_{oc}, but to be able to predict the effect on K_{oc} without the reliance on lab studies.

6.2.1 Stage 1 Reports

A request was made to Syngenta for some stage 1 data, the criteria for selecting the data being that the potential compounds and the K_{oc} values in the reports were "typical". "Typical" data meant that it compared to the original dataset and they would show a similarity with the results in the Group A dataset. The reason for wanting "typical" compounds was the results would be more representative of the majority of compounds and therefore any model could potentially be extrapolated further to include a wider range of compounds. The reports returned were for a benzazoles chemical group. The general structure of the Stage 1 compounds used in this study is shown in Figure 6.1

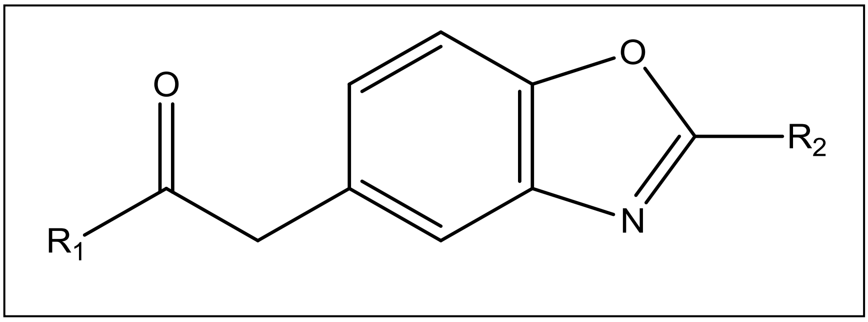


Figure 6.1 General anonymised structure for Stage 1 benzazole compounds used in this study. R_1 and R_2 indicates locations of functional groups and molecular fragments.

Each of the 15 potential compounds were tested on four different soil types, the properties are listed in Table 6.1.

Table 6.1 Soil properties for benzazoles Stage 1 study

Soil	USDA Textural Classification	%OM	%OC	%Sand	%Silt	%Clay	pH ($0.01MCaCl_2$)	CEC (meq/100g)
Rocky Mount	Sandy Loam	0.9	0.52	72	20	8	5.8	3.2
Leland	Loam	0.8	0.47	46	39	15	5.3	7.7
18 Acres	Sandy Loam	4.9	2.84	51	26	23	6.9	16.9
East Anglia	Loam	3.6	2.09	85	5	10	7.4	8.4

6.3 Stage 1 Data Analysis

This study used a range of multivariate statistical techniques including principal component analysis, multiple regression and analysis of variance. These techniques were explained in Chapter 2, Section 2.2.

6.3.1 PCA: Connectivity Parameters and Molecular Fragments

Table 6.2 First five principal components for Stage 1 Compounds, tested for connectivity parameters and molecular fragments. Eigenvalues are: PC1: 16.12, PC2: 10.86, PC3: 3.56, PC4: 0.94, PC5: 0.57.

	PC1	PC2	PC3	PC4	PC5
K_{oc}	0.030	-0.167	0.341	-0.106	-0.247
Ki0	0.180	-0.160	0.217	0.008	0.133
Ki1	0.181	-0.178	0.165	-0.018	0.155
Ki2	0.239	-0.555	0.068	0.069	0.169
Ki3	0.151	-0.233	0.018	0.088	0.244
Ki4	0.171	-0.210	0.000	-0.058	0.214
Ki5	0.166	-0.221	-0.069	-0.079	0.030
Ki6	0.186	-0.193	-0.047	-0.139	-0.120
Ki7	0.198	-0.175	0.004	-0.177	-0.072
Ki8	0.175	-0.201	-0.100	-0.094	-0.120
KiCP30	0.106	-0.084	0.427	0.266	0.014
KiCP31	0.174	-0.202	0.006	0.221	0.159

KiCP32	0.237	-0.066	-0.019	0.030	0.190
KiCP33	0.175	-0.169	-0.277	0.059	-0.104
KiCP34	0.234	0.008	-0.130	-0.048	-0.295
KiCP35	0.239	0.070	-0.078	0.015	-0.067
KiCP36	0.216	-0.006	-0.215	0.057	-0.231
KiCP37	0.229	0.065	-0.152	0.076	-0.069
KiCP38	0.224	0.078	-0.163	0.092	-0.006
KiCP39	0.235	-0.010	-0.154	0.038	0.145
KiCP40	0.068	-0.010	0.474	0.236	-0.069
KiCP41	0.144	0.192	0.200	0.119	0.121
KiCP42	0.146	0.240	0.084	0.023	-0.042
KiCP43	0.145	0.243	0.073	0.016	-0.036
KiCP44	0.143	0.246	0.053	0.001	-0.024
KiCP45	0.142	0.247	0.047	-0.003	-0.021
KiCP46	0.144	0.245	0.062	-0.009	-0.022
KiCP47	0.143	0.246	0.052	0.012	-0.026
KiCP48	0.143	0.246	0.045	0.020	-0.026
KiCP49	0.165	0.221	0.056	-0.025	0.056
C1C	0.200	0.142	-0.118	-0.111	-0.111
O1C	0.063	0.126	0.141	-0.753	0.410
C1O1C1C	0.084	-0.139	0.288	-0.333	-0.553
Variance Explained (%)	49	82	93	95	97

The eigenvalues show that there are five principal components worth exploring, explaining 97% of the variance (Table 6.2). The first principal component has positive loadings for the second order connectivity parameter: *0.239* (Ki2) and the third order cluster, fifth order path connectivity*: 0.239* (KiCP35). The second principal component had a negative loading for the second order connectivity: *-0.555* (Ki2). The third principal component has a positive loading for K_{oc}: *0.341* and the fourth order cluster, zero order path connectivity: *0.474* (KiCP40). The fourth principal component has a high negative loading on the molecular fragment O1C: *-0.753*. The fifth principal component has a negative loading on K_{oc}: *-0.247* and a positive loading on the molecular fragment O1C: *0.410*. As seen in the previous modelling studies (Chapters 3 and 4), it appears that it is a mixture of low order connectivities, path and cluster connectivities, and molecular fragments that are important for K_{oc}.

The lower order connectivity parameters, like the Ki2 parameter are usually associated with molecular size (Worrall and Thomsen 2004) also Lohninger. (1994) showed that the lower order connectivities can also be associated with molecular volume, which is important for sorption. Gramatica *et al*. (2000) indicated that an increase in size of a compound leads to increased hydrophobic effects with a compound tending to bind with the soil organic matter more strongly. The path and cluster connectivity parameters (KiCP35 and KiCP40) represent the structural complexity of the molecule, *e.g.* the degree of branching: an increased degree of branching in a molecule will restrict microbial degradation (Worrall, 2001). The molecular fragment (O1C) has been identified in other studies as being important to sorption (Lohninger, 1994). The molecular fragment analysed in this study can be related to the type of bonding and solubility of the compounds.

As the aim of this study was to predict K_{oc}, the two principal components with the highest loadings for K_{oc} were selected to make a scatter plot (Figure 6.2).

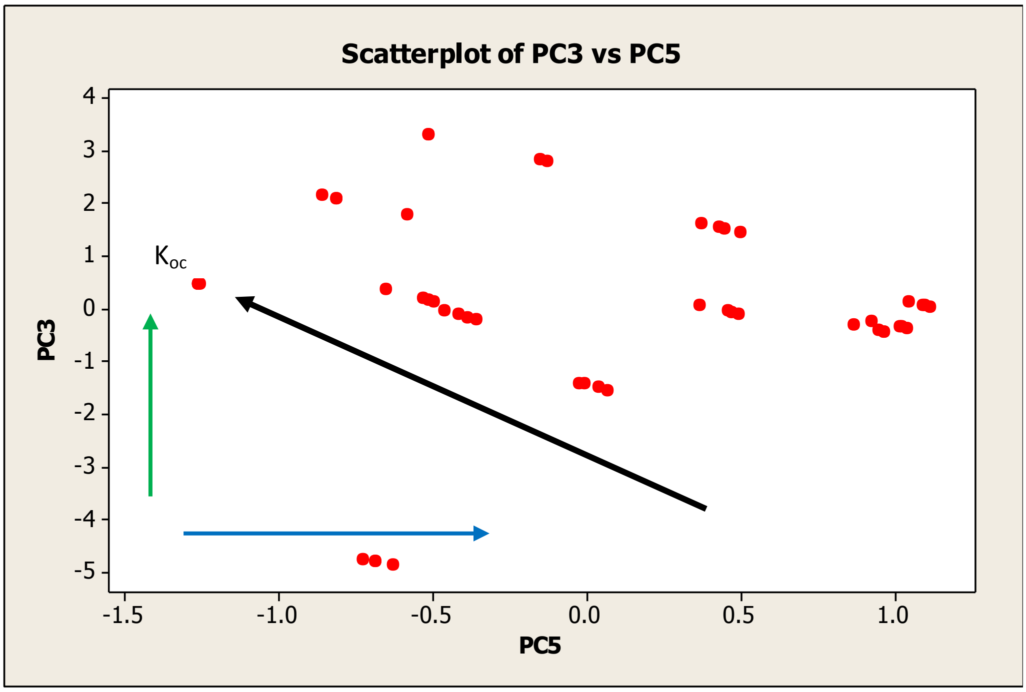


Figure 6.2 Scatter plot of principal component 3 versus principal component 5 for Stage 1 Compounds, analysed by connectivity parameters and molecular fragments. Black arrow represents the direction of increasing K_{oc} values. Green arrow represents influence of 4^{th} order cluster, zero order path connectivity parameter. Blue arrow represents influence of molecular fragment O1C.

Figure 6.2, represents the interaction between principal components three and five, which relates to K_{oc} and KiCP40 and K_{oc} and O1C respectively. Based on the loadings in Table 6.2, the K_{oc} is increasing in the direction of the black arrow, so this would put the compounds with the highest K_{oc} values in the top left region of the graph (Figure 6.2). The green arrow represents the loading on the KiCP40 parameter. As the KiCP40 parameter is related to the structural complexity of the molecule, it would suggest that the branched compounds have a higher K_{oc}. The blue arrow represents the loading on the molecular fragment O1C. In a QSAR model the hydroxyl group O1C, was found to decrease the sorption coefficient (Lohninger, 1994), which would imply that the presence of the O1C fragment in a compound decreases adsorption. As the O1C fragment is polar, the compound would be quite soluble in water, decreasing its ability to adsorb to the soil. Figure 6.1 appears to agree with the results of Lohninger (1994) as the loading on the O1C fragment is increasing in the direction of the lower K_{oc} values.

6.3.2 PCA: Molecular Properties

Table 6.3 First five principal components for Stage 1 Compounds, tested for molecular properties. Eigenvalues: PC1: 7.18, PC2: 5.56, PC3: 1.84, PC4: 1.52, PC5: 0.82.

	PC1	**PC2**	**PC3**	**PC4**	**PC5**
K_{oc}	0.167	0.051	0.271	-0.561	-0.193
dipolm	0.017	0.372	0.361	0.038	-0.098
totalE	0.095	0.087	0.528	0.458	-0.048
NHOMO	-0.028	0.456	-0.102	-0.046	0.350
HOMO	0.109	0.392	-0.226	0.088	-0.426
LUMO	-0.015	-0.475	0.195	-0.031	0.362
NLUMO	-0.065	-0.373	0.245	0.102	-0.632
V_{sav}	0.364	-0.071	-0.106	0.033	-0.017
A_{sas}	0.344	0.126	-0.103	0.111	-0.139
V_{vdW}	0.359	-0.103	-0.122	0.06	0.012
A_{vdW}	0.370	-0.042	-0.068	-0.027	-0.016
ΔHhyd	0.218	0.107	0.456	0.198	0.294
logP	0.308	0.165	0.194	-0.218	0.004
refractivity	0.335	-0.163	-0.142	0.173	0.082
polarisability	0.328	-0.129	-0.188	0.246	0.037
mass	0.259	-0.103	0.125	-0.508	0.038
Variance Explained (%)	45	67	79	88	93

The eigenvalues suggest there are five principal components. The first five principal components explain 93% of variation in the data (Table 6.3). The first principal component shows positive loadings for A_{vdW}: *0.370* and V_{sav}: *0.364*. The second principal component has a positive loading on NHOMO: *0.456* and a negative loading on LUMO: *-0.475*. The third principal component has a high positive loading on totalE: *-0.528* and ΔHhyd: *0.456*. The fourth principal component has a high negative loading for K_{oc}: *-0.561* and mass: *-0.508*. The fifth principal component shows high negative loading for NLUMO: *-0.632*.

The V_{sav} is a size descriptor but in hydrophobic compounds it can be related to the leaching ability of the compound, where it is inversely related to the aqueous solubility (Worrall and Thomsen, 2004). The ΔHhyd is also inversely related to solubility, meaning an increase in the change in hydration energy may lead to an increase in adsorption, whereas the dipole moment is proportional to the solubility (Worrall and Thomsen, 2004). In this case an increase in the dipole moment may lead to a decrease in adsorption. The LUMO (and NLUMO) are parameters that can be related to a hydrogen bonding term and the basicity of the compound (Kamlet *et al.*, 1987), similarly the NHOMO parameter can then be related to the acidity of the compound. The mass of the compound is linked to the adsorption ability of the compound. Gramatica *et al.* (2000) indicated that an increase in size of a compound leads to increased hydrophobic effects with a compound tending to bind with the soil organic matter. As above, the two principal components with the highest loadings for K_{oc} have been selected to make a scatter plot (Figure 6.3)

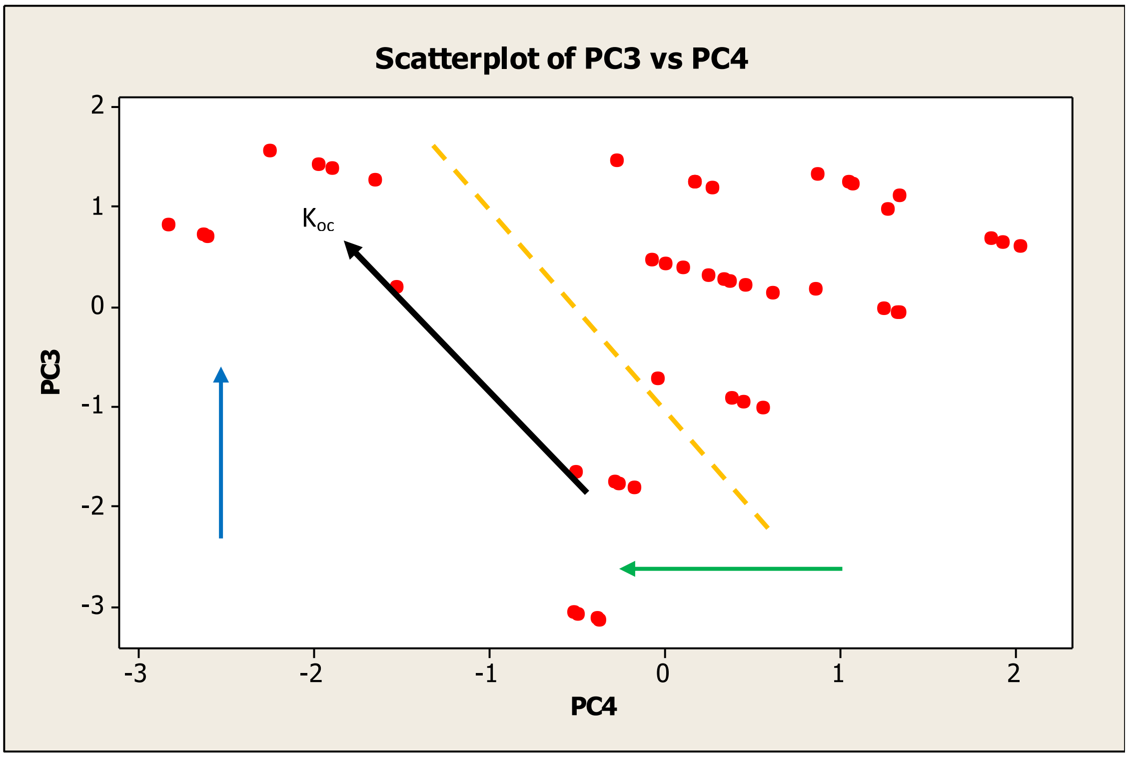


Figure 6.3 Scatter plot of principal component 3 versus principal component 4 for Stage 1 Compounds, analysed by molecular properties. Black arrow relates to the direction of increasing K_{oc} values. Blue arrow represents the influence of the total energy parameter. Green arrow represents the influence of the mass parameter. Orange dashed line shows divide between potential parent compounds and potential metabolite, with data points to the left of the line referring to metabolites.

Figure 6.3 represents the interaction between principal components 3 and 4, which relates to K_{oc} and totalE and K_{oc} and mass respectively. Based on the loadings in Table 6.3, the K_{oc} is increasing in the direction of the black arrow, so this would put the compounds with the highest K_{oc} values in the top left region of the graph. The blue arrow represents the loading on the totalE parameter, the total energy. The total energy can provide information about the bulkiness of the molecules (a size-related descriptor) (Dai *et al*. (1999). The green arrow represents the loading on the mass parameter. Based on the loading for the totalE and mass parameter, the mass or size of the compounds is increasing in the directions of the blue and green arrows, towards the top and left portion of Figure 6.3, where the higher K_{oc} values are found. Figure 6.3 supports the work of Gramatica *et al*. (2000) and shows that increasing mass is important in increasing adsorption. Dai *et al*. (1999) also showed that total energy was significant for K_{oc} and that the totalE parameter indicates that larger molecules adsorb better into soil.

Figure 6.3 also is unique in this study in that it is the only graph where there is a visible relationship between the parent compounds and the potential metabolites and the properties measured. The compounds to the left of the orange dashed line are solely potential metabolites, whereas the compounds to the right of the line are (with the exception of two compounds) parents. Figure 6.3 shows that the potential metabolites have higher K_{oc} values than the parent compounds, as the potential metabolites have a larger mass. The metabolites having a larger mass, does seem unusual and similar results to this were also found in the modelling work of Chapter 4, Section 4.6.2.

6.3.3 PCA: Soil Properties

Table 6.4 First five principal components for Stage 1 Compounds, tested for soil properties. Eigenvalues: PC1: 4.10, PC2: 2.95, PC3: 0.92, PC4: 0.03.

	PC1	PC2	PC3	PC4
K_{oc}	-0.142	0.118	-0.979	-0.077
% OM	0.493	-0.003	-0.035	-0.361
% OC	0.493	-0.003	-0.035	-0.361

% Sand	0.042	-0.578	-0.090	0.182
% Silt	-0.195	0.528	0.127	-0.510
% Clay	0.305	0.453	-0.027	0.604
pH	0.428	-0.277	-0.098	-0.116
CEC	0.420	0.303	-0.062	0.246
Variance Explained (%)	51	88	99	99

The eigenvalues show that there are four principal components worth exploring that explain 99% of the variance (Table 6.4). The first principal component has positive loadings for % organic matter: *0.493* and % organic carbon*: 0.493*. The second principal component has a high negative loading for the % sand content: *-0.578*. The third principal component has a very high negative loading on K_{oc}: *-0.979*. The fourth principal component has a high positive loading on the % clay content: *0.604* and a high negative loading on the % silt content: *-0.510*. As above, the two principal components with the highest loadings for K_{oc} have been selected to make a scatter plot (Figure 6.4)

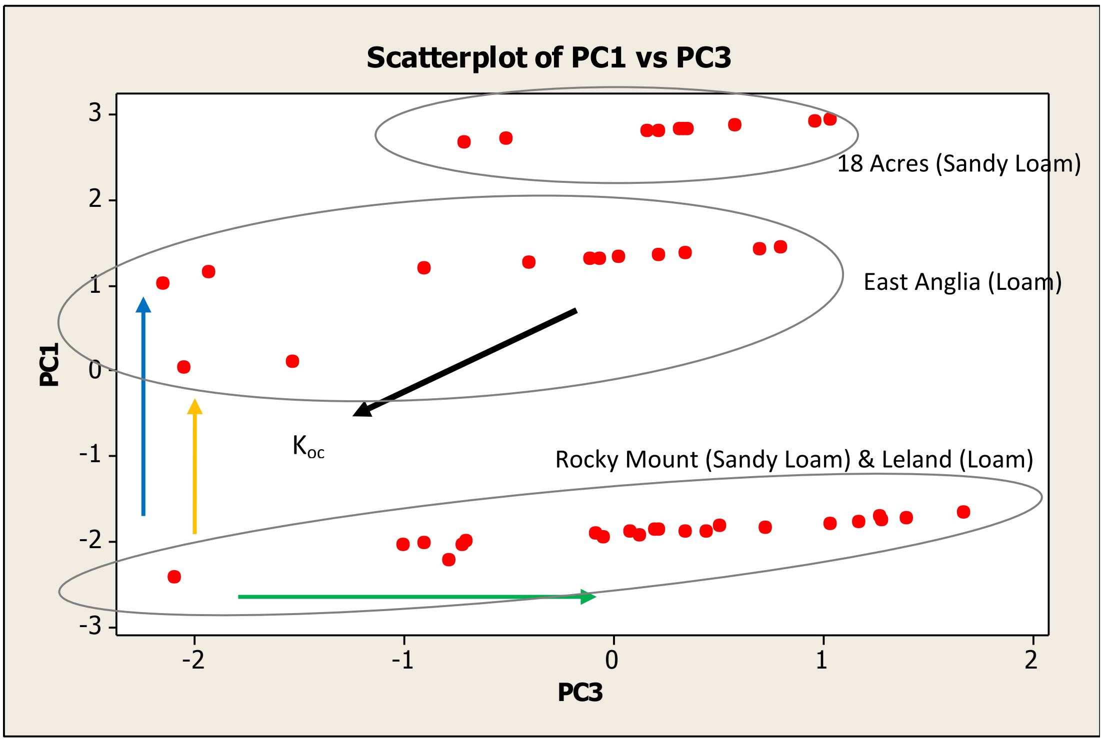


Figure 6.4 Scatter plot of principal component 1 versus principal component 3 for Stage 1 compounds, analysed by soil properties. Black arrow shows direction of increasing K_{oc} values. Blue arrow shows loading from % OC parameter. Green arrow shows influence from %silt parameter. Orange arrow represents loadings from pH. Groupings of soil types have been identified.

Figure 6.4 represents the interaction between principal components one and three, which relates to K_{oc} and percentage organic carbon content and K_{oc} and percentage silt content respectively. The black arrow relates to the loading on K_{oc}, and suggests that the lowest K_{oc} values are found in the lower left region of figure 6.4. The compounds appear to have grouped together according to the different soil types, so there will be one data point from each compound in each of the groups in Figure 6.4.

The blue arrow (Figure 6.4) represents the loading on the %OC parameter and shows that percentage organic carbon content is the reason why the data points have separated out into the groups of soil types. Based on Figure 6.4, it shows that the 18 Acres soil has the highest organic carbon content and the Leland and Rocky Mount have the lowest, which is confirmed by studying the soil characteristics in Table 6.1. As the %OC parameter has a positive loading, with the highest organic carbon contents found in the 18 Acres soil, Figure 6.3 suggests that organic carbon content was not a control on K_{oc}, for this group of benzazole compounds.

As soil organic carbon is usually considered important in influencing K_{oc} (Kah and Brown 2007) this result is initially surprising. However, this result is similar to the results in Chapter 5 (Section 5.7.2), where the MLR model showed that organic carbon content was negatively correlated with K_{oc}. As the compound tested in chapter five was part of an azole group (Hexaconazole was part of the triazole group), then it may share a similar structure with the benzazole compounds tested in this study, and therefore a similar binding mechanism. By looking at the original data for the Stage 1 Compounds and the azole group, they only appear to share similar electron acceptor-donor properties.

By looking at the loadings (Table 6.4), the next most important parameter was pH (orange arrow). Figure 6.4 suggests that the highest K_{oc} values were found on the more acidic soils. Work by Li *et al.*, (2003) has shown that compounds tended to bind more on acidic soils. The result in Figure 6.3 also supports the result in Chapter 5 (section 5.7.1 and section 5.7.2) which showed that the azole compound also showed greater adsorption on the acidic soils. The green arrow (Figure 6.4) represents the loading on the %silt parameter. It would appear that the percentage silt content is controlling the range of K_{oc} values in each group of soil type, with the higher K_{oc} values found on the soils with the lower percentage silt content. The influence of silt content shown in Figure 6.4 is an unusual result as generally silt content is shown to increase adsorption *e.g.* Kumar and Philip (2006).

6.4 Multiple Regression Model

The PCA has identified that there were a number of variables that were important for understanding controls on K_{oc} of this group of benzazole compounds, mainly those relating to the size and degree of branching in a compound, its solubility, and then an indication to type of soil where the highest adsorption may occur. Multiple regression (MLR) was used to model these parameters to predict K_{oc} of this group of benzazole compounds (Equation 9). Equation five has an R^2 value of 92%.

$$K_{oc} = 1067935 - 1971\ \%OM + 600\ CEC + 706\ totalE + 61930\ NHOMO + 32140\ LUMO - 889\ V_{sav} + 2477\ A_{vdW} - 5834\ \Delta Hhyd - 5246\ NsglBnd - 503690\ {}^{3}_{8}\chi^{c}_{p}$$

(Equation 9)

Table 6.5 The significant parameters of MLR model for predicting K_{oc} of Stage 1 compounds, and the regression coefficients and R^2 value for the model.

Predictor	**Coefficient**	**Standard Error (±)**		
Constant	1067935	153609		
% OM	1971	551		
CEC	600	185		
totalE	706	68		
NHOMO	61930	9651		
LUMO	32140	4027		
V_{sav}	889	84		
AvdW	2477	239		
ΔHhyd	5834	591		
NsglBnd	5246	846		
KiCP38	503690	69282	S	**R^2**
			3563	**91.80%**

The two soil parameters give an indication to the types of soils where the benzazole compounds would show the highest adsorption. The organic matter content has already been shown in the PCA work to be not important for adsorption. The cation exchange capacity (CEC) can be an indication of clay content, and would also suggest strong adsorption to clay minerals via a cation exchange process. So these two parameters suggest that the soil controls on the benzazole compounds would be low organic carbon content, but high clay content. As mentioned previously, high clay content in the soil would have many suitable adsorption sites for a pesticide to bind to via cation exchange.

The LUMO parameter can be related to the basicity of the compound and therefore would be a contributing factor to adsorption, as most soils have significant acid sites (Famini and Wilson. 1997). As the specific structures of the stage 1 compounds has been anonymised (Figure 6.1), then this study is unable to identify the details in the structure such as the functional groups that may aid in binding to the soil. However, as the LUMO parameter has been identified as being significant to predicting K_{oc} then it may be that there are basic functional groups in the structure that would easily protonate and be able to undergo cation exchange with the clay minerals. The acidity of the soil was also shown to be important in the PCA work in Section 6.4.1.3. If the soil had a low pH then there would be an abundance of humic acids present. As discussed previously, the carboxyl and phenol groups present in the humic acids can bind with structurally similar functional groups in the pesticide structure, either covalently, or if the pesticide is charged then via ionic bonding. It is a possibility in the case of these stage 1 compounds, that there are carboxyl and phenol groups present in the specific structures.

The ΔHhyd can be thought of as a control on the solubility of the compound. In Figure 6.1, there is a ketone group present in the general structure. The carbon-oxygen double bond in a ketone group is relatively polar and therefore would be able to form hydrogen bonds with water molecules, increasing the solubility of the compound (Ramsden, 2000). However, there may also be other functional groups present in the anonymised structures that would have a lower solubility due to not being strong enough to form hydrogen bonds. The A_{vdW} can be considered as size descriptors meaning that increasing the size of the compound, contributes to an increase in adsorption.

6.5 Summary

The results of this study have shown that it is possible to create a predictive K_{oc} model for a group of compounds. The model suggests that it was size and solubility of the compound that is controlling K_{oc}. The model developed in Equation 9, has important implications for Syngenta and their pesticide development process. Out of all of the models created in this study to predict K_{oc}, the stage 1 model was the only one that was successful, in that based on the R^2 values it explained the largest amount of variation (91.8%).

6.6 Further Thought

The reasons for the relative success in the stage 1 model needs to be considered. As the stage 1 compounds have a common structure, it would be expected that there would be some correlation between the structural properties. However, as the results of the Group A and B study in Chapter 3 showed, it appears that adsorption behaviour is compound or chemical group specific. Therefore it seems likely that any predictive models that are developed will have to be tailored to specific chemical groups. It is possible that the positive result in Equation 9 is indeed due to the quality and quantity of the stage 1 data that was outlined in the introduction to this study. Any groups of compounds with a similar structure would be likely to produce a successful predictive model. However due to their experimental nature combined with their importance in the early stages of the registration process; and the quality and suitability of the data in the reports, the stage 1 compounds make a better choice for developing predictive models. It is therefore proposed that further research into the stage 1 compounds is considered to test this theory.

7. VALIDATION AND DISCUSSION

7.1 Introduction

This chapter will validate the predictive K_{oc} models created for the Stage 1 compounds (Chapter 6). The metabolite models (Chapter 4, Section 4.6.2) have been validated as they have been tested to determine the correct classification of data. The model discussed in this chapter is a new model based on a separate validation study. The data for the model was divided into a training set and validation set on a random selection of two thirds training and one third validations, which is convention for this type of validation.

7.2 Stage 1 Model Validation

The new stage 1 model (Equation 10), which was based on data from Chapter 6 was created from the training data was validated against 19 compounds from the validation set.

$$\mathrm{K_{OC}} = 4086180 - 2170\ \%\mathrm{OM} + 797\ \mathrm{CEC} + 574\ \mathrm{totalE} + 112857\ \mathrm{NHOMO} + 156762\ \mathrm{HOMO} + 91075\ \mathrm{LUMO} + 1142\ \mathrm{V_{vdW}} - 3752\ \Delta\mathrm{Hhyd} - 4131\ \mathrm{refractivity}$$

(Equation 10)

When tested, Equation 10 tended to overestimate K_{oc}, predicting higher K_{oc} values for 12 out of 19 compounds. Overestimating K_{oc} in a screening model is particularly a problem for the compounds that are very mobile, like some of the compounds in this dataset. Some of the observed K_{oc} values (Compound 10 in Table 7.1) were extremely low, classifying them as very high to high mobility based on the McCall Scale (McCall *et al.*, 1980), but the model predicted K_{oc} values, classifying them as immobile. By using this model in a screening situation, then these mobile compounds could potentially be wrongly taken through to the next stage of development.

The mean average percentage error between the observed and predicted K_{oc} was calculated as 155406%. The predicted K_{oc} vs. observed K_{oc} is plotted in Figure 7.1.

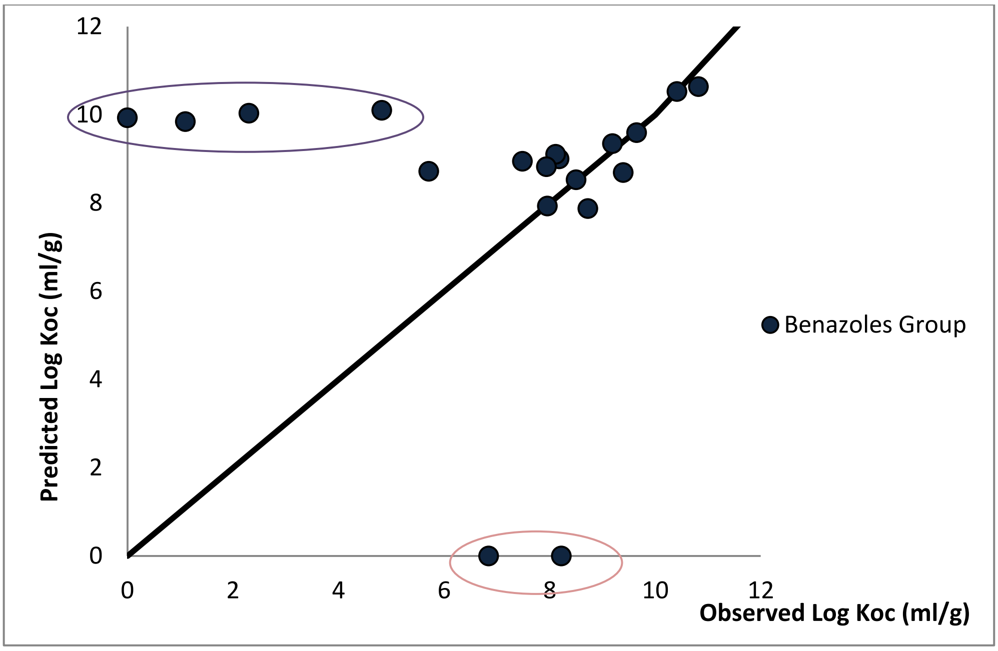


Figure 7.1 Predicted Log K_{oc} versus observed Log K_{oc} for Equation 10. Black line is a 1:1 line, and shows the trend that would be expected if observed and predicted values were the same. Purple circle represents the compound that does not have a good fit with the data. This extremely mobile was removed from the model. Pink circle highlights compounds that have been under predicted by the stage 1 model.

Figure 7.1 shows that for certain data points the model is not a good fit, particularly the four data points in the purple circle. These four data points correspond to the same compound, but tested on four different soil types (Compounds 10a, b, c and d in Table 7.1). Compound 10 is also extremely mobile as indicated by the observed K_{oc} values. The structural properties of Compound 10 show that it has a small total energy and V_{vdW} compared to the rest of the compounds in the validation data set. The total energy is related to the bulkiness of the compound (Dai *et al*. 1999). The V_{vdW} is a size related descriptor and also a control on solubility (Worrall and Thomsen, 2004). Therefore this model is not suitable for soluble compounds with a small mass. It also backs up other work in this study suggesting that size and solubility are controls on adsorption.

The two data points in the pink circle relate to compounds that have been predicted by the model to have extremely low K_{oc} values. One of these compounds (Compound 4a in Table 7.1) had quite a low observed K_{oc} regardless of the prediction (939 ml/g). However the other of these compounds (Compound 5b in Table 7.1) had an observed value that would class it as slightly mobile (3722ml/g) but the predicted value was -1166 ml/g. Both of these compounds have by chance been validated twice due to being tested on different soil types (Compound 4b and Compound 5a in Table 7.1). On the other occasion these two compounds are a good fit with the model, showing a mean average percentage error between the observed and predicted K_{oc} of 57% and 2% (Table 7.1). As the two compounds are structurally similar to each other and having been tested before and showing a good fit, it is therefore assumed that the low predicted values are due to soil properties. By looking at the experimental data for these compounds and comparing those to the variables in Equation 10, the %OM and CEC variables are significant. Equation 10 suggests that higher K_{oc} values are related to a low %OM and high CEC values. One of the soils tested does have a low OM content of just 0.9% but also has a low CEC of 3.2 meq/100g. The other soil had 3.6 %OM and a CEC of 8.4 meq/100g, which according to Equation 9 are both related to lower K_{oc} values.

Compound 10 that was in the purple circle was removed from the data set and the mean average percentage error between the observed and predicted K_{oc} was calculated again as 132%. Now the model can on average predict K_{oc} values to just over one order of magnitude. The model now overestimates K_{oc} for 8 out of the 15 compounds and underestimates K_{oc} for 7. The new predicted vs. observed K_{oc} values are plotted in Figure 7.2.

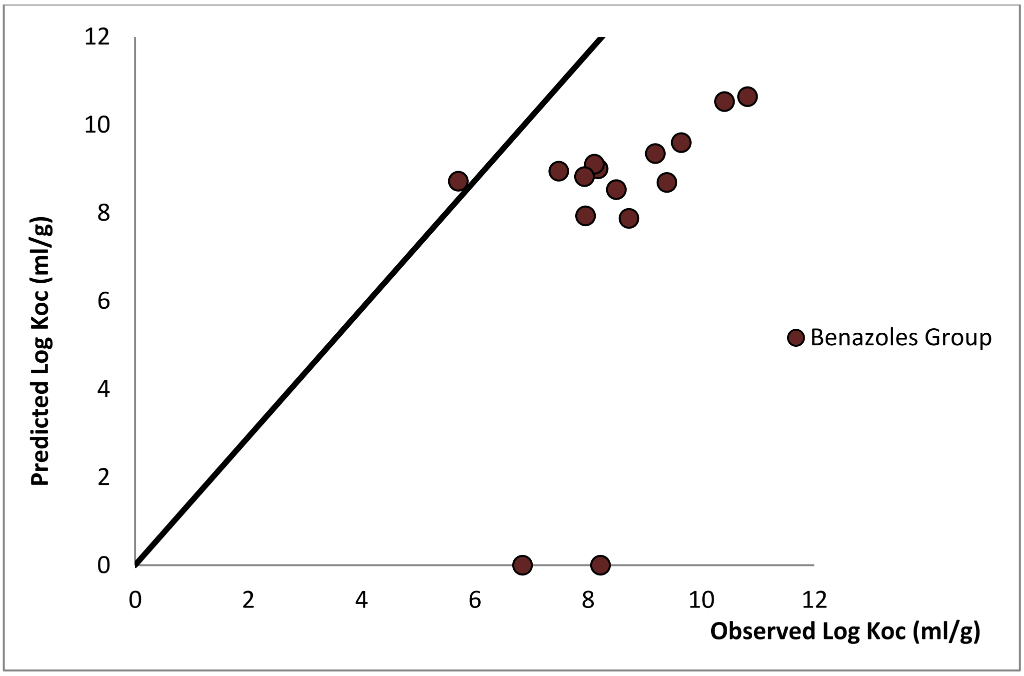


Figure 7.2 Recalculated predicted Log K_{oc} versus observed Log K_{oc} values for Equation 10, having removed data for Compound 10. Black line represents line of equal observed and predicted K_{oc} values.

7.2.1 Stage 1 Model Discussion

The results of both Chapter 6 and the validation of Equation 10 show that the structural properties that are significant for predicting K_{oc} in a benzazoles group are can be grouped into size, solubility and type of bonding. There is also the influence of some soil properties that are significant for K_{oc} although this model is chemical group specific so it is likely that the soil properties that are significant in this model are not significant for other compounds. Structural properties like size are important for adsorption as the size of the molecule can influence the way the compound is transported through the soil (van der Bruggen *et al*., 2002). Famini and Wilson, (1997) suggest that the larger the compound, then the better it is adsorbed to the soil, with the larger compounds being less soluble. The adsorption of pesticides with larger molecular sizes may be influenced by van der Waals forces. Van der Waals bonds are additive and so their contribution increases with the size of the interacting pesticide (Gevao *et al.*, 2000). The HOMO and LUMO parameters in Equation 10 may represent adsorption by charge transfer complexes formed by the electron donor-acceptor mechanisms (Gevao *et al.*, 2000). Charge transfer mechanisms create a relatively strong adsorption between the pesticide and soil (Moreno-Castilla, 2004) which may be represented by the high K_{oc} values in Table 7.1.

The main assumption for the validation of Equation 10 is that the experimental K_{oc} values are correct. It is known that there is variability in K_{oc} values (Dubus *et al*., 2003) so it is hard to validate against a value that is not constant. Considering experimental K_{oc} is not constant, building a model (Equation 10) that can predict K_{oc} to just over one order of magnitude could be seen as a success. The experimental K_{oc} values have been calculated from soil properties that are more variable whereas Equation 10 is based on structural properties that are less variable. Therefore it could be argued that Equation 10 may provide a more accurate value for K_{oc} than those estimated by experimental values.

7.3 Comparison of Stage 1 Model against EPA Model

The US Environmental Protection Agency (EPA) developed the EPI-suite software to provide a screening tool for organic compounds. The stage 1 compounds were imported into the program as mol files to predict their K_{oc} values as a comparison to the model created in this study. The EPA model provides two methods for estimating K_{oc}: an estimation using the first order molecular connectivity index (MCI); and estimation using log K_{ow} (the octanol-water partition coefficient). For both methods the mean average percentage error was calculated, for the MCI method it was calculated to be 86%, and for the log K_{ow} method was 656%. The estimated K_{oc} values for both methods are plotted against the observed K_{oc} values and the stage 1 model values Figure 7.3.

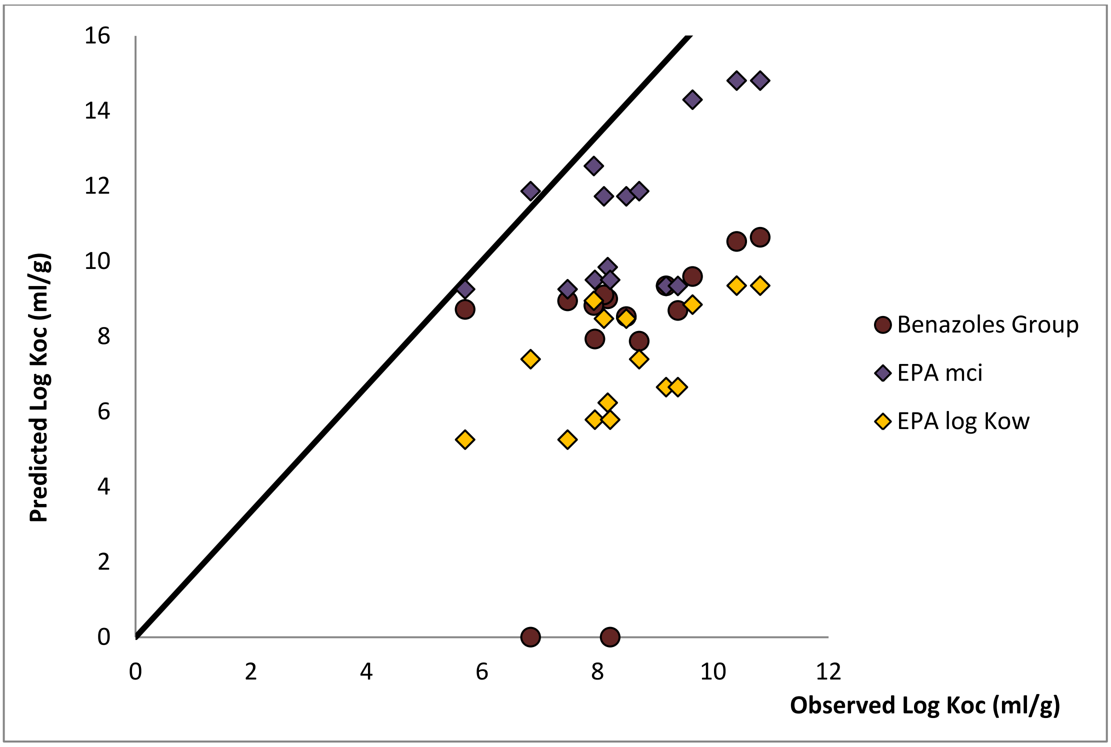


Figure 7.3 EPA predicted Log K_{oc} values, compared to observed Log K_{oc} values and the Log K_{oc} values predicted by the Stage 1 Model.

As the log K_{ow}, method produced a high mean average percentage error; this method will be disregarded from the discussion. Also, similar work in the literature generally prefers the use of molecular indices over K_{ow} (*e.g*. Meylan *et al*., 1992). The MCI method shows that although on average it can estimate K_{oc} to within one order of magnitude; it also overestimates the K_{oc} value for most of the compounds (Figure 7.3), although the EPA admits that the model does produce higher than expected values of exposure to err on the side of safety (EPIWIN, US EPA). To decide which model is preferred out of MCI or Stage 1 depends on the level of accuracy required. Selected compounds from Table 7.1 are used as an example of how this decision could be made.

Table 7.1 Compounds from the validation set and their K_{oc} values and percentage differences between predicted and observed values. Compound 10 was removed from validation and therefore not tested against EPA values.

Compound	**Observed K_{oc} (ml/g)**	**Stg1 K_{oc} (ml/g)**	**% difference**	**EPA K_{oc} (ml/g)**	**% difference**
Compound 1	3567	8118	56	18860	81
Compound 2a	3342	9011	63	124100	97
Compound 2b	4923	5057	3	124100	96
Compound 3	2806	6798	59	278100	99
Compound 4a	939	-1171	180	142400	99
Compound 4b	6154	2633	-134	142400	96
Compound 5a	2857	2787	-3	13430	79
Compound 5b	3722	-1166	419	13430	72
Compound 6a	1781	7691	77	10480	83
Compound 6b	302	6126	95	10480	97
Compound 7a	9806	11473	15	11440	14
Compound 7b	12026	5955	-102	11440	-5
Compound 8	15499	14713	-5	1623000	99
Compound 9a	33288	37371	11	2698000	99
Compound 9b	50000	14768	-20	2698000	98
Compound 10a	*-21*	*20612*	*-98252*	*x*	*x*
Compound 10b	*124*	*24415*	*19590*	*x*	*x*
Compound 10c	*10*	*22851*	*228407*	*x*	*x*
Compound 10d	*3*	*18897*	*629808*	*x*	*x*

If hypothetically using the McCall Scale (McCall *et al.*, 1980) as criteria for deciding which compounds should be taken to the next stage, then around half of the compounds would be classed as immobile (Observed K_{oc} > 5000 ml/g) which would probably be a good reason for putting them through to the next stage of development (Table 7.1). Both of the predictive models give estimated K_{oc} values that would also indicate that the compounds are likely to be immobile, although the EPA Model mostly over predicts them. In the case of Compound 7b the Stage 1 Model actually under predicts K_{oc} by around 100%, however it still returns a K_{oc} value that would indicate immobility so in this case the environmental fate of the compound would probably not be cause for concern.

Out of the two models the Stage 1 Model does have the worst root mean average percentage error, but for most of the compounds in the validation set the Stage 1 model actually has the smallest individual percentage difference. The smaller individual difference means that the Stage 1 model predicts K_{oc} values that are more realistic (realistic in terms that it is closer to what would be expected from a lab study). Having more realistic values is important, particularly in the case of Compounds 2b and 5a (Table 7.1). Compounds 2b and 5a have observed K_{oc} values that would put them in the category of slight mobility. The Stage 1 model estimates K_{oc} values with a 3% difference. These realistic K_{oc} values mean that the two compounds might not be suitable for the next stage of development. In contrast, for Compounds 2b and 5a the EPA Model estimates K_{oc} values with a 79% and 96% difference. These EPA K_{oc} values are unrealistic and may result in these compounds being wrongly taken into development.

If specific predicted K_{oc} values are not that important then the EPA model may be the best choice overall; however based on individual K_{oc} values in the validation set this study suggests that the Stage 1 Model should be used for prediction of this chemical group. In comparison to the EPA model which uses only the first order molecular connectivity index, the Stage 1 Model is constructed from a range of molecular properties that are interpretable in terms of adsorption. The Stage 1 model also includes soil parameters (%OM and CEC) so can provide an estimate of K_{oc} for specific sites.

7.4 Suitability of K_{oc} Models

The results of this validation have shown that the Stage 1 Model can offer a screening model for K_{oc} in the early stages of compound development for a particular group of low mobility compounds. However, when considering the use of predictive models one of the issues raised is the suitability of K_{oc} as a predictive tool. When creating and validating the models it is assumed that the original K_{oc} values obtained by the batch experiments are the correct values. However, Wauchope *et al*. (2002) stated that the batch method will tend to overestimate short term adsorption and under estimate long term adsorption. Wauchope *et al*. (2002) also estimated that a batch experiment will probably vary from the true average K_d value in a field of the same soil by a factor of up to two. Using K_{oc} also assumes that the main control on adsorption is organic carbon, which is not true for all compounds as the work in Chapter 5 shows.

This study shows that much of the control on K_{oc} was size and solubility related. Significant parameters related to solubility included the dipole moment, hydration energy and Van der Waals volume, with the more soluble compounds being less adsorbing. Along with partitioning and sorption, solubility can be thought of as one of the main controls on the environmental fate of pesticides (Thomsen *et al*., 1999), and there have been studies in the literature that have been successful in predicting aqueous solubility from molecular structure *e.g.* Patil, (1994) and Delgado, (2002). Therefore based on the results of this study it may be more appropriate to use structural properties to predict aqueous solubility, and use solubility as an indicator of the potential behaviour of pesticides, or at least to consider solubility alongside K_{oc}.

Any predictive K_{oc} models will need to be chemical class specific. Work in the literature (e.g. Müller, 1997) has already identified the need for compound class specific models and is supported by the work in this study. Using solubility as another indicator of environmental fate may reduce the problem of different compounds adsorbing by different mechanisms, and the need for individual chemical group models.

7.4.1 Is it possible to predict K_{oc}?

Taking the above concerns relating to K_{oc} into consideration, and assuming that K_{oc} is still the parameter that is to be predicted, then this study has raised some practical concerns relating to how transferrable the Stage 1 Model is. The benzazoles group analysed in Chapter 6 and validated in Equation 10, share some soil adsorption mechanisms with Hexaconazole from the triazoles group that was analysed in Chapter 5, mainly low organic matter and acidic soils. To test how far the Stage 1 model could be used in predicting K_{oc}, some compounds from the triazoles group were substituted into the model. The root mean average percentage error between observed K_{oc} and predicted K_{oc} for each compound in the azole group were calculated (Table 7.2).

Table 7.2 Calculated root mean average percentage errors for Hexaconazole and other compounds from triazole chemical group.

Compound	root mean average percentage error
Hexaconazole	2916%
triazole compound 1	9895046%
triazole compound 2	1400323043%
triazole compound 3	2487426%
triazole compound 4	4618297%

Table 7.2 shows that generally the benzazole Stage 1 Model is unreliable at predicting K_{oc} for compounds other than those it was built for. The exception is Hexaconazole which has an average percentage error which is relatively small compared to the other compounds. The smaller percentage error for Hexaconazole may be due to Hexaconazole and the benzazoles sharing a similar adsorption mechanism for binding to soils. Therefore, when building K_{oc} models, a possible method could be to not build them by chemical group which may be restrictive and would result in many different models, but instead to build them by adsorption mechanism allowing for more flexibility.

As well as the general science surrounding K_{oc} prediction being a problem, this study found there was the issue of quantity of suitable data posing a problem for confidently predicting K_{oc}, particularly for Chapter 5 and Chapter 6. The results of Chapter 5 Section 5.7.2 and 5.7.3 suggest that soil properties tested are not as important as the compounds own structural properties for predicting and influencing K_{oc}. However, the data validated by Equation 10 suggests that soil properties do actually have some effect on K_{oc}, and soil properties can make the difference between classifying mobile and immobile compounds (Table 7.1, Compound 4a and 4b). The problem this study had was that the results of Chapter 5 were only tested on a single compound. Likewise, Equation 10 was only built for benzazole compounds. The exclusivity of these results means it is hard to evaluate the effects of soil properties on other compounds.

The OECD Guidelines (OECD 2000) advise that soils for batch equilibrium adsorption studies should be cover a range of soil properties that would be typical of temperate geographical zones. However, it may be useful if Syngenta were to also use a control soil that remained constant across all compounds tested to try and create a base value for K_{oc} for compounds to be validated against. Inevitably debate would occur as to the characteristics of this control soil and whether it is appropriate for all compounds, but a control would at least be an attempt at direct validation and comparisons of K_{oc} between different compounds.

7.5 Summary

The results of this study suggest that predicting K_{oc} is possible for a specific group of compounds. This study would advise that further models should be created to assess if it is possible to predict K_{oc} for other compounds and other chemical groups/adsorption mechanisms. For chemical group specific models, for practical reasons (outlined in Chapter 6) the most appropriate data is likely to be the Stage 1 reports.

8. CONCLUSION

The aims of this project were to:

1. Understand the controls on pesticide adsorption in soil.

Aim 1 was to investigate how K_{oc} varies between soils and compounds. By understanding controls on K_{oc} there is the potential for reducing pesticide movement from soils into groundwater, which is important for groundwater quality and pollution.

As indicated by Principal Component Analysis and the multiple regression models, controls on adsorption of compounds are generally molecular size and solubility parameters, with larger, less soluble compounds being more adsorbing. Other important parameters were connectivity parameters meaning compounds with branched structures were also more adsorbing.

2. Develop models to predict K_{oc}.

Aim 2 was to investigate if K_{oc} can be predicted from structural compounds. Predictive models have the potential to reduce the need for experimental K_{oc} values in the early stages of the pesticide development process.

This study has developed a range of models based on structural properties to predict pesticide behaviour. The models show that:

- General adsorption behaviour can be modelled into high and low adsorption categories.
- Metabolite adsorption potential relative to its parent can be predicted, with the mass of metabolites being a key parameter in mobility.
- K_{oc} can be predicted within a range of just over an order of magnitude for a group of benazole compounds.

Experimental work suggested that for Compound A, the soil properties measured were not as important as the structural properties of the compound in predicting K_{oc}. Although, the significant properties in the QSAR models did not always directly relate to the soil properties given in the batch equilibrium studies, the adsorption of pesticides can be linked to soil via their molecular structures and their ability to undergo processes such as protonation and cation exchange. However the Stage 1 model validation suggested that soil properties did have some influence on the prediction of K_{oc}. Therefore soil properties should be included in the models on a case by case basis.

The results of this study, shown in Chapter 3 and Chapter 6, report that adsorption behaviour is chemical group specific. A stage 1 screening model has shown that it is possible to predict K_{oc} for a specific group of high adsorption compounds. However, this model does not transfer well to other groups of compounds. Based on the results of this study, currently K_{oc} can be predicted but only in group specific models.

This study recommends that:

- To increase the range of compounds that have a K_{oc} prediction, further chemical class or adsorption mechanism specific models should be created.
- To attempt a direct comparison between compounds, a control soil should be considered in adsorption experiments.
- To reduce the problem of different adsorption mechanisms between compounds, using solubility QSARs as indicators of environmental fate could be considered.

Based on the recommendations, potential future work would involve both data based work and lab work. Relating to Chapter 5, it is recommended that the batch equilibrium experiment is repeated, either using the same range of soils that were used in the Chapter 5 experiment on different compounds, or testing Compound A for a different range of soil characteristics. The aim of this experimental work would be to further understand the effects of soil on K_{oc}, and to identify the appropriate soil characteristics to include in the models for each compound. Included in the experimental work should be a control soil so that direct K_{oc} comparisons can be made.

The main priority for future work should be building a database of K_{oc} models for chemical groups or adsorption mechanisms. Although any group of compounds with a similar structure can be used, for convenience and availability of data it is recommended that the stage 1 reports are targeted.

Future work relating to QSARs may also be attempted. If solubility data is available then this should be used along with K_{oc} for a better understanding of environmental fate. Depending on the success of K_{oc} and solubility QSARs, the same technique could be applied to other parameters important for pesticide fate such as half life or ecotoxicology.

Overall, this study has identified the influence the molecular structure of a pesticide has on its environmental fate. The research presented in this thesis, although still in the preliminary stages, does offer both environmental and financial justifications. Through analysis of a large dataset of pesticide adsorption studies, molecular properties that would have an effect on adsorption and mobility in soil have been identified. These significant molecular properties appear to have an effect on, or be influenced by solubility; and therefore reiterate the link between solubility, adsorption and environmental fate. This study also highlights the complexity involved in attempting to understand the interactions between soil properties and molecular properties, when predicting environmental fate. Financially, there is the opportunity to make the pesticide development process more efficient by using QSAR models to screen potential compounds in the early stages of development. However, the QSARs presented in this study still need further refinement.

APPENDIX A: HOW TO MAKE YOUR MODEL

1. Drawing the compound

This study used ISIS Draw (MDL Information Systems) to draw the chemical structure of all compounds. Other chemical structure drawing packages could be used as long as the structure of the compounds can be converted into a .mol file. A .mol file is needed to import the structure into other packages to calculate structural properties. As mentioned in the methodology the .mol files are in the form of text files with rows of numbers and letters that correspond to elements within the compounds structure.

The compound can also be drawn in a program like HyperChem. Using HyperChem to draw the compound would be particularly suitable if only molecular (semi-empirical or quantum) parameters are needed. If other structural parameters are needed the best option is to use ISIS Draw, this is because HyperChem does not have an option for converting to .mol files.

2. Obtaining Molecular Properties

This study used HyperChem v8.0 (Hypercube Inc.) to calculate the molecular parameters. The full list of parameters calculated in this study is in the methodology Chapter 2 Section 2.1.1. Regardless of whether the structure has been drawn in HyperChem or drawn in another package and imported, the structure has to be optimised. This is because the structure appears as a hydrogen-reduced structure.

In the following instructions the first word in each section is the title of the drop down menu and is followed by the command.

2.1 Optimising the Structure

1. Build > Add Hydrogen
2. Build > Model Build
3. Setup > Molecular Mechanics > MM+
4. Compute > Geometry Optimisation

2.2 Obtaining QSAR Properties

1. Properties > QSAR Properties > mass, polarisability, refractivity, hydration energy, logP, surface area, volume
2. Properties > Dipole Moment
3. Properties > Total Energy

2.3 Obtaining Orbital Data

1. Setup > Semi-emipirical > Extended Hückel
2. Compute > Single Point
3. Orbitals > HOMO, LUMO, NHOMO, NLUMO

3. Obtaining Connectivity Parameters and Molecular Fragments

This study used Topix (http://www.lohninger.com/topix.html) although there are similar packages that would do the same job. Using the .mol files that were created in ISIS Draw, the compounds can be grouped together in an .sdf file. An .sdf file is the input file to Topix and is a combination of .mol files. An advantage of using an .sdf file is that a number of compounds can be calculated at the same time.

3.1 Creating .sdf Files

1. Notepad > File > Open > 'compoundname.mol'
2. 'compoundname.mol' > Type name of compound at the top of the row of numbers/letters (*this is to identify the compound in the output file if using multiple compounds*) > Save As > 'compoundname.sdf'

If wanting to use multiple compounds in the same input file, then:

3. Notepad > File > Open > 'compoundname2.mol'
4. 'compoundname2.mol' > Type name of compound2 > highlight all text > Copy > Open > 'compoundname.sdf' > paste > Save

3.2 Obtaining Properties in Topix

1. Topix > Load Structure File > 'compoundname.sdf'
2. Connectivity Parameters Tab > select check boxes > Kier-Hall
3. Fragments Tab > select check boxes > Atom counts, bond counts, augmented atoms
4. Calculate Descriptors

When the calculations are complete, the output file is created and can be read in notepad, but is best used in Excel where the values can be copied straight into the new database of properties you are creating:

5. Excel > Open > 'compoundname.asc'

4. Creating K_{oc} Models

This study used statistics package Minitab v14 to build all the models, but other similar programs could be used. By following steps 1, 2 and 3 it should have created a dataset, with a list of compounds with their molecular properties; connectivity parameters; and molecular fragments.

To make a predictive K_{oc} model, the compounds will need to be sorted into chemical groups, with a separate predictive model for each group of compounds. All the data for that particular chemical group should be copied and pasted into a new Minitab document.

A multiple regression model tries to find the relationship between the structural properties and K_{oc}. The predictive model in Chapter 6 was built using information from a Stage 1 report for a particular group of compounds (*e.g.* Benazoles). This information also included their K_{oc} values, as obtained from a batch experiment. These K_{oc} values were necessary to build the initial model for the Benazoles group. The advantage is that if there are any new compounds developed that belong in this chemical group, their structural parameters can be inputted directly into the model. However, the disadvantage is the model is chemical group specific, and if there are any new compounds developed that are not Benazoles, then this particular model cannot be used.

It therefore recommended that the Syngenta records are researched to find structural and K_{oc} information for groups of compounds in order to build up a set of models for different chemical groups. Based on the results of Chapter 6, it is advised that Stage 1 data is a suitable place to start developing a range of models.

4.1 To obtain K_{oc} value for a compound in a chemical group – e.g. FLOW CHART 1

If a model has been created for the relevant chemical group then the appropriate structural properties of the new compound can be substituted into the model. This process is best done in a program like Excel.

If a model is not available for the relevant chemical group and there is structural and K_{oc} data available then a new model will have to be created, if it is anticipated that more compounds in the same group will be developed. See Section 4.2

4.2 To build a new K_{oc} model

1. Minitab > Statistics > Regression > Multiple Regression
2. Response > K_{oc}
3. Predictors > calculated structural parameters
4. Parameters should be added or removed from model, one at a time in a stepwise process based on the p value as shown in the Minitab printout.

Once the models have been developed then a K_{oc} value for any new compound related to that group should be able to be calculated as described in Section 4.1

5. Metabolite Model – *e.g.* FLOW CHART 2

A model for predicting adsorption potential of metabolites has already been created (Equation 7) and is included with these instructions:

$$\ln\left(\frac{\theta}{1-\theta}\right) = 1.84 - 27.3\,{}^{3}_{2}\chi^{c}_{p} + 43\,{}^{0}\chi - 9.13\,{}^{3}_{8}\chi^{c}_{p} - 3.32\mathrm{C1C}$$

(Equation 7)

5.1 Calculating percentage change in metabolite properties

Steps 1, 2, and 3 should be followed to obtain structural properties for parent and metabolite compounds. The percentage change in metabolite properties relative to its parent should be calculated as follows:

((value of parent property – value of metabolite property)/ value of parent property)

The new percentage change values are the variables that will be substituted into the adsorption potential model.

5.2 Calculating the exponential equation and probabilities

Using Excel, the equation for the adsorption potential model can be calculated for each metabolite using the following formula:

$$\mathrm{EXP}(1.84 - 27.3 * {}_{2}^{3}\chi_{p}^{c} + 43 * {}^{0}\chi - 9.13 * {}_{8}^{3}\chi_{p}^{c} - 3.32 * \mathrm{C1C})$$

The calculated value is then used to calculate the predicted probability, which is how the adsorption potential is assessed. The predicted probability can be calculated in Excel:

$$\frac{\mathrm{EXP}(1.84 - 27.3 * {}_{2}^{3}\chi_{p}^{c} + 43 * {}^{0}\chi - 9.13 * {}_{8}^{3}\chi_{p}^{c} - 3.32 * \mathrm{C1C})}{1 + \mathrm{EXP}(1.84 - 27.3 * {}_{2}^{3}\chi_{p}^{c} + 43 * {}^{0}\chi - 9.13 * {}_{8}^{3}\chi_{p}^{c} - 3.32 * \mathrm{C1C})}$$

REFERENCES

Albers, C.N., Banta, G.T., Hansen,P., Jaconsen. O.S. (2009) The influence of organic matter on sorption and fate of Glyphosate in Soil - Comparing different soils and humic substances. Environmental Pollution. 157 2865-2870.

Arias-Estévez, M., López-Periago, E., Martínez-Carballo, E., Simal-Gándara, J., Mejuto, J-C., García-Río,L. (2008) The mobility and degradation of pesticides in soils and the pollution of groundwater resources. Agriculture, Ecosystems and Environment. 123. 247-260.

Baker, J.R., Mihelcic, J.R., Sabljic, A. (2001) Reliable QSAR for estimating K_{oc} for persistent organic pollutants: correlation with molecular connectivity indices. Chemosphere. 45. 213-221.

Behrendtt, H., Brüggemann, R. (1993) Modelling the fate of organic chemicals in the soil plant environment: Model study of root uptake of pesticides. Chemosphere. Vol. 27 (12). 2325-2332.

Bintein, S., Devillers, J. (1994) QSAR for organic chemical sorption in soils and sediments. Chemosphere. Vol. 28 (6) 1171-1188.

British Geological Survey Geology Maps (2012) (Accessed 26/08/12)
http://www.bgs.ac.uk/discoveringGeology/geologyOfBritain/viewer.html

Cooper, J., Dobson, H. (2007) The benefits of pesticides to mankind and the environment. Crop Protection. 26. 1337-1348.

Copping, L.G., Hewitt. H.G (1998) Chemistry and Mode of Action of Crop Proteciton Agents. The Royal Society of Chemistry.

Coupe, R.H., Kalkhoff, S.J., Capel, P.D., Gregoire, C. (2011) Fate and transport of glyphosate and aminomethylphosphonic acid in surface waters of agricultural basins. Pesticide Management Science 68: 16-30.

Clausen, L., Fabricius, I. (2001) Atrazine, Isoproturon, Mecoprop, 2,4-D, and Bentazone adsorption onto iron oxides. Journal of Environmental Quality. 30. 858-869.

Dai, J., Sun, C., Han, S., Wang. L. (1999) QSAR for Polychlorinated Organic Compounds (PCOCs). Prediction of Partition Properties for PCOCs Using Quantum Chemical Parameters. Bulletin of Environmental Contamination and Toxicology. 62. 530-538.

Dai., J., Xu, M., Wang., L. (2000) Prediction of octanol/water partitioning coefficient and sediment sorption coefficient for Benzaldehydes by various molecular descriptor. Bulletin of Environmental Contamination and Toxicology. 65. 190-199.

De Jonge, H., Wollesen de Jonge, L. (1999) Influence of pH and solution composition on the sorption of glyphosate and prochloraz to a sandy loam soil. Chemosphere. Vol 39, 5. 753-763.

De Schamphelaire, M., Spanoghe, P. Brusselman, E., Sonck, S. (2007) Risk assessment of pesticide spray drift damage in Belgium. Crop Protection. 26. 602-611.

Delgado, E.J. (2002) Predicting aqueous solubility of chlorinated hydrocarbons from molecular structure. Fluid Phase Equilibria. 199. 101-107.

Directive 2000/60/EC Establishing a framework for community action in the field of water policy. Official Journal of the European Communities. L327 1-72.

Dubus, I. G., Brown, C. D., Beullke, S. (2003) Sources of uncertainty in pesticide fate modelling. Science of the Total Environment. 317. 53-72.

Edward-Jones, G. (2008) Do benefits accrue to ‘pest control’ or ‘pesticides’: A comment on Cooper and Dobson. Crop Protection. 27. 965-967.

EC Directive 91/414/EEC The placing of plant protection products on the market. Official Journal of the European Communities. L230. 1-32.

EC Directive 98/83/EC On the quality of water intended for human consumption. Official Journal of the European Communities. L330 32-54.

EC Regulation 1107/2009 Concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC. Official Journal of the European Communities L309. 1-50.

EPIWIN (US Environmental Protection Agency)
http://www.epa.gov/oppt/exposure/pubs/episuitedl.htm

Ertli, T., Marton, A. Földényi, R. (2004) Effect of pH and the role of organic matter in the adsorption of isoproturon on soils. Chemosphere. 57. 771-779.

European Commission (2003) Directorate E1- Plant Health. Sanco/221/2000 -rev.10-final. Guidance document on the assessment of the relevance of metabolites in groundwater of substances regulated under council directive 91/414/EEC.

Farenhorst, A. (2006) Importance of soil organic matter fractions in soil-landscape and regional assessments of pesticide sorption and leaching in soil. Soil Science Society of America. 70. 1005-1012.

Fowler, J., Cohen, L., Jarvis, P. (1998) Multivarate Analysis Chapter 18 in Practical Statistics for Field Biology. Wiley.

Gattuso, D.J. (2000) Undestanding the benefits of pesticides. Consumers' Research Magazine. 83 (2). 34-35.

Gevao., B., Semple., K.T., Jones., K.C. (2000) Bound pesticide residues in soils: A review. Environmental Pollution. 108. 3-14.

Ghafoor, A., Jarvis, N.J., Thierfelder, T., Stenström, S. (2011) Measurements and modelling of pesticide persistence in soil at the catchment scale. Science of the Total Environment. 409. 1900-1908.

Gooddy, D.C., Chilton, P.J., Harrison, I., (2002) A field study to assess the degradation and transport of Diuron and its metabolites in a calcareous Soil. Science of the Total Environment. Vol. 297, 1-3. 67-83.

Gooddy, D.C., Mathias, S.A., Harrison, I., Lapworth, D.J., Kim, A.W. (2007) The significance of colloids in the transport of pesticides through chalk. Science of the Total Environment. Vol. 385, 1-3. 262-271.

Gramatica, P., Corradi, M.,Consonni, V. (2000) Modelling and prediction of soil sorption coefficients of non-ionic organic pesticides by molecular descriptors. Chemosphere. 41. 763-777.

Guth, J.A., Reischmann, F.J., Allen, R., Arnold, D., Hassink, J., Leake, C.R., Skidmore, M.W., Reeves, G.L. (2004) Volatilisation of crop protection chemicals from crop and soil surfaces under controlled conditions – Prediction of volatile losses from physio-chemical properties. Chemosphere. 57. 871-887.

Haria, A.H., Hidnett, M.G., Johnson, A.C. (2003) Mechanisms of groundwater recharge and pesticide penetration to a chalk aquifer in Southern England. Journal of Hydrology. 275. 122-137.

Hillocks, R.J. (2012) Farming with fewer pesticides: EU pesticide review and resulting challenges for UK agriculture. Crop Protection. 31. 85-93.

Hiscock, K. (2005) Hydrogeology Principles and Practice. Blackwell Science Ltd.

Holvoet, K.M.A., Seuntjens, P., Vanrolleghem, P.A. (2007) Monitoring and modelling pesticide fate in surface waters at the catchment scale. Ecological Modelling. 209. 53-64.

Hornsby, A.G., Wauchope, R.D., Herner, A.E. (1996) Pesticide Properties in the Environment. Springer New York.

Howell, D.C. (1997) Multiple Regression Chapter 15 in Statistical Methods for Psychology. Duxbury Press.

HyperChem 8.0 (Hypercube Inc.) http://www.hyper.com/Products/HyperChemProfessional/ProFeatures/tabid/362/Default.aspx

Johnson, A.C., White, C., C Lal Bhardwaj. (2000) Potential for Isoproturon, Atrazine and Mecoprop to be degraded within a chalk aquifer system. Journal of Contaminant Hydrology 44: 1-18.

Johnson, A.C., Besien, T.J., Lal Bhardwaj, C., Dixon, A., Gooddy, D.C., Haria, A.H. (2001) Penetration of herbicides to groundwater in an unconfined chalk aquifer following normal soil applications. Journal of Contaminant Hydrology. 53. 107-117.

Jolliffe, I.T. (2010) Principal Component Analysis. Springer London.

Juraske, R., Cateils, F. Vijay, A., Muńoz, P. Antón, A. (2009) Uptake and persistence of pesticides in plants. Measurements and model estimates for Imidacloprid after foliar and soil application. Journal of hazardous materials. 165. 683-689.

Kah, M., Brown, C.D. (2007) Prediction of the adsorption of ionisable pesticides in soils. Journal of Agricultural and Food Chemistry. 55. 2312-2322.

Kamlet, M. J, Doherty, R.M, Abraham, M.H., Carr, P.W., Doherty, R.F., Taft, R.W., (1987) Linear Solvation Energy Relationships. 41. Important differences between aqueous solubility relationships for aliphatic and aromatic solutes. Journal of Physical Chemistry 91. 1996-2004.

Kjaer, J., Ernsfsen, V., Jacobsen, O.H., Hansen, N., Wollensen de Jong, L., Olsen, P. (2011). Transport models and pathways of the strongly sorbing pesticides Glyphosate and Pendimethalin through structures drained soils. Chemosphere. 84. 471-479.

Koplin, D.W., Barbash, J.E., Gilliam, R.J. (2000) Pesticides in groundwater of the United States, 1992-1996. Groundwater. 38. 858-863.

Koplin, D.W., Schnoebelen, D.J. Thurman, E.M. (2004) Degradates provide insight to spatial and temporal trends of herbicides in ground water. Ground Water 42, 4: 601-608

Kumar, M. Philip, L. (2006) Adsorption and desorption characteristics of hydrophobic pesticide Endosulfan in four Indian soils. Chemosphere. 62 1064-1077.

Lapworth, D.J., Gooddy, D.C. (2006) Source and persistence of pesticides in a semi-confined chalk aquifer of South East England. Environmental Pollution. 144. 1031-1044.

Li,K., Liu,W., Xu,D., Lee,S. (2003) Influence of organic matter and pH on Bentazone sorption in soils. Journal of Agricultural and Food Chemistry. 51. 5362-5366.

Lloyd, D. (1989) A First Course in Organic Chemistry. John Wiley and Sons Ltd.

Lohninger. H (1994) Estimation of soil partition coefficients of pesticides from their chemical structure. Chemosphere 29, 8: 1611-1626.

Matthews, G.A. (2008). Attitudes and behaviours regarding use of crop protection products – A survey of more than 8500 smallholders in 26 countries. Crop Protection. 27. 834-846.

McCall, P.J., Laskowski, D.A, Swann, R.L., Dishburger, H.J. (1980) Measurement of sorption coefficients of organic chemicals and their use in environmental fate analysis. Test protocols for environmental fate and movement of toxicants; Proceedings of a Symposium Association of Official Analytical Chemists, 94th Meeting. Pages 88-109.

Meylan, W., Howard, P.H., Boethling, R.S. (1992) Molecular topology/fragment contribution method for predicting soil sorption coefficients. Environmental Science and Technology. 26. 1560-1567.

Moreno-Castilla, C. (2004). Adsorption of organic molecules from aqueous solution on carbon materials. Carbon. Vol. 42 (1) 83-94.

Morgan. R.P.C. (2001) A simple approach to soil loss prediction: A revised Morgan-Morgan-Finney Model. CATENA Vol. (44) 4. 305-322.

Moris, P., Alexandre, I., Roger, M., Remacle J. (1995) Chemiluminesence assays of organophosphorous and carbamate pesticides. Analytica Chimica Acta (302) 53-59.

Müller, M. (1997) Quantum chemical modelling of soil sorption coefficients: Multiple linear regression models. Chemosphere. 35. 365-377.

Nesse, W.D. (2000) Introduction to Mineralogy. Oxford University Press.

Novak, S.M., Portal., J.M., Schiavon, M. (2001) Effects of soil type on Metachlor losses in subsurface drainage. Chemosphere. Vol. 42 (3). 235-244.

OECD (2000) OECD guidelines for the testing of chemicals: Adsorption – desorption using a batch equilibrium method.

Oerke, E.C., Dehne, H.W. (2004) Safeguarding production losses in major crops and the role of crop protection. Crop Protection. 23. 275-285.

Oerke, E.C. (2006) Crop losses to pests. Journal of Agricultural Science. 144. 31-43.

Oliver, D.P., Kookana, R.S., Quintana, B. (2005) Sorption of pesticides in tropical and temperate soils from Australia and the Philippines. Journal of Agricultural and Food Chemistry. 53. 6420-6425.

Oliver, S., Kuet, S. (1999) 'Compound A'. Adsorption and desorption of 'Compound A' in six soils. Zeneca Agrochemicals.

Paszko, T. (2012) Effect on the adsorption of Carbendazim in Polish mineral soils. Science of the Total Environment. Vols 435-436. 222-229.

Patil, G.S. (1994) Prediction of aqueous solubility and octanol-water partition coefficient for pesticides based on their molecular structure. Journal of Hazardous Materials. 36. 35-43.

Pimentel, D., Acquay, H., Biltonen, M., Rice, P., Silva, M., Nelson, J., Lipner, V., Giordano, S., Horowitz, A., D'Amore, M. (1992) Environmental and economic costs of pesticide use. BioScience Vol. 42 (10). 750-760.

Pires, J.C.M., Martins, F.G., Sousa, S.I.V., Alvim-Ferraz, M.C.M., Pereira, M.C. (2008) Selection and validation of parameters in multiple linear and principal component regressions. Environmental Modelling and Software. 23. 50-55.

Pykh, Y.A., Malkina-Pykh, I.G. (1997). POLMOD.PEST – The model of pesticide dynamics in the elementary ecosystems. Ecological Modelling. 98. 215-236.

Ramsden, E.N. (2000). A-Level Chemistry. Nelson Thornes (Publishers) Ltd.

Reddy, K.N. , Locke, M.A (1994a) Prediction of soil sorption (K_{oc}) of herbicides using semi-empirical molecular properties. Weed Science. 42. 453-461.

Reddy, K.N., Locke, M.A. (1994b) Relationships between molecular properties and Log P and soil sorption (K_{oc}) of substituted Phenylureas: QSAR models. Chemosphere. 28 (11). 1929-1941.

Reichenberger, S., Bach, M., Skitschak, A., Frede, H.G. (2007) Mitigation strategies to reduce pesticide inputs into ground and surface water and their effectiveness; A review. Science of the Total Environment. 384. 1-35.

Sabljić, A., Güsten, H., Verhaar, H., Hermens, J. (1995) QSAR modelling of soil sorption. Improvements and systematic of Log K_{oc} vs. Log K_{ow} correlations. Chemosphere. Vol. 31 (11/12) 4489-4514.

Senesi, N. (1992) Binding mechanisms of pesticides to soil humic substances. The Science of the Total Environment. 123/124. 63-76.

Shaw, P.J. A. (2003) Multivariate Statistics for the Environmental Sciences. John Wiley and Sons Ltd.

Singh, N. (2005) Mobility of four triazole fungicides in two Indian soils. Pesticide Management Science 61. 191-196.

Soil Map of England and Wales: Scale 1:250, 000 (1983) Soil Survey of England and Wales. Harpenden.

Spark, K.M., Swift, R.S. (2002) Effect of soil composition and dissolved organic matter on pesticide sorption. The Science of Total Environment. 298. 147-161.

Tariq, M.I., Afzal, S., Hussain, I., Sultana., N. (2007) Pesticide exposure in Pakistan: A review. Environment International. 33. 1107-1122.

Thomas, P.K,. Lane, M.C.G. (1996) Glyphosate Acid: Adsorption and desorption properties in five soils. Zeneca Agrochemicals.

Thomsen, M., Rasmussen, A.G., Carlsen, L. (1999) SAR/QSAR approaches to solubility, partitioning, and sorption of phthalates. Chemosphere. Vol. 38 (11). 2613-2624.

Tomlin, C.D.S. (1997) The Pesticide Manual: A world compendium. 11th edition, British Crop Protection Council. Farnham, UK.

Tixier, C., Bogaerts, P., Sancelme, M., Bonnemoy, F., Twagilimana, L., Cuer, A., Bohatier J., Veschambre, H., (2000) Fungal biodegradation of a Phenylurea herbicide, Diuron : Structure and toxicity of metabolites. Pest Management Science. Vol. 56 (5). 455-462.

Van Beinum, W., Beulke, S., Brown, C.D. (2005) Pesticide sorption and diffusion in natural clay loam aggregates. Journal of Agricultural and Food Chemistry. 53. 9146-9154.

Van der Brüggen, B., Braeken, L., Vandecasteele, C. (2002) Evaluation of parameters describing flux decline in nanofiltration of aqueous solutions containing organic compounds. Desalination. 147. 281-288.

Van der Linden, A.M.A., Tiktak, A., Boesten, J.J.T.I, Leijnse, A. (2009) Influence of pH dependent sorption on simulated pesticide leaching. Science of the Total Environment. 407. 3415-3420.

Van der Werf, H.M.G. (1996) Assessing the impact of pesticides on the environment. Agriculture, Ecosystems and Environment. 60. 81-96.

Von Oepen, B., Kördel, W., Klein, W. (1991). Sorption of nonpolar and polar compounds to soils: Processes, measurements and experience with the applicability of the modified OECD-guideline 106. Chemosphere. Vol.22 (3-4). 285-304.

Wang, C.J., Liu, Z.Q. (2007) Foliar uptake of pesticides – Present status and future challenge. Pesticide Biochemistry and Physiology. 87. 1-8.

Wang. F. (2009) Factor analysis and principal component analysis. International Encyclopaedia of Human Geography. 1-7.

Ward, R.C., Robinson, M. (2000) Principle of Hydrology. McGraw Hill International (UK) Ltd.

Wauchope, R.D., Yeh, S., Linders, J.B.H.J., Kloskowski, R., Tanaka, K., Rubin, B., Katayama, A. Kördel, W., Gerstl, Z., Lane, M., Unsworth, J.B. (2002) Pesticide soil sorption parameters: Theory, measurement, uses, limitations and reliability. Pesticide Management Science. 58. 419-445.

Weber, W.J., McGinley, P.M., Katz, L.E. (1991) Sorption phenomena in subsurface systems: Concepts, models and effects on contaminant fate and transport. Water Research 25 (5) 499-528.

Wilson, C., Tisdell, C. (2001) Why farmers continue to use pesticides despite environmental, health and sustainability costs. Ecological Economics. 39. 449-462.

Worrall, F. (2001). A molecular topology approach to predicting pesticide pollution of groundwater. Environmental Science and Technology 35, no. 11 2282-2287.

Worrall, F., Thomsen, M., (2004). Quantum vs. topological descriptors in the development of molecular models of groundwater pollution by pesticides. Chemosphere 54 585-596.

Yariv, S., Cross, H. (2002). Organo-clay complexes and interactions. Marcel Dekker, Inc.

Yaron, B. (1989) General principles of pesticide movement to groundwater. Agriculture, Ecosystems and Environment. 26. 275-297.