Revision of the Statistical Method to Identify Pesticide Chemicals with Potential to Move through Soil to Groundwater in California

Study 287

John Troiano, Murray Clayton, Craig Nordmark
Environmental Monitoring Branch
Groundwater Protection Program

Jonathan Sullivan
Pesticide Evaluation Branch
Chemistry Program

Department of Pesticide Regulation
California Environmental Protection Agency

July 2024

EXECUTIVE SUMMARY

Background

Detections of pesticide residues in groundwater in the late 1970s resulted in enactment of the Pesticide Contamination Prevention Act (PCPA) in 1985 by the California legislature. As indicated by its title, the goal of the legislation is to protect California's groundwater resources from further contamination. When first adopted, the PCPA mandated that the California Department of Pesticide Regulation (DPR) establish the Groundwater Protection List (GWPL) — a list of pesticides that have the potential to pollute groundwater — and provided very specific requirements for how the chemicals would be identified. The PCPA required DPR to develop threshold values, termed specific numerical values (SNV), which were developed for several physical-chemical properties to identify the potential for a chemical to move to groundwater. Threshold values were calculated using a statistical comparison of the distributions of data for each chemical property between pesticide chemicals recognized as known leachers and non-leachers (Johnson, 1991). All chemicals used to develop the SNVs had agricultural applications that could result in movement to groundwater. Chemicals identified as leachers were those that have been detected in well water samples, and chemicals identified as non-leachers were those that were sampled for but not detected. The SNVs developed by Johnson in 1991 were used to determine which pesticides are placed on the GWPL. DPR's Groundwater Protection Program (GWPP) uses the GWPL, codified in Title 3 of the California Code of Regulations section 6800(b), to prioritize pesticide chemicals for subsequent well sampling studies. Since the Johnson (1991) report, additional well monitoring studies have determined the presence or absence of pesticide chemicals in California's groundwater, leading to a revision of chemicals identified by DPR as known leachers and non-leachers. The revised categorization of leacher and non-leacher chemicals provided a basis for this proposed revision of the SNV procedure. Furthermore, revisions to the PCPA in 2015 allowed greater flexibility for refining the SNV process used to identify potential leachers and required the identification of pesticide degradation products with the potential to contaminate groundwater. These changes to the PCPA allowed for the investigation of additional chemical properties not specified in the original PCPA and of an alternative and more discriminating statistical approach to identifying potential leachers.

Methodology

The SNV threshold values developed by Johnson (1991) were generated using the lists of leacher and non-leacher pesticide chemicals that were established from well monitoring studies conducted throughout the contiguous United States (U.S.). For this present study, analyses were limited to well monitoring data from studies conducted within California by DPR and other agencies because of greater reliability of the data generated from these studies. Specifically, there was: 1) greater access to analytical quality control data, 2) knowledge of the condition of the wells sampled, and 3) the ability to demonstrate a connection between analytical results obtained from a well sample and pesticide applications made near the wellhead.

The SNV values developed by Johnson (1991) relied on data generated for five empirically-measured chemical properties that reflect mobility and persistence in soil. Data for these empirical chemical properties were derived from laboratory or field studies. As required by the PCPA, when pesticide registrants submit products with new active ingredients for California registration, they submit these data to DPR. In addition to several of the original properties used in the SNV process, some additional empirically-derived chemical-specific properties were investigated in this revision of the process. Various quantum-derived, chemical-specific properties based on the molecular structure and activity of the chemical were also investigated. The quantum- or molecular-chemical properties are calculated by modeling software and result in less variability than the empirical properties that can be influenced by inconsistencies in environmental conditions, study management practices, and measurement errors.

The SNV procedure is statistically univariate in nature because it individually compares five chemical-specific properties to each of five corresponding threshold values. Exceedance of a threshold value related to at least one mobility-related, chemical-specific property and one persistence-related, chemical-specific property identifies a chemical as having potential to leach to groundwater. As an alternative methodology to the SNV approach, a multivariate approach was investigated to discriminate between the known leacher and non-leacher chemicals. This multivariate procedure generates a single threshold value utilizing a function containing multiple chemical-specific properties where exceedance of this threshold value identifies the chemical as having potential to leach to groundwater.

Results

Revision of Leacher and Non-Leacher Chemicals

The revised categorization of leacher and non-leacher chemicals resulted from additional well sampling data generated in California since the report by Johnson (1991) and from improvements in chemical analytical methodology. The following changes were made to the known leacher and non-leacher lists of chemicals used by Johnson (1991):

- Eleven chemicals were removed from consideration due to lack of available physicochemical properties data:
    - Dieldrin, DDT, DDD, DDE, ametryne, chlordane, heptachlor, lindane, naled, silvex, toxaphene
- Fourteen chemicals were identified as leachers based on detections in subsequent well sampling studies:
    - Parent chemicals — norflurazon, hexazinone, imidacloprid, tebuthiuron
    - Degradation products — aldicarb sulfone, aldicarb sulfoxide, alachlor ethanesulfonic acid (ESA), metolachlor ESA, metolachlor oxanilic acid (OXA), deethylatrazine, deethylsimazine, diaminochlorotriazine, desmethylnorflurazon, 2,3,5,6-tetrachloroterephthalic acid
- Six chemicals were reclassified from leacher to non-leacher status:

    - Analytical chemistry updates separating parent from its degradate — chlorthal-dimethyl (DCPA)
    - Subsequent well sampling conducted with no detection — alachlor, carbofuran, cyanazine, fonofos, oxamyl
- Six chemicals were identified as non-leachers based on lack of detections in subsequent well sampling studies despite ongoing agricultural use in the sampling areas:
    - Iprodione, methomyl, napropamide, oryzalin, oxyfluorfen, thiobencarb

From the well monitoring studies conducted in California, 25 chemicals have been detected and subsequently identified as leachers. These include many previously identified by Johnson (1991), but also include the four parent chemicals and ten degradation products discussed above. However, only 18 of these chemicals were eventually used in this study as known leachers because the remaining seven chemicals had incomplete datasets of physicochemical properties (Table 1).

Thirty-two chemicals have been identified as non-leachers. These include many previously identified by Johnson (1991) along with the six reclassified and six newly identified parent chemicals indicated above. However, in this study only 24 of these chemicals were used as non-leachers because the remaining chemicals had incomplete datasets of physicochemical properties (Table 2).

Revised Statistical Method to Determine a Chemical’s Leaching Potential

Investigations into the quantum- and empirically-derived chemical properties resulted in 29 chemical-specific properties for potential inclusion in the multivariate analysis. All 29 chemical properties were subjected to statistical testing prior to inclusion in the multivariate analysis. The statistical testing ensured that the property values were normally distributed, that actual differences between known leacher and non-leacher chemicals existed for each property, and that the chemical properties were not highly correlated. The latter would identify potentially redundant information as reflected in a correlation analysis. Eleven of the 29 chemical properties were eventually identified for inclusion in the initial multivariate analysis using a canonical discriminant analysis (CDA) procedure. In this procedure, the 11 chemical properties were analyzed together as a group to determine if a significant difference existed between the distributions of the CDA scores for the known leacher and non-leacher chemicals. This initial CDA analysis indicated a highly significant difference between the scores for the leacher and non-leacher groups. A threshold value to identify a leacher was then generated from the distribution of scores for the leacher chemicals. Since the distribution of the leacher group was shifted to the right of the distribution of the non-leacher group due to its higher scores, the threshold value was determined as the lower prediction limit at the 95 percent confidence level of the leacher distribution. The 11-parameter multivariate model predicted the correct leaching status of all members in the leacher group of chemicals, and all but two members of the non-leacher group of chemicals. The error rate of misclassification was 2 out of 42 chemicals, or approximately 5%.

Subsequent analyses determined whether a CDA model containing fewer chemical properties could also effectively discriminate between the known leacher and non-leacher groups of chemicals. Sequential reduction of the least influential chemical properties identified a model utilizing five chemical properties that had the same misclassification rate (5%) as the full model with 11 chemical properties. Cross-validation of the CDA models using formal assessments, termed leave-one-out and leave-pair-out analyses, indicated that the five-parameter model was highly stable compared to models containing fewer than five parameters. The five chemical properties in this model are the empirically-derived properties of soil aerobic half-life, carbon-normalized soil adsorption coefficient (K_{oc}), and quantum-calculated properties of dipole moment, energy of highest occupied molecular orbital (E_{HOMO}), and maximum electrostatic potential (MaxElPot). Consistent with findings from other studies investigating the fate of chemicals in the environment, a measure of a chemical's environmental persistence is reflected by soil aerobic half-life and E_{HOMO}. Potential for soil mobility is reflected by K_{oc}, dipole moment, and MaxElPot. Inherent variability in chemical-specific property values derived from empirical experimentation, such as those used exclusively in the SNV process, do not exist with the quantum or molecular properties. The dipole moment, E_{HOMO}, and MaxElPot are chemical properties whose values are calculated from the molecular structural information of a chemical and therefore provide added stability to the CDA model.

Comparison to the SNV Process

The SNV procedure utilizes compounding conditions from two categories of chemical properties to predict a chemical's leaching status. One category relates to soil mobility and includes water solubility and K_{oc}. The other relates to environmental persistence and includes aerobic and anaerobic soil metabolism half-lives and hydrolysis half-life. A pesticide is classified as having leaching potential if one threshold value from a chemical property in both categories is exceeded.

Johnson (1991) reported that with the SNV procedure, 14 of 27 non-leacher pesticides were misclassified as predicted leachers. Conversely, 4 of 23 leacher pesticides were misclassified as predicted non-leachers. Together, this resulted in 18 of 50 chemicals misclassified by the SNV procedure, giving an overall misclassification rate of 36%.

As indicated above, the new multivariate model with five chemical properties misclassified two non-leacher chemicals out of the combined total of 42 known leacher and non-leacher chemicals, giving an overall misclassification rate of approximately 5%. The procedure developed in this present study significantly improves the identification of a chemical's potential to leach to groundwater.

Process to Identify the Potential Leaching Status of a Chemical Using the Multivariate Procedure

1. In the presence of multiple values, utilize the median value for the soil aerobic half-life and K_{oc} of the chemical.

2. Calculate the dipole moment, E_{HOMO}, and MaxElPot values of the chemical from its molecular structure in the equilibrium state. The Spartan '20 computer program was used in this study to derive these values.

3. Derive the Multivariate Leaching Value (MLV) for the chemical as:

$$\mathrm{MLV} = 1.9349\,(\log_{10}\text{ soil aerobic half-life}) - 1.2197\,(\log_{10} K_{oc}) + 0.1562\,(\text{dipole moment}) - 0.9140\,(E_{HOMO}) + 2.2728\,(\log_{10}\text{ MaxElPot})$$

4. Compare the MLV score for the chemical to the Multivariate Leaching Threshold (MLT) value of 14.4706. The MLT value is derived from the distribution of MLV scores for the group of known leacher chemicals. If the MLV ≥ MLT, then the chemical is identified as a member of the leacher group of chemicals with potential to move into California's groundwater.

TABLE OF CONTENTS

INTRODUCTION

Background for the Development of a Procedure to Identify Potential Groundwater Contaminants

In the late 1970’s, chemical ingredients in pesticide products used for agriculture were detected in California groundwater resulting in the enactment of the Pesticide Contamination Prevention Act (PCPA) of 1985. As implied by its name, the goal of the legislation is to protect California’s groundwater resources from further contamination by agriculturally-applied pesticides. One key directive in the legislation mandates that the California Department of Pesticide Regulation (DPR) identify pesticide active ingredients registered in California that have a potential to move to groundwater. In the initial legislation, the PCPA identified specific environmental fate properties that must be used to identify a chemical’s potential to leach to groundwater. Based also on their agricultural use patterns, specific pesticide active ingredients were then identified as potential leachers and added to the Groundwater Protection List (GWPL) [Title 3 of the California Code of Regulations (3CCR) section 6800(b)], which references the California Food and Agricultural Code (FAC) sections 13144, 13145, and 13149. As required by the PCPA, DPR’s Groundwater Protection Program (GWPP) conducts well water sampling studies for chemicals on the GWPL to determine their presence in California’s groundwater. The design of well sampling studies for chemicals on the GWPL includes additional data describing patterns of pesticide use, agricultural management practices, and information on potential geographical vulnerability for movement of chemicals to groundwater. DaSilva (2018) provides an example of a well sampling study design to monitor for the potential presence of a pesticide in groundwater.

The GWPL is periodically updated in regulation to include chemicals in newly registered pesticide products. The procedure to determine if a chemical qualifies for listing on the GWPL was developed from physical and chemical properties identified as important determinants for the potential of a chemical to leach to groundwater. The approach originally outlined in the PCPA was developed from observations in an earlier United States Environmental Protection Agency (U.S. EPA) report by Cohen et al. (1984). That report listed water solubility, soil adsorption coefficient normalized for organic carbon content of soil (K_{oc}), Henry’s law constant (K_h), speciation, hydrolysis half-life, soil photolysis half-life, and soil metabolism half-life as important chemical characteristics. Ranges were provided for values of the chemical properties thought to identify a pesticide active ingredient as a potential leacher. These values were estimates and were not derived through an analytical procedure. A more rigorous approach was developed by DPR combining the chemical properties into two groups, where one group reflected a chemical’s potential for mobility through soil and a second group reflected persistence in the environment. A threshold value was developed for each chemical property using a statistical procedure. Exceedance of at least one threshold value from each of the soil mobility and environmental persistence groups identified the chemical as a potential leacher. The active ingredient underwent further investigation with respect to its agricultural use patterns to qualify for placement on the 3CCR section 6800(b) list.

Wilkerson and Kim (1986) originally developed the statistical approach to determine the threshold values for each chemical property. They identified two lists of chemicals. One was denoted as leachers and contained pesticide chemicals detected in well water samples. The second was denoted as non-leachers and contained pesticide chemicals with use patterns that facilitated opportunity for movement to groundwater but where residues had not been detected in well water samples. Chemicals on the lists reflected detections or non-detections in groundwater resulting from general agricultural use; chemical detections originating from point sources were not included on the leacher list because point source contamination does not necessarily indicate a chemical's persistence or mobility in the soil environment. Chemical-specific values for the pesticide properties specified in the PCPA were then collected for each chemical. University of California staff with expertise in environmental toxicology independently reviewed the well monitoring and chemical properties data to authenticate each chemical's leacher or non-leacher status and to verify the data used in the statistical analysis. Upon normalization of the distribution of values to base 10 logarithm, a hypothesis test evaluated potential statistical differences between means for the leachers and non-leachers for each chemical property. For those properties indicating statistical difference between the two groups, a threshold value was calculated as the antilogarithm of the logged value at either the 10 or 90 percent confidence interval of the distribution of leacher chemicals. Selection of the 10 or 90 percent confidence interval depended on the location of the distribution of the chemical property for the leacher group relative to the non-leacher group (Wilkerson and Kim, 1986; Johnson, 1988, 1989, and 1991). The threshold values were designated Specific Numeric Values (SNVs) and enacted into regulation (3CCR section 6804) in 1989. Several amendments to the regulations followed with the final amendment occurring in 1991. A pesticide active ingredient is currently designated a potential leacher if it exceeds at least one soil mobility SNV threshold value and at least one environmental persistence SNV threshold value. Johnson (1991) calculated the misclassification rate of the SNV process based on reclassification of the chemicals on both lists. The SNV process misclassified four of 23 leachers as non-leachers and 14 of 27 non-leachers as leachers. In total, 18 of 50 pesticides were misclassified giving an overall misclassification rate of 36%.

Revision of the Process to Identify Chemicals with High Leaching Potential

Since development of the SNVs in 1991, many well water sampling studies have been conducted in California and other states to determine the presence of chemicals in groundwater. The PCPA was updated in 2015 and allows greater flexibility in the process used to identify chemicals as potential leachers and added the requirement that degradation products also be evaluated for their potential to contaminate groundwater. This current revision of the process includes:

- Updating the lists of chemicals determined as known leachers or non-leachers.
- Expanding the list of chemical properties that could reflect chemical leaching potential.
- Updating the statistical methodology used to predict and classify chemicals as potential leachers or non-leachers.

Updating Chemicals Listed as Leachers or Non-Leachers

Chemical properties data used to develop the SNV procedure were based on pesticides from groundwater monitoring studies conducted prior to the Wilkerson and Kim (1986) report. Since then, subsequent monitoring studies have detected additional pesticide residues in California well water. Many of these studies were conducted by GWPP to determine potential groundwater contamination by chemicals on the 3 CCR section 6800(b) list. As required by the PCPA, GWPP developed and maintains a database for all groundwater sampling of pesticides conducted in California by state or other government agencies. When pesticide residue detections are reported in well water by another agency, GWPP investigates the veracity of the detections and the extent of the groundwater contamination. If submitted laboratory quality assurance and quality control (QA/QC) data indicate that the well water sample is the source of the detection, the well is typically resampled and nearby wells may be sampled to determine the areal extent of groundwater contamination.

Recent improvement in chemical analytical methodology has revealed earlier misidentification of chemicals on the lists of known leachers and non-leachers used by Wilkerson and Kim (1986) and Johnson (1988, 1989, and 1991). For example, chlorthal-dimethyl (DCPA) was initially identified as a leacher by Wilkerson and Kim (1986). However, improvement in the analytical methodology to identify parent and degradation products later determined that the original detections of DCPA were 2,3,5,6-tetracloroterephthalic acid (TPA), which is a degradate of DCPA (Monohan et al., 1995). The propensity for TPA, rather than its parent, to move to groundwater was supported by results of a national well sampling program conducted between 1988 and 1990 by U.S. EPA (U.S. EPA, 1992; U.S. EPA, 2008). In that study, TPA was the most widely-detected chemical in groundwater; no residues of the parent DCPA were detected. This result was further supported by domestic well water sampling conducted in California by GWPP where TPA residue was detected and DCPA was not (Ando, 1992; Ruud, 2021). Consequently, for this current study DCPA was reclassified to known non-leacher status while TPA was classified as a known leacher.

Expanding the List of Chemical Properties

As initially prescribed in the PCPA, the SNV threshold values were generated through analysis of physical and chemical properties data required to be submitted to U.S. EPA and DPR for pesticide product registration. Revision of the PCPA in 2015 allowed for inclusion of additional chemical properties and alternative statistical methodology to develop threshold indicator(s) as long as the revised process was subject to independent peer review [FAC section 13145(e)]. Previous investigations have identified alternative chemical properties that effectively relate to a chemical’s leaching potential. Randic (1975) and Worrall (2001) used measures of molecular connectivity to identify pesticide chemicals with greater leaching potential. Following those investigations, Worrall and Thomsen (2004) used a combination of quantum and topological variables to further identify chemicals with greater leaching potential. For this present revision, the evaluation included investigating chemical-specific quantum properties. Topological and connectivity properties were not included because the multivariate analysis utilized in this study is a parametric procedure that requires data for each property to approximate a normal distribution. The dataset for many topological and connectivity properties are not normally distributed. Base 10 logarithm transformations of these data were either unsuccessful or often not possible because of the large proportion of zero values in the datasets.

Updating the Statistical Methodology

The SNV procedure utilizes a univariate approach where, for each new chemical submitted for registration in California, the median value of each chemical property is compared to its respective SNV threshold value. For this revision of the SNV procedure, the use of a multivariate statistical approach to distinguish between leaching and non-leaching chemicals is proposed. In a multivariate approach, discrimination between groups is based on a single value generated from two or more chemical property values. For example, Gustafson (1989) generated a single score for each pesticide, denoted as the Groundwater Ubiquity Score (GUS), from its K_{oc} and field- or laboratory-derived soil half-life values. The function that generated the scores also identified threshold GUS values that discriminated between the same known leacher and non-leacher pesticides previously identified by Wilkerson and Kim (1986). The GUS index is an example of a single pesticide-specific score calculated from two chemical properties where the score is compared against threshold scores identifying a chemical as a potential leacher, transitional leacher, or non-leacher. Similarly in this present revision, a multivariate statistical approach produced an equation to derive a single score generated from multiple chemical properties for each chemical tested. The chemical's score is then compared against a threshold score identifying the chemical as a potential leacher or non-leacher. The updated groupings of known leacher and non-leacher chemicals provide the basis for the development of an effective multivariate model of chemical properties.

METHODOLOGY

Well Monitoring Data Identifying Chemicals as Leachers or Non-Leachers

Two important aspects of a well water sampling study that are evaluated to determine if reported detections result from nonpoint-source agricultural applications are:

1) Wells are sampled in areas where pesticide use has occurred, providing a physical connection between agricultural applications and detection in a well sample. The connection between agricultural pesticide use and subsequent detection of residues in groundwater is evident by the extent of reported detections of pesticides in DPR's Well Inventory Database (DPR, 2020) that have exclusive agricultural uses; and

2) QA/QC data generated during the study validate the results, assuring that the pesticide detection is associated with the well water sample. The QA/QC data should rule out laboratory or field contamination as a source of the reported detection. For example, detection limits for chemical analytical methodology are relatively low with values frequently reported in the parts-per-trillion range (ng/L). At these levels, residues on glassware or residual chemicals in chromatographic columns used for sample analysis could lead to detections of low concentrations where the source is from laboratory equipment and not from an actual environmental sample, such as well water.

Initially, data from well sampling conducted throughout the U.S. were to be utilized for determining pesticide chemical potential to move to groundwater due to agricultural use. However, verification of detections reported in studies conducted outside of California was problematic because:

- Inspection of laboratory QA/QC data indicated potential issues with determining well water samples as sources of detections.
- Most of the wells were not resampled to verify detections.
- Previous agricultural pesticide use around the well was not verified.
- The type of well, the well's condition, or the structural integrity of the well was not reported.

Lack of data on the type and condition of a well is problematic. For example, detections in extremely shallow monitoring wells may not translate to detections in domestic well water, which is typically deeper. In addition, detections could be due to point source contamination resulting from damaged or poor well construction rather than from nonpoint-source agricultural applications.

Based on these factors, this current study uses California well monitoring data included in GWPP's Well Inventory Database (DPR, 2020). This database contains DPR's well monitoring data, and monitoring conducted by other agencies such as the California State Water Resources Control Board (SWRCB), and the U.S. Geological Survey (USGS). The USGS initiated nationwide well water sampling for pesticide residues and other constituents in groundwater through the National Water-Quality Assessment Program (NAWQA). In California, the program was expanded to address a mandate by the State legislature for SWRCB to provide an assessment of the condition of all groundwater basins in California. This program was designated as the Groundwater Ambient Monitoring Assessment (GAMA) Program. DPR created the Well Inventory Database to comply with the PCPA, which requires all state agencies to submit well water sampling data to DPR when measuring for the presence of pesticide chemical residues.

Studies conducted by DPR provide a high level of confidence in a chemical identified as detected or not-detected in groundwater because:

- Well sampling is conducted in areas of known pesticide use — DPR maintains a database of spatial applications of agricultural pesticides and well sampling is targeted to areas where pesticides are known to be used.
- Single-family domestic wells are targeted for sampling — many single-family domestic wells are located in rural areas surrounded by potential nonpoint-source, agriculturally-applied pesticides and have water drawn from shallow, more vulnerable aquifers.
- Well sampling is conducted in areas of known vulnerability — wells are targeted in areas that have been identified as vulnerable to pesticide leaching.
- Extensive QA/QC samples are taken:
    - Reported detections in primary samples are verified by analysis of backup samples.

  - Field-collected blank samples are analyzed to determine if detections in well water samples are the result of potential contamination at the sampling site due to handling and sample collection procedures or during transportation of the sample.
  - Laboratory QA/QC samples are analyzed to provide information on the potential for laboratory contamination.

Owing to the large number of chemical analyses conducted for the GAMA Program, design elements for well sampling and study QA/QC included:

- Multi-chemical screen analysis — samples taken were subject to analytical screens that provide data on a large list of inorganic and organic chemicals thereby requiring a more robust and intensive QA/QC methodology.
- Streamlined QA/QC data — wells with detections were not resampled, nor were backup samples available for analyses due to the large number of chemical analyses conducted. Field-blank water samples accounted for only 10% of the total well water samples collected from random wells.

Fram and Stork (2019) recently revised previously reported detections in well water samples collected under the GAMA Program for the years 2004 through 2018. For their report, the initial well sampling dataset was subject to an updated five-step QA/QC procedure. The revised procedure included information generated over time from laboratory blank samples and was shown to affect the outcome of well water sampling data generated by the NAWQA Program (Medalie et al., 2019). Application of the revised QA/QC methodology for the GAMA Program resulted in changing the designation of some detections to non-detected status. The modified data from that report were used in this current study to determine the detected or non-detected status of pesticide chemicals. However, some of the chemicals in the revised report had sparse detections and were subjected to further investigation (Troiano and Clayton, 2022).

Pesticide Physical, Chemical, and Quantum-Calculated Properties Investigated

U.S. EPA requires studies on various physical and chemical properties of pesticide active ingredients prior to federal registration. The PCPA mandates that product registrants submit these studies to DPR as a requirement for active ingredients in pesticide products going through registration in California. Data for some of these physical and chemical properties including aerobic and anaerobic soil metabolism half-life, hydrolysis half-life, water solubility, and K_{oc} are used in the SNV procedure developed in 1991. Other environmental fate data of active ingredients required for registration of pesticide products by U.S. EPA and DPR that were included in this revision, but not utilized in the SNV process, were terrestrial field dissipation (TFD) half-life, octanol-water partition coefficient (K_{ow}), vapor pressure (VP), and Henry's law constant (K_h). The study data required by U.S. EPA and DPR for registration purposes are typically derived experimentally and thereby designated as empirical data in this report. Data submitted to satisfy registration requirements of a pesticide product in California were the primary source for the empirical chemical properties data. Where data were unavailable from registration submissions, such as for many of the degradation products and some active ingredients of pesticides registered prior to the PCPA and exempted from data submission, reliance on other data sources included U.S. EPA Re-registration Eligibility Documents (RED), open scientific literature, and the University of Hertfordshire's Pesticides Properties Database (PPDB). PPDB is a European-based database used to support risk assessment and risk management largely for registration purposes (https://sitem.herts.ac.uk/aeru/ppdb/en/atoz.htm). Documentation for the data submitted to PPDB has been generated in individual reports by European member states.

Some empirical properties were not included in this investigation because of insufficient data or because the test protocol utilized in the study indicated that the property would not discriminate between the known leacher and non-leacher chemicals. For example, measures for anaerobic soil metabolism half-life and aqueous and soil photolysis half-lives were not included because these data were not available for a substantial portion of the degradation products. Hydrolysis half-life was not included in this investigation because the U.S. EPA test protocol only requires a 30-day study duration period. Many pesticides do not hydrolyze sufficiently within 30 days to calculate a meaningful hydrolysis dissipation rate. Consequently, hydrolysis is often noted as stable with no numeric value indicated and thus has limited discriminatory power.

The chemical properties investigated were further broadened to include molecular properties of chemicals derived from computational quantum modeling methods. These computations require three-dimensional (3-D) conformation as input and describe many aspects of molecular structure and activity. For example, surface area and volume descriptors are obtained directly from the molecular wave function as calculated from the graphical surface of the electron density map. To derive the quantum properties, the molecular structure of the chemical was imported into the Spartan '20 computer program (available at: http://wavefun.com) in spatial data file notation to simulate its 3-D structure in the equilibrium state. Chemical property calculations based on molecular structure and activity have been previously investigated for use as predictors of environmental fate (Worrall, 2001; Worrall and Thomsen, 2004).

Seven empirical and 22 molecular chemical properties were considered for the analysis, as summarized below:

- Empirical properties: water solubility, K_{oc}, laboratory soil aerobic half-life, TFD half-life, K_{ow}, K_h, and VP.
- Molecular properties: PolarArea75, AccPolarArea75, PolarArea100, AccPolarArea100, PolarArea125, AccPolarArea125, PSA, Dipole, MaxElPot, MinElPot, Polarizability, Molecular weight, SurfArea, SurfVolume, AccSurfArea, E_{HOMO}, Hardness, E_{LUMO}, Electronegativity, EAU, MinLocIonPot, and Conformers.

Appendix I contains a description of each chemical property referenced in this report and its associated abbreviation. Appendix II contains the value used for each chemical property with a reference for the source of the value.

Statistical Methodology

The statistical approach used in this study consisted of five sequential steps which are expanded on in the following sections:

1. Screening each chemical property for distribution normality, testing statistical differences for each chemical property between the known leacher and non-leacher groups of chemicals, and correlation testing between chemical properties to reduce redundancy in the dataset prior to multivariate analysis testing.
2. Canonical Discriminant Analysis (CDA) to identify a statistically significant and efficient multivariate model from the chemical properties that discriminates between the known leacher and non-leacher chemicals.
3. Calculation of a CDA score for each chemical, denoted as its Multivariate Leaching Value (MLV), which is calculated using chemical property values specific to the chemical.
4. Use of the statistical prediction interval to calculate a universal MLV threshold score, denoted as the Multivariate Leaching Threshold (MLT), where a chemical's MLV is compared to the MLT to determine its predicted leaching or non-leaching classification.
5. Cross-validation of candidate multivariate models using the 'leave-one-out' and 'leave-pair-out' analyses.

1. Screening Chemical Properties to Include in the Multivariate Statistical Analysis

There were 29 chemical properties to potentially include in the CDA analysis. This number of predictor variables was greater than the number of chemicals, or group members classified as leachers, which violated the prerequisites of a discriminant function analysis. Therefore, a reduction of chemical properties was required to produce a valid CDA analysis. Reduction in the number of properties was based on sequential evaluation of each chemical property, first testing for distribution normality of the data, then measuring potential to discriminate between the leacher and non-leacher classified chemicals, and lastly determining redundancy in information provided between the properties. The CDA procedure in the Statistical Analysis System (SAS, 2008) used in this revision was based on parametric statistics and required tested variables to approximate a normal distribution (PROC CANDISC, SAS, 2008). Each empirical and quantum or molecular chemical property was subject to a test of distribution normality with separate tests conducted within the leacher and non-leacher groups (PROC CAPABILITY, SAS, 2008). When the main test statistic, Shapiro-Wilk, was marginally significant for just one of the groups (leachers or non-leachers), inspection of quantile-quantile (Q-Q) plots indicated the extent of deviation from normality. A chemical property was not rejected if the Q-Q plot of the dataset illustrated only marginal deviation from normality. Where necessary, data transformed to base 10 logarithms (Log_{10}) were also subject to normality tests. A chemical property was excluded in subsequent analyses if the raw or transformed data did not approximate a normal distribution.

Chemical properties that approximated a normal distribution were subject to Student's *t*-tests (PROC T-TEST, SAS, 2008) for hypothesis testing of the means between the leacher- and non-leacher-classified chemicals. Typically, a *t*-test *P*-value of ≤ 0.05 indicates statistical significance. However, for this analysis a relaxed *P*-value of ≤ 0.15 was chosen for statistical significance when testing if a chemical property was retained for further statistical analysis. This relaxed significance level ensured inclusion of chemical properties that, despite having marginal insignificance in the *t*-test, may provide power of discrimination when used in the multivariate analysis.

Lastly, chemical property values that were normally distributed and with a *t*-test *P*-value ≤ 0.15 between the leacher and non-leacher groups were entered into a correlation analysis to identify potential redundancy between the properties. The correlation analysis was conducted on the combined data from the leacher and non-leacher groups. Based on a Pearson's correlation coefficient of ≥ |0.7|, several groups of highly correlated chemical properties were observed. A representative chemical property from within each correlated group was chosen based on the strength of its *t*-test as indicated by the *P*-value significance, and how representative it was of the group of correlated variables. The chemical property with the greatest number of Pearson's correlation coefficients ≥ |0.7| was considered most representative of the group — the remaining chemical properties in the correlated group were excluded from further analyses. Chemical properties with coefficients < |0.7|within the whole correlation matrix were considered sufficiently uncorrelated and were retained in the study.

2. Canonical Discriminant Analysis

All chemical properties retained from the final correlation analysis were included in the initial CDA multivariate procedure. Canonical discriminant analysis is a multivariate technique that is used to construct a discrimination space for maximum separation of class variables. Accordingly, relationships can be established between a categorical variable and a group of independent variables (Zhao and Maclean, 2000). The CDA procedure (PROC CANDISC, SAS, 2008) by default standardizes the data for each variable to zero mean and unit variance, then implements a multivariate analysis of variance which provides an overall test for significant difference between the categorical leacher and non-leacher groups of chemicals. Resulting standardized coefficients are back transformed to their original variable space to be applied as coefficients to the raw chemical properties data. Upon verification of significant statistical differences, subsequent CDA analyses were conducted to determine if fewer chemical properties were as effective in discriminating between the leacher and non-leacher groups. Each CDA analysis provided a set of coefficients denoted as the Total-Sample Standardized Canonical Coefficients (standardized coefficients) that indicated the relative influence of each chemical property to the discrimination process: the magnitude of the coefficients represents the relative importance of the variable in discriminating the leachability category, and the sign determines the direction of the correlation. The chemical property with the lowest standardized coefficient in terms of absolute value was excluded from the next iteration of the chemical-property reduction process.

The effectiveness of each model during the chemical-property reduction process was tested by comparing the predicted leaching or non-leaching status of each chemical to its known or correct status and assessing the misclassification rate. Predicted leaching status of the chemicals by each CDA model required recalculation of both the MLV for every chemical and the universal MLT. Calculation of the MLV for each chemical and the universal MLT is described in the following sections. The misclassification result from the initial, or full 11-parameter model, was the standard to compare the effectiveness of the CDA models generated from reduced numbers of chemical properties. Here the objective was to identify models with the fewest chemical properties while also minimizing the misclassification rate.

3. Calculation of the Multivariate Leaching Value for Each Chemical

The CDA procedure produces a set of Raw Canonical Coefficients (raw coefficients), each of which is associated with a chemical property. A single canonical score, denoted as the MLV, was calculated for each chemical:

$$MLV_i = \sum_{j=1}^{N_j} C_j \, V_{j,i} \qquad \text{(Eq. 1)}$$

where MLV_i is the single canonical score, C_j is the raw canonical coefficient of the jth predictor variable, $V_{j,i}$ is the value of the jth predictor variable of the ith chemical.

4. Calculation of the Multivariate Leaching Threshold

The distribution of MLVs for the leacher group of chemicals provided the basis for developing a statistic to discriminate between leacher and non-leacher chemicals. This was possible because the categorical variable was encoded in the CDA model such that the leacher chemicals, in general, are expected to have a higher MLV. Unlike the SNV procedure, which utilized five threshold values unique to each of five chemical properties, this revised procedure produced a single threshold value from the MLVs. The threshold values for the SNV procedure were derived from the distributions for the chemical properties of the leacher chemicals. They were derived for each chemical property by selecting either the 10 or 90 percent confidence interval from their distribution of values. Similarly, the threshold value developed in this current study was also derived from the distribution of leacher chemicals. However, a more formal statistic was used to derive this value. A lower prediction limit was derived (Hahn and Meeker, 1991) at the 95 percent confidence level whereby a future chemical (a chemical not included in this study) will be included in the leacher category if its MLV equals or exceeds the lower prediction limit threshold. This value, denoted as the MLT, was calculated (PROC CAPABILITY, SAS, 2008) as the threshold statistic:

$$\mathrm{MLT} = X_{mL} - (t_{(1-\alpha:\, n-1)})(1/M + 1/n)^{1/2}(S_L) \qquad \text{(Eq. 2)}$$

Where:
MLT = the threshold value for membership in the known leacher distribution
X_{mL} = the mean MLV for the known leacher distribution
$t_{(1-\alpha:\, n-1)}$ = the Student's *t* statistic with alpha set at 0.05
M = the number of future predictions set at 1
n = the number of observations in the known leacher distribution
S_L = the standard deviation of MLVs in the known leacher distribution

5. Cross-Validation of the Multivariate Discriminant Models

Two formal cross-validation methods applicable to small datasets were used to evaluate the most promising CDA models. The first method, termed leave-one-out (LOO) (Huberty, 1994), required one of the 42 chemicals used in the study to be removed from the leacher or non-leacher group and the CDA model regenerated using the remaining 41 chemicals. A new MLV was computed for the chemical removed from the regenerated CDA model which represents the independent test chemical for use in cross-validation. A new MLT was derived from the recalculated MLVs for members in the known leacher group. To maintain its full independent status, the test chemical was not included in the recalculation of the MLT. The test chemical was then reclassified as a leacher or non-leacher by the regenerated CDA model with respect to the revised MLT. For each of the most promising CDA models evaluated, the procedure described above was repeated for all 42 leacher and non-leacher chemicals. Robust model validation was related to a low misclassification rate of the test chemicals. The LOO procedure precisely simulated the basic functionality and intended use of the CDA model where a new independent chemical, simulated here as the test chemical, would be identified as a potential leacher or non-leacher.

The second cross-validation method, termed leave-pair-out (LPO), was similar to the LOO procedure but required removal of two chemicals from the combined group of leacher and non-leacher chemicals. The CDA model was then regenerated using the remaining 40 chemicals. A new MLV was computed for each of the two chemicals removed from the regenerated CDA model which represent the independent test chemicals for use in cross-validation. A new MLT was derived from the recalculated MLVs for members in the known leacher group. The test pair were then reclassified as a leacher or non-leacher with respect to the revised MLT. To maintain their full independent status, the test chemicals were not included in recalculation of the MLT. For each candidate CDA model evaluated, the LPO procedure described above was repeated to account for every possible pair of chemicals to represent the test chemicals. Again, robust model validation was related to a low misclassification rate of the test chemicals. LPO has been reported as a less biased validation procedure for quantifying classification error compared to LOO analyses (Montoya Perez, et al., 2019). For this current study, the LPO procedure provided 861 independent test pairs of chemicals for each candidate CDA model evaluated as opposed to just 42 test chemicals for the LOO procedure. The main limitation of the LPO procedure was its multiplicative analytical intensity with 861 regenerated cross-validation models for each candidate CDA model evaluated. Extending this cross-validation methodology to more than two independent test chemicals became computationally infeasible.

SAS programming code that was developed to provide tests for distribution normality, *t*-test statistics and graphics, correlation analyses, CDA multivariate analysis, and LOO and LPO cross-validation are available upon request.

RESULTS

Revision of the Leacher and Non-Leacher Chemicals

Data from well water monitoring studies conducted in California were used for the development of the lists of leacher (Table 1) and non-leacher (Table 2) chemicals. These lists of chemicals provide information on application site, the number of detections or non-detections, (the latter in relation to the number of wells sampled and statewide use of the chemical), and the sampling or reporting agency. Some listed chemicals were not included in this current analysis because the empirical properties data for these compounds were unavailable or incomplete. These chemicals are identified in Tables 1 and 2, accordingly. Of the 42 chemicals used in this analysis, the majority were soil-applied herbicides; however, there were also 11 soil-applied insecticides, four herbicide degradates, two soil-applied legacy fumigants, and one insecticide degradate.

Chemicals detected in California's groundwater from agricultural use are the basis for inclusion in the known leacher group (Table 1). The sheer number of detections for most of the chemicals indicate widespread movement to groundwater. Chemicals with fewer than 50 reported detections were verified by further investigation. Leaching status for each chemical was confirmed upon reports of residues in two or more wells located within adjacent square-mile sections of land. Detection of these residues in wells located in other agricultural areas added further evidence that the chemical moved to groundwater due to nonpoint-source agricultural use.

Chemicals not detected in California groundwater, but with use patterns that enabled opportunity for movement to groundwater, are the basis for inclusion in the known non-leacher group (Table 2). Two chemicals in this group, alachlor and pendimethalin, have isolated detections in groundwater, with one and two detections, respectively, out of 2,994 well water samples analyzed. These detections were considered questionable and not the result of agricultural use (Troiano and Clayton, 2022).

Analysis of well water sampling data resulted in several changes to the original leacher and non-leacher chemicals upon which the SNV procedure was based. The following are changes to the known leacher- and non-leacher-identified chemicals used by Johnson (1991):

- Older chemicals no longer included because of lack of available empirical properties data required for the CDA modeling process due to data exemption at the time of registration, voluntary withdrawal, or registration cancellation of the chemical prior to the deadline to submit the data:
    - Dieldrin, DDT, DDD, DDE, ametryne, chlordane, heptachlor, lindane, naled, silvex, toxaphene
- Chemicals added as known leachers due to detections in subsequent well sampling studies:
    - Parent chemicals: norflurazon, hexazinone, imidacloprid, tebuthiuron

  - Degradation products: aldicarb sulfone, aldicarb sulfoxide, alachlor ESA, metolachlor ESA, metolachlor OXA, deethylatrazine, deisopropylatrazine, diaminochlorotriazine, desmethylnorflurazon, 2,3,5,6-tetrachloroterephthalic acid

- Chemicals changed from leacher to known non-leacher status:
  - Analytical chemistry update that separated parent from degradates — chlorthal-dimethyl (DCPA)
  - Well sampling in California with no detection — alachlor, carbofuran, cyanazine, fonofos, oxamyl

- Chemicals added as known non-leachers due to no detections in subsequent well sampling studies despite ongoing agricultural use:
  - Iprodione, methomyl, napropamide, oryzalin, oxyfluorfen, thiobencarb

Screening of Chemical Properties to Include in the Multivariate Statistical Analysis

Use of the screening steps outlined in the Methodology section resulted in Log_{10} transformation of all seven empirical properties to approximate normal distributions (Table 3). For the quantum- or molecular-derived chemical properties, Shapiro-Wilk tests on the raw values indicated distribution normality for 11 of the 22 properties. For several other molecular chemical properties, a normal distribution was indicated for either the leacher or non-leacher distribution, but not for both. In this case, Q-Q plots for E_{LUMO}, EAU, SurfArea, SurfVolume, MinElPot, AccPolarArea75, PSA, and Polarizability graphically indicated that the raw data did approximate a normal distribution. For Conformers, Q-Q plots graphically indicated that Log_{10} transformation was acceptable in approximating distribution normality of both the leacher and non-leacher chemicals. Appendix III illustrates the comparison of the Q-Q plots for these chemical properties. Shapiro-Wilks tests for MaxElPot indicated that Log_{10} transformation was successful in approximating distribution normality of both the leacher and non-leacher chemicals. AccSurfArea was discarded from the study because neither the raw nor transformed data for this chemical property approximated distribution normality. In summary, only MaxElPot and Conformers were transformed by Log_{10} to satisfy distribution normality. The remaining molecular property values were maintained in their raw state.

Student's *t*-tests followed the normality tests to indicate potential for discrimination between the leacher and non-leacher groups for each chemical property (Table 3). Six of the seven empirical chemical properties and 13 of the 22 remaining molecular chemical properties met the *t*-test relaxed significance level *P*-value of ≤ 0.15. In Table 3, chemical properties that were not retained for further consideration are identified by an asterisk corresponding to the step in which they failed — failing to meet either distribution normality or the *t*-test *P*-value of ≤ 0.15.

The remaining 19 chemical properties were subjected to a correlation analysis to measure potential redundancy in information. Several groups of highly correlated chemical properties were identified where the Pearson correlation coefficient for paired comparisons was ≥ |0.7| (Table 4). One group was comprised of Log_{10}-transformed values for K_{oc}, solubility, and K_{ow}. Log K_{oc} was selected because it had the strongest *t*-test result. A second group indicated high correlation between Log_{10}-transformed values for laboratory soil aerobic half-life and terrestrial field dissipation half-life. Log aerobic half-life was selected because it had a strong *t*-test result and because aerobic half-life studies are conducted under more controlled and uniform conditions, which would theoretically provide more stable data with less potential variation. A third group consisted of strong correlations between various molecular PolarArea and AccPolarArea properties. PolarArea 100 was selected because it had the highest overall correlation with all other members of the correlated group. Lastly, a fourth group indicated correlation between molecular SurfArea, SurfVolume, and Polarizability. SurfArea was selected because it had the strongest *t*-test result within the correlated group. The seven chemical properties of $LogK_h$, Dipole, E_{HOMO}, LogMaxElPot, MinLocIonPot, Hardness, and LogConformers were retained because each had relatively low correlation with all other chemical properties. These resulting 11 chemical properties chosen for testing in the initial CDA analysis are identified by hash signs (#) in the column and row headers in the correlation analyses results (Table 4).

Multivariate CDA Analysis and Model Selection

The initial CDA model incorporating all 11 chemical properties, denoted as the full model, indicated highly significant differences ($P < 0.001$) for all four multivariate measures of statistical significance, indicating a strong discrimination between the distributions of the known leacher and non-leacher chemicals (Table 5A). Graphical comparison of the distributions of MLV scores indicated a distinct separation between the two groups of chemicals (Figure 1). In the full model, the MLV for each chemical was calculated by summing the products of the raw canonical coefficients associated with each of the 11 chemical property values. Comparing each chemical's MLV against the MLT revealed that only methomyl and 1,3-dichloropropene (1,3-D), members of the non-leacher group, were misclassified as predicted leacher chemicals. As described earlier, the MLT was calculated from the distribution of MLVs from the leacher group of chemicals as the estimated lower prediction limit at the 95 percent confidence level for including an additional chemical into its distribution. From a groundwater contamination perspective, misclassification of a non-leacher as a leacher has fewer adverse consequences than misclassification of a leacher as a non-leacher. Therefore, based on misclassification of only two non-leacher chemicals as leacher chemicals, the full model is judicious and practically robust.

New CDA models were developed using reduced numbers of chemical properties. This procedure was used to identify if a simpler, reduced-parameter model was as effective as the full 11-parameter model in discriminating between the known leacher and non-leacher chemicals. In the full model, the standardized coefficient associated with LogConformers was smaller than those associated with the other chemical parameters in terms of absolute value, indicating it was the least influential in the CDA analysis for discriminating between the leacher and non-leacher chemicals (Table 6). Therefore, this property was excluded in the next reduced-parameter CDA model. Leacher and non-leacher classification of chemicals for the reduced-parameter model with ten chemical properties was the same as the full model, again with only methomyl and 1,3-D misclassified as predicted leachers. Further sequential reduction of chemical properties from the CDA analysis resulted in continued misclassification of only methomyl and 1,3-D. This was observed from the full model, identified as CDA model No. 1, through the four-parameter CDA model, identified as CDA model No. 8 (Table 7). The three-parameter model (CDA model No. 9) resulted in the additional misclassification of the chemical napropamide (Table 7).

Candidate models chosen from Table 7 for cross-validation analysis using the LOO and LPO procedures were CDA models No. 7 (two misclassifications), No. 8 (two misclassifications), and No. 9 (three misclassifications). These selected models featured the fewest chemical properties coinciding with either no additional or only one additional misclassification compared to the full model. Expanded results for each of the CDA models and for the reduced-parameter selection procedure are reported in Appendix IV.

Model Cross-Validation

Cross-validation using the LOO methodology required sequential removal of every chemical from each of the three CDA candidate models (No. 7, 8, and 9; Table 7). Forty-two models were regenerated for each candidate model to allow for every leacher and non-leacher chemical to represent the independent test chemical. The test chemical for each regenerated model was reclassified as a leacher or non-leacher based on its updated MLV and the revised MLT.

The misclassification rate for candidate model No. 7 was 4.8% because two chemicals (methomyl and 1,3-D) were misclassified as leacher chemicals out of a total of 42 leacher and non-leacher chemicals used in the study. The LOO analysis of candidate model No. 7 also resulted in a 4.8% misclassification rate because every independent test chemical was correctly classified except for when methomyl and 1,3-D were the test chemicals. Interestingly, when bentazon or EDB was the independent test chemical there was greater separation between the leacher and non-leacher chemicals resulting in 1,3-D being correctly classified and only methomyl misclassified.

The misclassification rate for candidate model No. 8 was 4.8%, again because only methomyl and 1,3-D were misclassified as leacher chemicals. The LOO analysis of candidate model No. 8 resulted in a 9.5% misclassification rate because four chemicals (methomyl, 1,3-D, napropamide, and norflurazon) were misclassified when they were the independent test chemicals. The remaining 38 test chemicals in the LOO analysis of candidate model No. 8 were correctly classified.

Finally, the misclassification rate for candidate model No. 9 was 7.1% because three chemicals (methomyl, 1,3-D, and napropamide) were misclassified out of the total of 42 leacher and non-leacher chemicals used in the study. The LOO analysis of candidate model No. 9 resulted in a 14.3% misclassification rate of the independent test chemicals because six chemicals (methomyl, 1,3-D, napropamide, bentazon, norflurazon, and carbofuran) were misclassified when they were the test chemicals.

Misclassification of norflurazon in the LOO analysis of candidate model No. 8, and norflurazon and bentazon in the LOO analysis of candidate model No. 9 as non-leachers is noteworthy because these two chemicals are known leachers and have frequently been detected in California groundwater. Candidate model No. 9 was excluded from LPO cross-validation analysis based on its shift to a considerably greater misclassification of chemicals in the LOO analysis.

LPO cross-validation analysis was then conducted whereby all 42 leacher and non-leacher chemicals used in the study were arranged into every possible paired combination to produce 861 models regenerated for each candidate model (No. 7 and 8; Table 7). The independent test pair removed from each regenerated model was evaluated for leacher or non-leacher reclassification status. As stated above, candidate models No. 7 and 8 had a 4.8% misclassification rate because methomyl and 1,3-D were misclassified as leacher chemicals.

The LPO analysis of candidate CDA model No. 7 resulted in a 10.1% misclassification rate of the independent test chemicals. Most of these misclassifications were of methomyl (41 misclassifications) and 1,3-D (40 misclassifications) when they were paired with the remaining test chemicals, except for 1,3-D which was correctly classified when paired with EDB. Bentazon, EDB, and oxyfluorfen were each misclassified once when paired with another test chemical. Norflurazon was misclassified when paired with three other test chemicals.

The LPO analysis of candidate CDA model No. 8 resulted in a 17.8% misclassification rate of the independent test chemicals. Most of these misclassifications were methomyl, 1,3-D, and norflurazon, each with 41 misclassifications. In addition, napropamide had 29 misclassifications and bentazon had one misclassification when they were paired with various other independent test chemicals.

Although candidate models No. 7 and No. 8 only had two misclassifications (methomyl and 1,3-D), the LPO cross-validation analysis revealed a disproportionately higher misclassification rate for candidate model No. 8. The many misclassifications of the known leacher norflurazon as a non-leacher when cross validating candidate model No. 8 was particularly concerning when compared to model No. 7.

CDA candidate models No. 7 and 8 with five and four model input parameters, respectively, had only two misclassifications out of the 42 chemicals used to develop each model. Candidate model No. 9 with only three parameters had three misclassifications out of the 42 chemicals used to develop the model. However, the cross-validation results from the LOO and LPO analyses conclusively indicated that CDA candidate model No. 7 with the five model input parameters exhibited substantially greater consistency during reclassification of the independent test chemicals. Thus, candidate model No. 7 was identified as considerably more stable and reliable compared to the other two candidate models tested.

Multivariate Model to Determine Potential for Leaching

The CDA model generated from five chemical properties of dipole, LogMaxElPot, E_{HOMO}, $LogK_{oc}$, and LogAero was as effective as the full 11-parameter model in discriminating between the known leacher and non-leacher chemicals (CDA model No. 7, Table 7). This model produced a highly significant statistical difference in MLVs (raw canonical scores) between the known leacher and non-leacher distributions of chemicals as indicated by statistics (Table 5B) and the large graphical separation of the two distributions (Figure 2). LOO and LPO cross-validation analyses of this model revealed a low misclassification rate of the independent test chemicals for their leacher or non-leacher status. This contrasted with the models having four or three chemical properties (CDA models No. 8 and 9, Table 7) where a higher misclassification rate was indicated for the independent test chemicals, including misclassifications of known leachers as non-leachers. Equation 3 presents the raw canonical coefficients associated with the five chemical-specific properties to produce an MLV score used for evaluating a chemical for leaching potential:

$$\text{MLV} = 1.9349\ (\text{Log}_{10}\ \text{soil aerobic half-Life}) - 1.2197\ (\text{Log}_{10}\ K_{OC}) + 0.1562\ (\text{dipole moment}) - 0.9140\ (E_{HOMO}) + 2.2728\ (\text{Log}_{10}\ \text{MaxElPot}) \qquad \text{(Eq. 3)}$$

The chemical-specific MLV is then compared to the derived MLT value of 14.4706. An MLV score equal to or greater than the MLT value classifies the chemical as a predicted leacher.

DISCUSSION

The revised multivariate approach to identify a chemical's potential for movement to groundwater retained some of the methodology used in the original SNV procedure developed to comply with the PCPA (Wilkerson and Kim, 1986; Johnson, 1991). Specifically, both approaches identified potential leachers using statistics that compared the distributions of chemical-specific properties or property-derived scores between chemicals recognized as known leachers or non-leachers. In the presence of statistical differences between the distributions, membership of a chemical into the distribution of leachers was determined using a statistical interval. However, this revised approach includes the following improvements:

- Revision of the known leacher and non-leacher chemicals to incorporate a more rigorous body of well water monitoring data;
- Inclusion of additional chemical-specific properties to develop a model that distinguishes between the known leacher and non-leacher chemicals;
- Use of a multivariate analysis to develop an effective model that discriminates membership between the known leacher and non-leacher chemicals;
- Development of a single test score to identify a chemical's leaching potential whereby a chemical's MLV score is compared to the MLT value;
- Utilization of a more formal statistic for establishing a threshold value for discriminating between known leacher and non-leacher chemicals, namely a prediction limit at the 95 percent confidence level for the addition of a new chemical into the leacher distribution; and
- Inclusion of two cross-validation procedures for testing the consistency and stability of the multivariate model for identifying leacher and non-leacher chemicals.

With respect to development of the known leacher and non-leacher lists, only well water sampling data from studies conducted in California were evaluated. Compared to the original approach that was based on nationwide well water sampling data, this restriction provided greater confidence in the reporting of detections and non-detections because sampling was conducted in areas of known high pesticide use and known vulnerability to pesticide movement to groundwater. Use of these data resulted in a substantial difference from the listing of chemicals categorized as known leachers and non-leachers by Johnson (1991). For example, some chemicals previously identified as leachers were reclassified as non-leachers. Most of these changes were based on the more recent targeted well sampling studies conducted by GWPP, but updated chemical analytical methodology also accounted for some changes.

The multivariate statistical approach used to identify an effective model indicated that a model containing five chemical-specific properties was extremely effective in discriminating between the revised known leacher and non-leacher chemicals. The properties included in this model were empirically-derived values for Log_{10} soil aerobic half-life and Log_{10} K_{oc}, and the calculated molecular parameters dipole moment, E_{HOMO}, and Log_{10} MaxElPot. These variables relate to the current theoretical basis and general understanding of the processes responsible for chemical movement to groundwater. Soil aerobic half-life and E_{HOMO} relate to persistence of a chemical in the soil environment. E_{HOMO} directly relates to the ionization potential (i.e., the energy necessary to remove an electron from the neutral atom) and is a measure of the susceptibility of the molecule to chemical breakdown (Karelson et al., 1996). K_{oc}, dipole, and MaxElPot characterize potential mobility of a chemical in the soil environment and the latter two indicate the polarity and hydrogen bonding capacity of a molecule, respectively — both properties strongly linked to water solubility. With respect to mobility of a chemical, previous studies have established dipole and MaxElPot parameters as explanatory variables for empirical measures of water solubility, soil sorption, and K_{ow} (Mamy, 2015; Rathi, et al., 2020; Weng, et al., 2002; Doucette, 2003). The inclusion of chemical persistence and mobility properties in the multivariate CDA model reflects the specified grouping of chemical properties related to pesticide movement to groundwater as specified in the original PCPA legislation. Appendix V presents an in-depth discussion of how dipole, E_{HOMO}, and MaxElPot values reflect potential persistence and mobility of a chemical in soil.

The enhanced performance of the proposed CDA model with five chemical properties is attributable to its multivariate-based development, a more robust list of known leacher- and non-leacher-classified chemicals, inclusion of additional chemical properties data, and stability in the measurement and calculation of these property values. With respect to the last attribute, soil aerobic half-life and K_{oc} are determined from studies conducted in pre-conditioned soil under controlled laboratory conditions resulting in more measurement stability compared to other half-life data generated in native soil under actual field conditions, such as from terrestrial field dissipation studies. Nonetheless, since uncertainty in chemical-specific values for soil aerobic half-life and K_{oc} exists, median values for these properties were used for the CDA model-development set of chemicals. Propagation of these uncertainties with respect to evaluation of new chemicals by the CDA model would likely increase their potential for misclassification. Dipole, MaxElPot, and E_{HOMO} are calculated from 3-D conformational data and chemical-specific regeneration of their values is invariant under the same quantum modeling construct.

The misclassification rate of the proposed five-chemical-property model was approximately 5% where the leaching status of only two of the 42 chemicals were misclassified. In contrast, the SNV procedure by Johnson (1991) utilizing a univariate statistical approach had a reported chemical misclassification rate of 36%. However, contributing to this relatively high misclassification rate was the misidentification of several known leacher and non-leacher chemicals used for model development. With respect to the SNV process, the multivariate-based CDA model developed in this study provides a more accurate estimate of an agricultural chemical's potential to leach to California's groundwater. Furthermore, cross-validation of the selected CDA model using the LOO and LPO analyses indicated no significant loss in model performance or stability when predicting the leaching status of a chemical or pair of chemicals that were independent of the chemicals used for model development. This finding was particularly compelling because these analyses simulated the precise functionality of the proposed model as mandated by the PCPA as it applies to evaluating the leaching potential of new, independent chemical active ingredients or degradates.

CONCLUSIONS

With improved discrimination between the revised known leacher and non-leacher chemicals, and inherent stability indicated in the LOO and LPO cross-validation procedures, the five-parameter multivariate model developed in this study provides improved capability in identifying potential groundwater leachers compared to the SNV process. The new model requires data for soil aerobic half-life, K_{oc}, and three molecular properties representing the values for dipole, E_{HOMO}, and MaxElPot.

The following procedure is recommended for identifying leaching potential of a specific chemical:

1. In the presence of multiple values, utilize the median value for the soil aerobic half-life (days) and K_{oc} (cm^3/g) of the chemical.

2. Calculate the molecular dipole moment (Debye), E_{HOMO} (eV), and MaxElPot (kJ/mol) values from the chemical structure while in its equilibrium state using the Spartan '20 software utilized in this study. Appendix VI provides guidance for generating the molecular variables values for use in the proposed multivariate model.

3. Derive the MLV for the chemical as:

$$\text{MLV} = 1.9349\,(\text{Log}_{10}\text{ soil aerobic half-life}) - 1.2197\,(\text{Log}_{10} K_{oc}) + 0.1562\,(\text{dipole moment}) - 0.9140\,(E_{HOMO}) + 2.2728\,(\text{Log}_{10}\text{ MaxElPot})$$

4. Compare the MLV for the chemical to the MLT value of 14.4706. The MLT value was derived from the distribution of MLV scores for the group of known leacher chemicals. If the MLV ≥ MLT then the chemical is identified as a member of the leacher group of chemicals and has the potential to move to California's groundwater.

REFERENCES

Ando, C. 1992. Survey for chlorthal-dimethyl residues in well water of seven California counties. Environmental Monitoring Branch, Department of Pesticide Regulation, Sacramento, California. EH 92-01.

Cohen, S.Z., S.M. Creeger, R.F. Carsel, and C.G. Enfield. 1984. Potential for pesticide contamination of groundwater resulting from agricultural use. In (R.F. Krueger and J.N. Seiber, eds): Treatment and Disposal of Pesticide Wastes. American Chemical Society Symposium Series 269. Pp 297-325.

DaSilva, A. 2018. Protocol for Groundwater Protection List monitoring for 2,4-dichlorophenoxyacetic acid herbicide. Environmental Monitoring Branch, Department of Pesticide Regulation, Sacramento, California. Available at: https://www.cdpr.ca.gov/docs/emon/pubs/protocol.htm?filter=grndwater (verified 24 July 2024).

DPR. 2020. Well Inventory Database. California Department of Pesticide Regulation, Sacramento, California. Available at: https://www.cdpr.ca.gov/docs/emon/grndwtr/well_inventory_database/index.htm (verified 24 July 2024).

Doucette, W.J. 2003. Quantitative structure-activity relationships for predicting soil-sediment sorption coefficients for organic chemicals. Environmental Toxicology and Chemistry, 22(8): 1771-1788.

Fram, M.S., and S.V. Stork. 2019. Determination of study reporting limits for pesticide constituent data for the California Groundwater Ambient Monitoring and Assessment Program Priority Basin Project, 2004–2018—Part 1: National Water Quality Schedules 2003, 2032, or 2033, and 2060: U.S. Geological Survey Scientific Investigations Report 2019–5107, 129 p., https://doi.org/10.3133/sir20195107 (verified 24 July 2024).

Gustafson, D.I. 1989. Groundwater Ubiquity Score: A simple method for assessing pesticide leachability. Env. Tox. Chem. Vol 8, pp 339-357, 1989.

Hahn, J.G., and W.Q. Meeker. 1991. Statistical Intervals: A Guide for Practitioners. John Wiley and Sons. 392 pp.

Huberty, C.J. 1994. Applied Discriminant Analysis. John Wiley & Sons, Inc. N.Y. 466 pages.

Johnson, B. 1988. Setting revised specific numerical values. November 1988. Environmental Monitoring Branch, Department of Pesticide Regulation, Sacramento, California. EH 88-12.

Johnson, B. 1989. Setting revised specific numerical values. October 1989. Environmental Monitoring Branch, Department of Pesticide Regulation, Sacramento, California. EH 89-13.

Johnson, B. 1991. Setting revised specific numerical values. April 1991. Environmental Monitoring Branch, Department of Pesticide Regulation, Sacramento, California. EH 91-06.

Karelson M., V.S. Lobanov, A.R. Katritzky. 1996. Quantum-chemical descriptors in QSAR/QSPR studies. Chem Rev. 1996 May 9; 96(3):1027-1044.

Medalie, L., M.W. Sandstrom, P.L. Toccalino, W.T. Foreman, R.C. ReVello, L.M. Bexfield, and M.L. Riskin. 2019. Use of set blanks in reporting pesticide results at the U.S. Geological Survey National Water Quality Laboratory, 2001–15: U.S. Geological Survey Scientific Investigations Report 2019-5055. Available at: https://doi.org/10.3133/sir20195055 (verified 24 July 2024).

Mamy, L. 2015. Prediction of the fate of organic compounds in the environment form their molecular properties: A review. Critical Reviews in Environmental Science and Technology, 45:1277-1377, 2015. Available at: https://www.tandfonline.com/doi/full/10.1080/10643389.2014.955627 (verified 24 July 2024).

Monohan, K., I.J. Tinsley, S.F. Shepherd, and J.A. Field. 1995. Quantitative determination of the acidic metabolites of Dacthal in groundwater by strong anion exchange solid-phase extraction. J. Agicult. Food Chem. 1995 34:2418-2423.

Montoya Perez, I., A. Airola, P.J. Boström, and T. Pahikkala. 2019. Tournament leave-pair-out cross-validation for receiver operating characteristic analysis. *Statistical methods in medical research* 28.10-11 (2019): 2975-2991. Available at: https://doi.org/10.1177/0962280218795190 (verified 24 July 2024).

Nordmark, C. 2001. Mini-Memo: GW01 and GW01 overview and results. Environmental Monitoring Branch, Department of Pesticide Regulation, Sacramento, California. 12/26/2001.

Randic, M.J. 1975. On characterization of molecular branching. J. Am. Chem. Soc., 97(1975), pp 3086-6615.

Rathi, P.C., R.F. Ludlow, and M.L. Verdonk. 2020. Practical high-quality electrostatic potential surfaces for drug discovery using a graph-convolutional deep neural network., J. Med. Chem. 2020, 63, 16, 8778–8790.

Ruud, N. 2021. Study 300: Groundwater monitoring for DCPA and its degradation products MTP and TPA. Environmental Monitoring Branch, Department of Pesticide Regulation, Sacramento, California. Available at: https://www.cdpr.ca.gov/docs/emon/pubs/ehapreps/study_300_2021.pdf (verified 24 July 2024).

SAS Institute Inc. 2008. SAS/STAT(R) User's Guide 9.2, Second Edition. SAS Institute Inc., Cary, N.C. Available at: http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_chap0_sect004.htm (verified 24 July 2024).

Troiano, J. and M. Clayton. 2022. Verification of pesticide residue detections in wells reported by U.S. Geological Survey Groundwater Ambient Monitoring Assessment Well Sampling Program for data collected between May 2004 and June 2018 using pesticide use data and landscape analysis. Environmental Monitoring Branch, Department of Pesticide Regulation, Sacramento, California.

Available at: https://www.cdpr.ca.gov/docs/emon/pubs/ehapreps/analysis_memos/gama_detection_analysis_2004-2018.pdf (verified 24 July 2024).

U.S. EPA. 1992. Another Look: National Survey of Pesticides in Drinking Water Wells. Phase 2 Report. USEPA, Washington D.C., USA. 570/9-91-020, January 1992.

U.S. EPA. 2008. Chapter 4: DCPA mono- and diacid degradates. A chapter from: Regulatory Determination Support Document for Selected Contaminants from the Second Drinking Water Contaminant Candidate List (CCL 2). EPA Report 815-R-08-012. Available at: https://www.epa.gov/sites/production/files/2014-09/documents/chapter_4_dcpa_mono-_and_di-acid_degradates.pdf (verified 24 July 2024).

Weng, L., E. Temminghoff and W. Van. 2002. Interpretation of humic acid coagulation and soluble soil organic matter using a calculated electrostatic potential. European Journal of Soil Science. 2002, 53, 1365-2389.

Wilkerson, M.R. and K.D. Kim. 1986. The Pesticide Contamination Prevention Act: Setting specific numerical values. Environmental Monitoring Branch, Department of Pesticide Regulation, Sacramento, California. EH 86-02.

Worrall. F. 2001. A molecular topology approach to predicting pesticide pollution of groundwater. Environ. Sci. Technol., 35 (11): 2282–2287. Available at: https://doi.org/10.1021/es001593g (verified 24 July 2024).

Worrall, F. and M. Thomsen. 2004. Quantum vs topological descriptors in the development of molecular models of groundwater pollution by pesticides. Chemosphere, 54(4): 585-596. Available at: https://doi.org/10.1016/S0045-6535(03)00705-7 (verified 24 July 2024).

Zhao, G. and A.L. Maclean. A comparison of canonical discriminant analysis and principal component analysis for spectral transformation. Photogrammetric Engineering & Remote Sensing, 66(7): 841-847.

TABLES

Table 1. Pesticide chemicals and degradation products (Deg) with verified detections in California groundwater prior to 2019 resulting from agricultural applications. Chemicals with insufficient physical/chemical properties data were excluded from the analysis.

Chemical	**Application Site of Chemical or Parent Chemical**	**Wells with Detections**	**Reporting Agency**	**Included in Analysis**
1,2-Dichloropropane (1,2-D)	Soil-applied fumigant	303	DPR, USGS, SWRCB	Yes
1,2,3-Trichloropropane	Impurity in soil-applied fumigant	379	SWRCB	No
2,3,5,6-Tetrachloroterephthalic acid (TPA) (Deg)	Soil-applied herbicide	35	DPR	Yes
3,4-Dichloroaniline (Deg)	Soil-applied herbicide	106	USGS	No
Alachlor ethanesulfonic acid (ESA) (Deg)	Soil-applied herbicide	34	DPR	No
Aldicarb sulfone (Deg)	Soil-applied insecticide	61	DPR, RWQB	Yes
Aldicarb sulfoxide (Deg)	Soil-applied insecticide	25	DPR, RWQB	No
Atrazine	Soil-applied herbicide	684	DPR, USGS, SWRCB	Yes
Bentazon	Rice herbicide	153	DPR, USGS, SWRCB	Yes
Bromacil	Soil-applied herbicide	323	DPR, USGS	Yes
Deethylatrazine (Deg)	Soil-applied herbicide	572	DPR, USGS	Yes
Deisopropylatrazine (Deg)	Soil-applied herbicide	549	DPR, USGS, Registrant	Yes
Diaminochlorotriazine (Deg)	Soil-applied herbicide	373	DPR, USGS, Registrant	No
Dibromochloropropane (DBCP)	Soil-applied fumigant	3399	DPR, USGS, SWRCB, Fresno, Kern	Yes
Diuron	Soil-applied herbicide	594	DPR, USGS, SWRCB	Yes
Ethylene dibromide (EDB)	Soil-applied fumigant	204	DPR, USGS, SWRCB, Kern	Yes
Hexazinone	Soil-applied herbicide	82	DPR, USGS	Yes
Imidacloprid	Soil-applied insecticide	13	DPR, USGS	Yes
Metolachlor ESA (Deg)	Soil-applied herbicide	81	DPR, USGS	Yes
Metolachlor oxanilic acid (OXA) (Deg)	Soil-applied herbicide	27	DPR, USGS	No
Norflurazon	Soil-applied herbicide	101	DPR, USGS	Yes
Desmethylnorflurazon (Deg)	Soil-applied herbicide	131	DPR, USGS	No
Prometon	Soil-applied herbicide	145	DPR, USGS	Yes
Simazine	Soil-applied herbicide	1328	DPR, USGS, Registrant	Yes
Tebuthiuron	Soil-applied herbicide	53	DPR, USGS	Yes

Table 2. Pesticide chemicals and degradation products (Deg) not detected in California groundwater prior to 2019 following agricultural applications. Chemicals lacking physical/chemical properties data were excluded from the analysis. Minimum detection limit (MDL) for groundwater analysis was 0.05 µg/L except for 1,3-dicloropropene and methyl bromide where the MDL was 0.1 µg/L.

Chemical	**Application Site of Chemical or Parent Chemical of Degradate**	**Statewide Application of Chemical or Parent Chemical of Degradate 1990-2005 (millions lbs)**	**Detections / Number of Wells Sampled**		**Included in Analysis**
			DPR	**USGS**	
1,3-Dicloropropene	Soil-applied fumigant	56.0	0/54	0/971	Yes
2,6-Diethlaniline (Deg)	Soil-applied herbicide	0.7	0/0	0/2994	No
3,5-Dichloroaniline (Deg)	Soil-applied fungicide	6.8	0/0	0/1000	No
Alachlor	Soil-applied herbicide	0.7	0/259	1/2994	Yes
Carbaryl	Soil-applied insecticide	9.2	0/123	0/2994	Yes
Carbofuran	Soil-applied insecticide	3.1	0/119	0/1241	Yes
Chlorthal-dimethyl (DCPA)	Soil-applied herbicide	6.7	0/279	0/2994	Yes
Cyanazine	Soil-applied herbicide	4.3	0/823	0/1000	Yes
Dimethoate	Soil-applied insecticide	8.6	0/120	0/2994	Yes
Disulfoton	Soil-applied insecticide	1.7	0/83	0/964	Yes
Disulfoton sulfone (Deg)	Soil-applied insecticide	1.7	0/0	0/1000	No
Ethoprop	Soil-applied insecticide	0.6	0/86	0/1000	Yes
Fenamiphos	Soil-applied insecticide	2.1	0/166	0/2959	Yes
Fenamiphos sulfone (Deg)	Soil-applied insecticide	2.1	0/61	0/2993	No
Fenamiphos sulfoxide (Deg)	Soil-applied insecticide	2.1	0/67	0/2341	No
Fonofos	Soil-applied insecticide	0.6	0/111	0/2994	Yes
Iprodione	Soil-applied fungicide	6.8	0/124	0/2992	Yes
Linuron	Soil-applied herbicide	1.4	0/244	0/840	Yes
Methomyl	Soil-applied insecticide	9.6	0/142	0/810	Yes
Methyl bromide	Soil-applied fumigant	215	0/9	0/931	Yes
Napropamide	Soil-applied herbicide	2.2	0/231	0/77	Yes
Oryzalin	Soil-applied herbicide	9.6	0/275	0/840	Yes
Oxamyl	Soil-applied insecticide	1.6	0/3	0/804	Yes
Oxyfluorfen	Soil-applied herbicide	7.2	0/2	0/1000	Yes
Pendimethalin	Soil-applied herbicide	6.5	0/0	2/2994	Yes
Phorate	Soil-applied insecticide	1.8	0/77	0/2994	Yes
Phorate oxygen analog (Deg)	Soil-applied insecticide	1.8	0/0	0/2958	No
Propyzamide	Soil-applied herbicide	1.9	0/162	0/2993	Yes
Thiobencarb	Soil-applied herbicide	8.7	0/273	0/1241	Yes
Trifluralin	Soil-applied herbicide	19.8	0/9	0/2994	Yes

Table 3. Shapiro-Wilk (W) testing for distribution normality of chemical properties within the leacher and non-leacher groups. Logarithm Base-10-transformed data were selected when indicating distribution normality in contrast to raw data. *T*-tests indicated differences between leacher and non-leacher chemicals for the means of each chemical property. Pesticide properties considered for inclusion in the canonical discriminant analysis (CDA) were dependent on distribution normality, generally when W > 0.05, and *t*-test probability at *P* ≤ 0.15. An asterisk (*) indicates property excluded from the study; placement of the * indicates the analysis step it failed.

Pesticide Property	**Raw Data**		**Log10 Transformed**		***T*-test (*t* ≤ *0.15*)**	**Considered Further for CDA Modeling**
	Leachers (W > *0.05*)	**Non-Leachers (W > *0.05*)**	**Leachers (W > *0.05*)**	**Non-Leachers (W > *0.05*)**		
Empirical Properties						
Solubility	<0.001	<0.001	0.78	0.61	0.02	Yes
K_{oc}	<0.001	<0.001	0.34	0.84	0.002	Yes
Aerobic	<0.001	<0.001	0.68	0.88	<0.001	Yes
TFD	0.001	<0.001	0.57	0.83	<0.001	Yes
K_{ow}	<0.001	<0.001	0.08	0.48	0.005	Yes
VP*	<0.001	<0.001	0.04	0.11	0.96*	No
K_{h}	<0.001	<0.001	0.1	0.05	0.08	Yes
Molecular Properties						
MolecularWt*	0.77	0.17			0.25*	No
E_{HOMO}	0.07	0.77			0.06	Yes
E_{LUMO}*	0.48	0.02			0.65*	No
Electronegativity*	0.66	0.22			0.44*	No
Hardness	0.27	0.13			0.13	Yes
EAU*	<0.001	0.27	0.002	0.66	0.76*	No
Dipole	0.19	0.84			0.05	Yes
Conformers	<0.001	<0.001	<0.001	0.64	0.03	Yes
SurfArea	0.10	0.006			0.05	Yes
SurfVolume	0.27	0.004			0.07	Yes
AccSurfArea*	0.02*	0.0009*	0.0024*	<0.001*		No
MinElPot*	0.009	0.82	<0.001	0.03	0.95*	No
MinLocIonPot	0.09	0.19			0.09	Yes
PSA*	0.10	0.007			0.62*	No
PolarArea75*	0.83	0.62			0.32*	No
AccPolarArea75*	0.49	0.07			0.34*	No
PolarArea100	0.08	0.42			0.06	Yes
AccPolarArea100	0.09	0.21			0.06	Yes
PolarArea125	0.14	0.70			0.04	Yes
AccPolarArea125	0.18	0.64			0.06	Yes
MaxElPot	0.53	0.03	0.24	0.07	0.04	Yes
Polarizability	0.25	0.004	0.15	<0.001	0.06	Yes

Table 4. Correlation coefficients between chemical properties for the combined set of leacher and non-leacher pesticides. The selection of properties was determined by criteria in Table 3. Asterisks correspond to correlation coefficients as |0.7| ≤ * < |0.8|, |0.8| ≤ ** < |0.9|, and |0.9| ≤ *** ≤ |1.0|. Hash sign (#) identifies chemical properties retained for canonical discriminant analysis model.

Pearson Correlation Coefficient, N=42
Prob > |r| under H0

	LogSolub	LogK_{oc}#	LogK_{ow}	LogAero#	LogTFD	LogKh#	PolarArea100#	AccPolarArea100	PolarArea125
LogSolub		**-0.872 <0.001	**--0.819 <0.001	-0.094 0.555	-0.012 0.942	-0.291 0.062	-0.025 0.875	-0.024 0.990	0.215 0.171
LogK_{oc}#			**-0.802 <0.001	0.097 0.542	-0.056 0.726	0.329 0.033	-0.048 0.762	-0.026 0.868	-0.270 0.084
LogK_{ow}				0.070 0.659	0.018 0.912	0.475 0.002	-0.314 0.043	-0.295 0.058	-0.524 <0.001
LogAero#					*0.760 <0.001	-0.235 0.134	0.302 0.052	0.279 0.074	0.275 0.078
LogTFD						-0.194 0.219	0.183 0.246	0.182 0.247	0.139 0.380
LogK_h#							-0.680 <0.001	-0.642 <0.001	*-0.749 <0.001
PolarArea 100#								***0.979 <0.001	***0.916 <0.001
AccPolar Area100									**-0.875 <0.001
AccPolar Area125	0.212 0.177	-0.240 0.125	-0.506 <0.001	0.242 0.123	0.116 0.466	*-0.731 <0.001	***0.906 <0.001	***0.901 <0.001	***0.981 <0.001
Dipole#	-0.002 0.988	-0.023 0.886	-0.340 0.028	0.290 0.062	0.041 0.795	-0.509 <0.001	0.670 <0.001	0.603 <0.001	0.646 <0.001
E_{HOMO}	-0.272 0.081	0.232 0.140	0.118 0.458	-0.027 0.865	-0.271 0.083	-0.284 0.069	0.193 0.221	0.084 0.595	0.215 0.172
LogMax ElPot#	0.097 0.542	-0.215 0.172	-0.335 0.030	0.111 0.486	0.257 0.101	-0.398 0.009	0.563 <0.001	0.539 <0.001	0.553 <0.001
MinLoc IonPot#	-0.192 0.224	0.138 0.385	0.078 0.623	0.313 0.043	0.548 <0.001	-0.094 0.555	0.212 0.177	0.278 0.075	0.124 0.433
Hardness	0.550 <0.001	-0.545 <0.001	-0.325 0.036	-0.158 0.319	-0.048 0.761	0.268 0.086	-0.380 0.013	-0.361 0.019	-0.248 0.114
Polarizability	-0.477 0.001	0.414 0.006	0.290 0.063	0.091 0.569	-0.047 0.768	-0.424 0.005	0.373 0.015	0.343 0.026	0.294 0.059
SurfArea#	-0.479 0.001	0.414 0.006	0.315 0.042	0.066 0.677	-0.061 0.697	-0.403 0.008	0.371 0.016	0.339 0.028	0.294 0.059
Surf Volume	-0.477 0.001	0.401 0.008	0.303 0.051	0.083 0.601	-0.049 0.757	-0.416 0.006	0.364 0.018	0.342 0.027	0.272 0.082
Log Conformers#	-0.145 0.359	0.070 0.659	0.128 0.418	-0.262 0.093	-0.440 0.004	-0.153 0.334	0.020 0.901	0.005 0.974	-0.034 0.832

	Dipole#	E_{HOMO}#	LogMaxElPot#	MinLocIonPot#	Hardness#	Polarizability	SurfArea#	SurfVolume	LogConformers#
AccPolar Area125	0.600 <0.001	0.129 0.417	0.502 <0.001	0.170 0.281	-0.236 0.132	0.304 0.050	0.301 0.053	0.289 0.063	-0.007 0.963
Dipole#		0.348 0.024	0.314 0.043	-0.035 0.826	-0.230 0.143	0.329 0.033	0.294 0.059	0.310 0.046	0.142 0.371
E_{HOMO}#			0.178 0.259	-0.320 0.039	-0.613 <0.001	0.572 <0.001	0.561 <0.001	0.532 <0.001	0.319 0.040
LogMax ElPot#				0.318 0.040	-0.316 0.041	-0.017 0.913	-0.013 0.933	-0.058 0.713	-0.340 0.028
MinLoc IonPot#					-0.284 0.069	0.058 0.716	-0.020 0.901	0.041 0.795	-0.449 0.003
Hardness#						-0.670 <0.001	-0.639 <0.001	-0.656 <0.001	-0.131 0.407
Polarizability							***0.983 <0.001	***0.994 <0.001	0.579 <0.001
SurfArea#								***0.983 <0.001	0.604 <0.001
Surf Volume									0.629 <0.001

Table 5A. Statistical significance of full 11-parameter multivariate canonical discriminant analysis model to discriminate between the leacher and non-leacher distributions of chemicals.

Multivariate Statistics and Exact F Statistics					
S=1 M=4.5 N=14					
Statistic	**Value**	**F Value**	**Num DF**	**Den DF**	**Pr > F**
Wilks' Lambda	0.27474251	7.20	11	30	<0.0001
Pillai's Trace	0.72525749	7.20	11	30	<0.0001
Hotelling-Lawley Trace	2.63977167	7.20	11	30	<0.0001
Roy's Greatest Root	2.63977167	7.20	11	30	<0.0001

Table 5B. Statistical significance of reduced five-parameter multivariate canonical discriminant analysis model to discriminate between the leacher and non-leacher distributions of chemicals.

Multivariate Statistics and Exact F Statistics					
S=1 M=1.5 N=17					
Statistic	**Value**	**F Value**	**Num DF**	**Den DF**	**Pr > F**
Wilks' Lambda	0.28613842	17.96	5	36	<0.0001
Pillai's Trace	0.71386158	17.96	5	36	<0.0001
Hotelling-Lawley Trace	2.49481207	17.96	5	36	<0.0001
Roy's Greatest Root	2.49481207	17.96	5	36	<0.0001

Table 6. Total standardized canonical coefficients for the full 11-parameter canonical discriminant analysis model. Absolute relative magnitude of each coefficient is indicative of the property's effectiveness in discriminating between the leacher and non-leacher chemical groups.

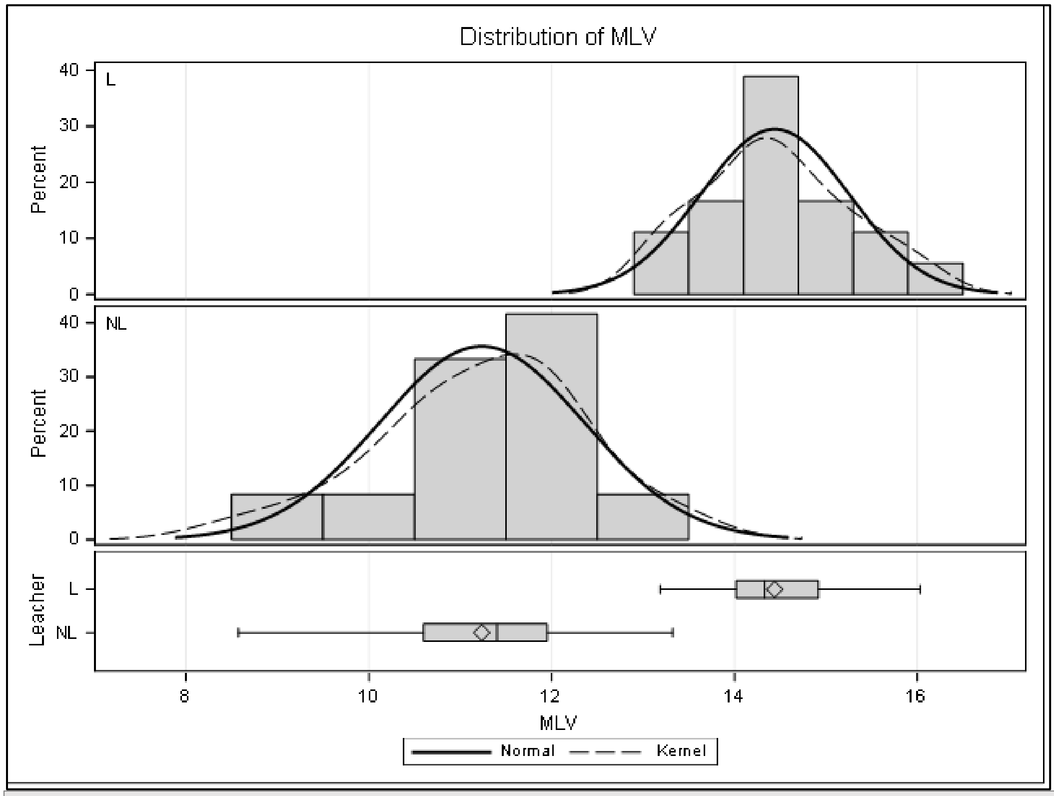
Chemical Properties	**Total Standardized Canonical Coefficient**
$LogK_{oc}$	-0.933
LogAero	1.281
$LogK_{h}$	0.220
PolarArea100	0.064
Dipole	0.425
E_{HOMO}	-0.404
LogMaxElPot	0.378
MinLocIonPot	0.071
Hardness	0.076
SurfArea	-0.166
LogConformers	0.026

Table 7. Misclassification of leacher and non-leacher chemicals resulting from development of the canonical discriminant analysis (CDA) models. Stepwise removal of chemical properties during the model-parameter-reduction process was ordered from the least to most influential property as identified by standardized canonical coefficients. A check mark symbol (✓) indicates inclusion of chemical property in model. A cross mark symbol (✗) indicates exclusion of property from model. An asterisk symbol (*) identifies the most promising models that were evaluated by the leave-one-out and leave-pair-out cross-validation analyses.

	Chemical Properties Included in CDA Model (Empirical and Molecular)												
CDA Model No.	**LogConformers**	**MinLocIonPot**	**PolarArea100**	**SurfArea**	**Hardness**	**LogK$_h$**	**Dipole**	**LogMaxElPot**	**E$_{HOMO}$**	**LogK$_{oc}$**	**LogAero**	**Number of Leacher Group Misclassifications**	**Number of Non-Leacher Group Misclassifications (Chemicals Misclassified)**
1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0	2 (Methomyl; 1,3-D)
2	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0	2 (Methomyl; 1,3-D)
3	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	0	2 (Methomyl; 1,3-D)
4	✗	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	0	2 (Methomyl; 1,3-D)
5	✗	✗	✗	✗	✓	✓	✓	✓	✓	✓	✓	0	2 (Methomyl; 1,3-D)
6	✗	✗	✗	✗	✗	✓	✓	✓	✓	✓	✓	0	2 (Methomyl; 1,3-D)
7*	✗	✗	✗	✗	✗	✗	✓	✓	✓	✓	✓	0	2 (Methomyl; 1,3-D)
8*	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	✓	0	2 (Methomyl; 1,3-D)
9*	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	0	3 (Methomyl; 1,3-D; Napropamide)

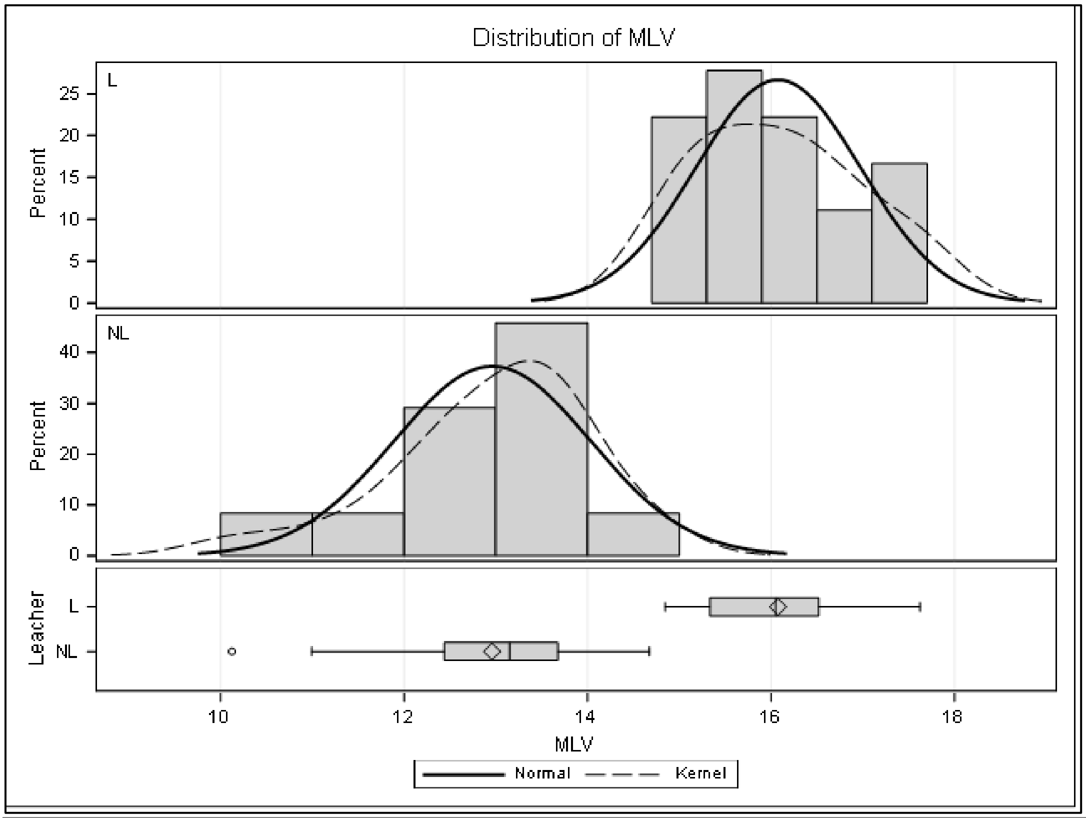
FIGURES

Figure 1. Distributions of Multivariate Leaching Values (MLVs) with associated fits of the normal and kernel distribution for the leacher (L) and non-leacher (NL) chemical groups of the full 11-parameter canonical discriminant analysis model. Box plots indicate the mean, median, interquartile range, and minimum and maximum MLVs of each chemical group. Numerical statistics indicate descriptive and *t*-test results between the groups.



Descriptive Statistics						
Chemical Group	**N**	**Mean**	**Std Dev**	**Std Err**	**Minimum**	**Maximum**
L	18	14.4394	0.8119	0.1914	13.1877	16.0303
NL	24	11.2354	1.1189	0.2284	8.5737	13.3242
Diff (1-2)		3.2040	1.0000	0.3118		
Equality of Variances						
Method	**Num DF**	**Den DF**	**F Value**	**Pr > F**		
Folded F	23	17	1.90	0.1786		
T-Test						
Method	**Variances**	**DF**	**t Value**	**Pr > |t|**		
Pooled	Equal	40	10.28	<0.0001		

Figure 2. Distributions of Multivariate Leaching Values (MLVs) with associated normal and kernel fits for the leacher (L) and non-leacher (NL) chemical groups of the five-parameter canonical discriminant analysis model comprised of Log soil aerobic half-life, LogK_{oc}, dipole, E_{HOMO} and LogMaxElPot. Box plots indicate the mean, median, interquartile range, and minimum and maximum MLVs of each chemical group. Numerical statistics indicate descriptive and *t*-test results between the groups.



Descriptive Statistics						
Chemical Group	**N**	**Mean**	**Std Dev**	**Std Err**	**Minimum**	**Maximum**
L	18	16.0751	0.8977	0.2116	14.8448	17.6204
NL	24	12.9603	1.0693	0.2183	10.1242	14.6740
Diff (1-2)		3.1148	1.0000	0.3118		
Equality of Variances						
Method	**Num DF**	**Den DF**	**F Value**	**Pr > F**		
Folded F	23	17	1.42	0.4639		
T-Test						
Method	**Variances**	**DF**	**t Value**	**Pr > |t|**		
Pooled	Equal	40	9.99	<0.0001		

APPENDICES

Appendix I — Description of Empirical and Quantum-Derived Chemical Properties Investigated

Table I-1. Abbreviation and description of each empirical chemical property.

Parameter	**Abbreviation Used in Report**	**Units[*]**	**Description**
Water Solubility	Solubility	mg/L	Amount of chemical that can dissolve in water at a specific temperature.
Soil Adsorption Coefficient Normalized for Soil Organic Carbon Content	K_{oc}	cm^3/g	Soil partition coefficient for the ratio of amount of chemical sorbed onto soil vs amount in solution adjusted for soil organic carbon content.
Laboratory-derived Soil Aerobic Half-life	Aerobic	days	Laboratory study conducted to determine the dissipation rate in soil.
Field-derived Soil Half-life	TFD	days	Half-life generated from Terrestrial Field Dissipation (TFD) studies. TFD studies are conducted with an end-use formulation (not the technical AI) to determine the extent of pesticide residue dissipation in the field under actual use conditions.
Octanol/Water Partition Coefficient	K_{ow}	none	Laboratory measure of the ratio of a chemical partitioned between octanol and water.
Vapor Pressure	VP	mPa	Pressure exerted by a vapor in thermodynamic equilibrium with the condensed phase either liquid or solid at a given temperature in a closed system.
Henry's Law Constant	K_h	atm m^3/mol	Ratio of the relative abundance of a chemical in the gas vs liquid phase.

[*] Units: atm = standard atmosphere; cm^3 = cubic centimeter; g = gram; L = liter; m^3 = cubic meter; mg = milligram; mol = mole; mPa = megapascal

Table I-2. Abbreviation and description of each quantum-derived chemical property.

Parameter	**Abbreviation Used in Report**	**Units***	**Description**
Molecular Weight	MolWt	amu	The sum of the atomic weights of the atoms contained in a molecule.
Energy of Highest Occupied Molecular Orbital	E_{HOMO}	eV	Energy of the highest occupied molecular orbital.
Energy of Lowest Unoccupied Molecular Orbital	E_{LUMO}	eV	Energy of the lowest unoccupied molecular orbital.
Energy of the Molecule	EAU	au	The energy of a hypothetical reaction that splits a molecule into its isolated nuclei and electrons.
Electronegativity	ElectNeg	eV	A chemical property that describes the tendency of an atom to attract electron density towards itself. Defined in terms of molecular orbital energies, i.e., c = -(EHOMO+ELUMO)/2.
Hardness	Hardness	eV	In density functional theory, the hard/soft principle describes the resistance of the electron cloud (density) of an atom or molecule to deformation in an electric field, e.g. from the charge of an adjacent ion or the partial charge of a dipole. Like electronegativity, defined in terms of molecular orbital energies, i.e., h = -(EHOMO-ELUMO)/2.
Dipole Moment	Dipole	Debye	A measure of the electron cloud distortion (resulting in a separation of positive and negative charge) due to the differences in the electronegativities of atoms in a molecule (designated by an arrow pointing toward the more electronegative side of the molecule).
Number of Conformers	Conformers	unitless	The number of stereoisomers that can be interconverted by rotations about single bonds only, i.e., the number of conformational isomers per molecule.

Parameter	Abbreviation Used in Report	Units*	Description
Polarizability	Polarizability	$10^{-30}m^3$	The ease of distortion of the electron cloud of a molecule by an electric field (e.g., due to the proximity of a charged or partially-charged reagent). The 1st principle method calculates polarizability as a second-order tensor (3x3 matrix), the diagonal elements of which vary in response to an applied electric field. The empirical method is derived from a regression equation containing molecular volume and hardness terms.
Minimum/ Maximum Electrostatic Potential	MinElPot MaxElPot	kJ/mol	An electrostatic potential is defined as the energy sensed by a point positive charge with respect to the nuclei and electrons of a molecule as a function of their location in the molecule. A surface for which the electrostatic potential is negative is indicative of a region of the molecule having stronger nucleophilic character. Conversely, a surface for which the electrostatic potential is positive is indicative of a region of the molecule having stronger electrophilic character. The maxima and minima represent the areas of a molecule having the highest nucleophilic and electrophilic character, respectively.
Minimum Local Ionization Potential	MinLocIonPot	kJ/mol	Reflects the relative ease of electron removal (i.e., ionization) at any location around a molecule. A surface of "low" local ionization potential demarks the areas that are most easily ionized.
Total Polar Surface Area	PSA	Å²	The surface area sum over all polar atoms—primarily oxygen and nitrogen—and their attached hydrogens.
Polar Surface Area	PolarArea75 PolarArea100 PolarArea125	Å²	The area in a molecule occupied by nitrogen and oxygen and any attached hydrogens, derived from the molecular electrostatic potential map. Values were calculated at three levels of electrostatic potential where 75 indicates values calculated at 75 kJ/mol or below -75 kJ/mol, 100 indicates values calculated at 100 kJ/mol or below -100 kJ/mol, and 125 indicates values calculated at 125 kJ/mol or below -125 kJ/mol.
Accessible Polar Surface Area	AccPolarArea75 AccPolarArea100 AccPolarArea125	Å²	Solvent-accessible polar regions on molecular surfaces. As for Polar Surface Area the values were calculated at 75 kJ/mol or below -75 kJ/mol, at 100 kJ/mol or below -100 kJ/mol, and at 125 kJ/mol or below -125 kJ/mol.

Parameter	**Abbreviation Used in Report**	**Units**[*]	**Description**
Surface Area/Surface Volume	SurfArea SurfVolume	Å²	Surface area and volume derived from first principles using electron density surfaces.
Accessible Surface Area	AccSurfArea	Å²	Solvent-accessible regions on molecular surfaces.

[*] Units: Å²= angstrom; amu = atomic mass unit; au = atomic unit of energy; eV = electron-volt; kJ = kilojoule; mol = mole

Appendix II — Values for Empirical and Quantum-Derived Chemical Properties Used in the Analysis

References for the Sources of Empirical Data are Included.

Quantum Data were Calculated using Spartan ’20 Software.

Table II-1. Values for empirical data with associated reference sources (Ref) where group designation L indicates that the chemical was determined to be a leacher and designation NL indicates that the chemical was determined to be a non-leacher. The median value was used where multiple values for a chemical property were available.

Chemical	**Group**	**Water Solubility (mg/L)**	**Ref**	**Soil K_{oc} (cm^3/g)**	**Ref**	**Lab Aerobic Half-life (Days)**	**Ref**	**K_{ow}**	**Ref**	**Vapor Pressure (mPa)**	**Ref**	**K_h (atm m^3/mol)**	**Ref**	**TFD (Days)**	**Ref**
1,2-Dichloropropane (1,2-D)	L	2700	Lewis et al. (2016)	50	Lewis et al. (2016)	140	Lewis et al. (2016)	105	Lewis et al. (2016)	1.94E+06	Lewis et al. (2016)	2.80E-03	Lewis et al. (2016)	700	USDA-ARS DB (2019)
2,3,5,6-Tetrachloroterephthalic acid (TPA)	L	5780	Lewis et al. (2016)	38	Hammer and White (2009)	154	Lewis et al. (2016)	134.9	Hammer and White (2009)	6.52E-03	Hammer and White (2009)	6.58E-13	Hammer and White (2009)	552	DPR (1995)
Aldicarb sulfone	L	10000	Lewis et al. (2016)	10	Lewis et al. (2016)	21	Lewis et al. (2016)	0.269	Lewis et al. (2016)	1.20E+01	Lewis et al. (2016)	2.83E-11	Lewis et al. (2016)	20	Lewis et al. (2016)
Atrazine	L	32.5	Spurlock (2008)	86.5	Spurlock (2008)	146	Spurlock (2008)	449.5	Spurlock (2008)	3.07E-02	Spurlock (2008)	2.00E-09	Spurlock (2008)	85.9	Spurlock (2008)
Bentazon	L	530	DPR (2019)	72.5	DPR (2019)	31	DPR (2019)	0.595	DPR (2019)	1.70E-01	Lewis et al. (2016)	7.11E-10	Lewis et al. (2016)	21	U.S. EPA (1994)
Bromacil	L	700	Spurlock (2008)	14.1	Spurlock (2008)	344	Spurlock (2008)	75.9	Spurlock (2008)	4.13E-02	Spurlock (2008)	1.50E-10	Spurlock (2008)	146	Spurlock (2008)
Deethylatrazine	L	2700	Lewis et al. (2016)	110	Lewis et al. (2016)	170	Lewis et al. (2016)	32.4	Lewis et al. (2016)	1.24E+01	Lewis et al. (2016)	1.53E-09	Lewis et al. (2016)	45	Lewis et al. (2016)
Deisopropylatrazine	L	980	Lewis et al. (2016)	130	Lewis et al. (2016)	102.5	Lewis et al. (2016)	14.1	Lewis et al. (2016)	2.81E+01	Kruger (1992)	9.67E-03	Lewis et al. (2016)	36	Krutz et al. (2010)
Dibromochloropropane (DBCP)	L	1230	Lewis et al. (2016)	100	Lewis et al. (2016)	360	Lewis et al. (2016)	2690	Lewis et al. (2016)	1.00E+04	Lewis et al. (2016)	1.90E-05	Lewis et al. (2016)	203	USDA-ARS DB (2019)
Diuron	L	36.4	Spurlock (2008)	540.2	Spurlock (2008)	372	Spurlock (2008)	697.5	Spurlock (2008)	9.20E-03	Spurlock (2008)	5.10E-10	Spurlock (2008)	114.5	Spurlock (2008)
Ethylene dibromide (EDB)	L	4150	Lewis et al. (2016)	87	Lewis et al. (2016)	70	Lewis et al. (2016)	91.2	Lewis et al. (2016)	1.49E+06	Lewis et al. (2016)	6.47E-04	Lewis et al. (2016)	70	Cohen et al. (1983)
Hexazinone	L	33000	Spurlock (2008)	45.2	Spurlock (2008)	225.5	Spurlock (2008)	15	Spurlock (2008)	2.00E-02	Spurlock (2008)	1.10E-12	Spurlock (2008)	138.5	Spurlock (2008)
Imidacloprid	L	514	Spurlock (2008)	289	Spurlock (2008)	997	Spurlock (2008)	3.7	Spurlock (2008)	2.00E-04	Spurlock (2008)	2.00E-15	Spurlock (2008)	58.9	Spurlock (2008)
Metolachlor ESA	L	212461	Lewis et al. (2016)	9	Lewis et al. (2016)	132	Lewis et al. (2016)	0.0129	Lewis et al. (2016)	2.15E+01	Bayless et al. (2008)	7.21E-16	U.S. EPA (2012)	70	Bayless et al. (2008)
Norflurazon	L	33.7	Spurlock (2008)	460	Spurlock (2008)	130	Spurlock (2008)	280	Spurlock (2008)	3.87E-03	Spurlock (2008)	3.40E-10	Spurlock (2008)	180	Spurlock (2008)
Prometon	L	393.5	Spurlock (2008)	100	Spurlock (2008)	459	Spurlock (2008)	492	Spurlock (2008)	1.03E+00	Spurlock (2008)	3.20E-09	Spurlock (2008)	246.5	Spurlock (2008)
Simazine	L	6.2	Spurlock (2008)	151.7	Spurlock (2008)	110	Spurlock (2008)	122	Spurlock (2008)	2.93E-03	Spurlock (2008)	5.40E-10	Spurlock (2008)	83.5	Spurlock (2008)
Tebuthiuron	L	2600	Spurlock (2008)	79.8	Spurlock (2008)	1220	Spurlock (2008)	63	Spurlock (2008)	3.47E-01	Spurlock (2008)	3.00E-10	Spurlock (2008)	690.5	Spurlock (2008)

Chemical	Group	Water Solubility (mg/L)	Ref	Soil K_{oc} (cm^3/g)	Ref	Lab Aerobic Half-life (Days)	Ref	K_{ow}	Ref	Vapor Pressure (mPa)	Ref	K_h (atm m^3/mol)	Ref	TFD (Days)	Ref
1,3-Dicloropropene	NL	2250	Spurlock (2008)	66	Spurlock (2008)	32.7	Spurlock (2008)	105.5	Spurlock (2008)	3.33E+06	Spurlock (2008)	1.60E-03	Spurlock (2008)	51.6	Spurlock (2008)
Alachlor	NL	240	Spurlock (2008)	127.5	Spurlock (2008)	15	Spurlock (2008)	1220	Spurlock (2008)	1.87E+00	Spurlock (2008)	2.50E-08	Spurlock (2008)	13.2	Spurlock (2008)
Carbaryl	NL	113	Spurlock (2008)	426.4	Spurlock (2008)	5.5	Spurlock (2008)	70.8	Spurlock (2008)	1.60E-01	Spurlock (2008)	2.70E-09	Spurlock (2008)	9.5	Spurlock (2008)
Chlorthal-dimethyl	NL	0.5	Spurlock (2008)	2565	Spurlock (2008)	25.8	Spurlock (2008)	199526	Spurlock (2008)	3.33E-01	Spurlock (2008)	2.20E-06	Spurlock (2008)	21.4	Spurlock (2008)
Carbofuran	NL	351	Spurlock (2008)	25.7	Spurlock (2008)	22.1	Spurlock (2008)	45.7	Spurlock (2008)	3.20E-02	Spurlock (2008)	5.10E-09	Spurlock (2008)	30.4	Spurlock (2008)
Cyanazine	NL	155	Spurlock (2008)	236.8	Spurlock (2008)	15.4	Spurlock (2008)	127	Spurlock (2008)	2.13E-04	Spurlock (2008)	6.60E-11	Spurlock (2008)	37.5	Spurlock (2008)
Dimethoate	NL	39800	Spurlock (2008)	10	Spurlock (2008)	2.4	Spurlock (2008)	5.1	Spurlock (2008)	2.40E-01	Spurlock (2008)	1.40E-11	Spurlock (2008)	7.8	Spurlock (2008)
Disulfoton	NL	1.2	DPR (2019)	490.4	DPR (2019); Rao and Davidson (1982)	15.6	DPR (2019)	8910	DPR (2019)	7.199408	DPR (2019)	1.60E-06	DPR (2019)	2.9	DPR (2019)
Ethoprop	NL	843	Spurlock (2008)	183.5	Spurlock (2008)	34.1	Spurlock (2008)	3890	Spurlock (2008)	50.662501	Spurlock (2008)	1.40E-09	Spurlock (2008)	23.3	Spurlock (2008)
Fenamiphos	NL	329	Spurlock (2008)	224.3	Spurlock (2008)	24.2	Spurlock (2008)	1635	Spurlock (2008)	3.1997369	Spurlock (2008)	3.40E-08	Spurlock (2008)	10	Spurlock (2008)
Fonofos	NL	16.9	Spurlock (2008)	894.3	Spurlock (2008)	62.8	Spurlock (2008)	8700	Spurlock (2008)	35.99704	Spurlock (2008)	6.50E-06	Spurlock (2008)	22.4	Spurlock (2008)
Iprodione	NL	12.2	DPR (2019)	700	Lewis et al. (2016)	56	DPR (2019)	1300	DPR (2019)	0.0133322	DPR (2019)	3.56E-09	DPR (2019)	83.9	DPR (2019)
Linuron	NL	77.2	Spurlock (2008)	653	DPR (2019); Hance (1973); Jury et al. (1987); Kookana et.al. (1990); Rao and Davidson (1982); Sanchez-Camazona et al. (2000)	48.9	Spurlock (2008)	1020	Spurlock (2008)	0.1866513	Spurlock (2008)	5.80E-09	Spurlock (2008)	65.9	Spurlock (2008)
Methomyl	NL	57900	Spurlock (2008)	40	Spurlock (2008)	46.2	Spurlock (2008)	1.2	Spurlock (2008)	6.5327961	Spurlock (2008)	1.90E-10	Spurlock (2008)	29.8	Spurlock (2008)
Methyl bromide	NL	17500	Spurlock (2008)	126.5	Spurlock (2008)	9.6	Spurlock (2008)	51.5	Spurlock (2008)	239.98027	Spurlock (2008)	1.60E-02	Spurlock (2008)	3.8	Spurlock (2008)
Napropamide	NL	74	Spurlock (2008)	667.9	Spurlock (2008)	455	Spurlock (2008)	2100	Spurlock (2008)	0.0226648	Spurlock (2008)	8.10E-10	Spurlock (2008)	10	Spurlock (2008)
Oryzalin	NL	2.6	Spurlock (2008)	886.7	Spurlock (2008)	63.3	Spurlock (2008)	5420	Spurlock (2008)	0.0013332	Spurlock (2008)	1.70E-09	Spurlock (2008)	121	Spurlock (2008)

Chemical	Group	Water Solubility (mg/L)	Ref	Soil K_{oc} (cm^3/g)	Ref	Lab Aerobic Half-life (Days)	Ref	K_{ow}	Ref	Vapor Pressure (mPa)	Ref	K_h (atm m^3/mol)	Ref	TFD (Days)	Ref
Oxamyl	NL	280000	Spurlock (2008)	31.6	Spurlock (2008)	10.7	Spurlock (2008)	0.4	Spurlock (2008)	30.664145	Spurlock (2008)	2.40E-10	Spurlock (2008)	9	Spurlock (2008)
Oxyfluorfen	NL	0.1	Spurlock (2008)	6601.4	Spurlock (2008)	291	Spurlock (2008)	29400	Spurlock (2008)	0.0333306	Spurlock (2008)	1.00E-06	Spurlock (2008)	175	Spurlock (2008)
Pendimethalin	NL	0.3	Spurlock (2008)	15000	Spurlock (2008)	126	U.S. EPA (1997)	152000	Spurlock (2008)	1.2532303	Spurlock (2008)	1.30E-08	Spurlock (2008)	42	Spurlock (2008)
Phorate	NL	29	Spurlock (2008)	538.4	Spurlock (2008)	3	Spurlock (2008)	8390	Spurlock (2008)	85.326317	Spurlock (2008)	7.60E-03	Spurlock (2008)	1.8	Spurlock (2008)
Propyzamide	NL	12.9	Spurlock (2008)	825	Spurlock (2008)	26.3	DPR (2019)	1565	Spurlock (2008)	0.0573286	Spurlock (2008)	9.80E-09	Spurlock (2008)	53.5	Spurlock (2008)
Thiobencarb	NL	27.5	Spurlock (2008)	594.7	Spurlock (2008)	37	Spurlock (2008)	29.3	Spurlock (2008)	2.9330921	Spurlock (2008)	2.70E-07	Spurlock (2008)	27.8	Spurlock (2008)
Trifluralin	NL	0.3	Spurlock (2008)	3532.5	Spurlock (2008)	189	Spurlock (2008)	118000	Spurlock (2008)	13.332237	Spurlock (2008)	6.10E-04	Spurlock (2008)	114.5	Spurlock (2008)

Table II-2. Values for quantum data where group designation L indicates that the chemical was determined to be a leacher and designation NL indicates that the chemical was determined to be a non-leacher.

Chemical	Group	MolWt	EHOMO	ELUMO	ElectNeg	Hardness	EAU	Dipole	Conformers	SurfArea	SurfVolume	AccSurfArea
1,2-Dichloropropane (1,2-D)	L	112.987	-10.4	2.07	4.164452	6.237235	-1038.268	0.65	3	125.2822	107.379	93.23
2,3,5,6-Tetrachloroterephthalic acid (TPA)	L	303.912	-9.27	0.13	4.571684	4.699785	-2447.476	0	4	232.8273	225.7189	162.74
Aldicarb sulfone	L	222.265	-9.31	0.96	4.176442	5.137934	-1083.446	6.61	54	229.0382	219.0886	154.86
Atrazine	L	215.688	-8.52	1.48	3.52198	4.998614	-1047.089	5.19	36	240.9175	222.0137	172.99
Bentazon	L	240.283	-8.73	0.21	4.259494	4.472007	-1121.365	4.49	24	228.6703	226.1664	161.5
Bromacil	L	261.119	-8.47	0.71	3.880076	4.585735	-3184.519	5.33	18	226.9503	220.3634	158.18
Deethylatrazine	L	187.634	-8.59	1.44	3.578858	5.014392	-968.4861	5.21	6	200.753	183.2629	149.47
Deisopropylatrazine	L	173.607	-8.61	1.44	3.586809	5.027262	-929.1784	5.28	6	184.1979	163.9341	141.88
Dibromochloropropane (DBCP)	L	236.334	-10.09	0.89	4.598338	5.489058	-5725.316	1.88	9	149.7239	136.8289	109.82
Diuron	L	233.098	-7.97	1.41	3.28009	4.686691	-1453.914	7.27	4	228.9071	216.2525	177.53
Ethylene dibromide (EDB)	L	187.862	-10.01	1.15	4.43026	5.57887	-5226.434	0	3	119.5127	99.97471	94.83
Hexazinone	L	252.318	-8.75	1.37	3.691024	5.063135	-838.5889	7.05	12	268.0428	268.6809	177.22
Imidacloprid	L	255.665	-9.06	0.52	4.270251	4.792077	-1233.01	8.12	18	245.5336	233.1247	183.24
Metolachlor ESA	L	329.417	-8.9	1.09	3.905705	4.997346	-1414.146	8.94	5832	310.9803	338.7759	197.08
Norflurazon	L	303.671	-8.15	0.07	4.037543	4.111077	-1461.561	7.59	4	266.2571	256.9084	195.22
Prometon	L	225.296	-8.08	2.38	2.850596	5.23276	-741.3193	3.18	72	267.881	251.5868	185.71
Simazine	L	201.661	-8.54	1.48	3.527289	5.009324	-1007.781	5.26	36	224.2235	202.6566	165.82
Tebuthiuron	L	228.32	-8.37	1.17	3.601704	4.770344	-1044.807	4.42	8	246.6778	239.8559	179.62
1,3-Dicloropropene	NL	110.971	-9.38	1.23	4.074967	5.302176	-1037.024	2.3	3	123.1526	100.5901	99.62
Alachlor	NL	269.772	-8.6	1.49	3.557665	5.046947	-1210.692	4.55	432	280.6873	292.0321	180.13
Carbaryl	NL	201.225	-7.61	0.84	3.383257	4.22423	-668.9088	3.14	4	220.7113	207.6871	174.54
Carbofuran	NL	221.256	-7.61	2.02	2.796716	4.813376	-746.5281	2.15	8	244.8915	232.7333	169.94
Chlorthal-dimethyl	NL	331.966	-9.16	0.29	4.434514	4.720683	-2526.065	0	4	274.0232	264.7568	189.2
Cyanazine	NL	240.698	-8.69	1.28	3.705218	4.982284	-1139.289	5.12	36	255.9952	241.7336	177.9
Dimethoate	NL	229.261	-8.71	1.4	3.654853	5.050165	-1615.781	1.77	486	216.9186	217.5841	149.65
Disulfoton	NL	274.41	-8.26	2.18	3.039173	5.22017	-2002.527	6.06	19683	280.1453	284.7407	181.32
Ethoprop	NL	242.344	-8.85	1.89	3.483636	5.370101	-1604.359	0.38	6561	278.8376	260.2836	177.93
Fenamiphos	NL	303.363	-8.14	1.55	3.29774	4.846589	-1528.413	3.74	972	349.77	307.5	219.3313
Fonofos	NL	246.335	-8.24	1.1	3.57099	4.668583	-1602.911	2.64	81	260.5236	256.7886	179.92
Iprodione	NL	330.171	-9.13	0.63	4.246641	4.880533	-1813.304	2.87	12	309.7647	299.4441	221.23
Linuron	NL	249.10	-8.07	1.31	3.37853	4.690964	-1529.056	6.42	72	240.8577	226.5673	182
Methomyl	NL	162.213	-8.16	1.99	3.086866	5.073062	-854.4956	4.1	18	188.9408	165.2837	143.47
Methyl bromide	NL	94.939	-9.74	2.09	3.823914	5.915406	-2613.814	2.31	1	76.18852	56.59675	68.76
Napropamide	NL	271.36	-7.46	0.9	3.28083	4.183127	-865.4086	4.1	324	297.5041	303.727	209.3
Oryzalin	NL	346.364	-8.55	-1.01	4.780038	3.76906	-1536.01	3.65	972	319.3905	329.0067	197.98
Oxamyl	NL	219.265	-8.44	1.18	3.631618	4.809358	-1062.428	3.55	36	246.1277	209.2071	168.2895
Oxyfluorfen	NL	361.703	-8.6	-0.05	4.324424	4.274865	-1693.087	6.9	24	318.9478	308.0492	216.17
Pendimethalin	NL	281.312	-8.27	-0.72	4.496152	3.773952	-971.4962	4.67	108	287.7683	293.7902	188.43
Phorate	NL	260.383	-8.31	1.59	3.361817	4.952663	-1963.221	2.88	6561	272.4863	265.6514	179.03
Propyzamide	NL	256.132	-9.1	0.36	4.370265	4.726335	-1513.992	4.12	6	259.4479	247.4607	189.99
Thiobencarb	NL	257.785	-8.55	1.31	3.617862	4.932611	-1455.08	4.16	162	278.3089	267.8525	196.99
Trifluralin	NL	335.282	-8.55	-1	4.77713	3.774081	-1269.116	0.61	81	302.1828	308.4823	182.82

Table II-2. Continued.

Chemical	Group	MinElPot	MinLoc IonPot	PSA	Polar Area75	AccPolar Area75	Polar Area100	AccPolar Area100	Polar Area125	AccPolar Area125	MaxElPot	Polarizability
1,2-Dichloropropane (1,2-D)	L	-66.11	50.14	0	15.05	13.28	1.99	1.97	0	0	109.51	46.39
2,3,5,6-Tetrachloroterephthalic acid (TPA)	L	-141.72	53.95	68.935	45.55	38.86	33.46	30.48	20.76	18.65	306.21	56.01
Aldicarb sulfone	L	-189.84	50.57	74.417	85.83	61.62	53.94	40.66	42.16	35.41	235.59	55.85
Atrazine	L	-178.04	45.73	41.935	62.1	40.86	34.03	19.34	17.12	7.95	197.06	56.02
Bentazon	L	-148.23	50.03	62.048	71.16	52.94	50.83	41.08	27.58	22.74	280.61	57.07
Bromacil	L	-201	46.33	37.313	65.77	46.79	46.96	32.29	32.79	22.82	273.91	56.32
Deethylatrazine	L	-177.85	46.12	56.781	66.13	49.83	39.33	28.19	22.74	16.3	223.64	52.85
Deisopropylatrazine	L	-178.35	46.12	57.011	68.86	53.1	41.02	30.07	23.8	17.27	220.68	51.35
Dibromochloropropane (DBCP)	L	-93.18	47.01	0	33.51	23.38	16.03	13.18	2.64	2.18	138.27	48.59
Diuron	L	-170.11	48.6	20.857	74.97	52.51	40.7	27.95	22.28	14.88	254.62	56.08
Ethylene dibromide (EDB)	L	-55.54	46.59	0	23.55	19.54	7.74	7.74	0	0	116.1	45.93
Hexazinone	L	-228.96	46.02	34.593	83.58	58.41	46.53	33	28.96	21.42	131.25	60.08
Imidacloprid	L	-192.02	46.8	66.502	126.97	94.7	96.33	73.02	62.02	48.76	204.04	57.11
Metolachlor ESA	L	-190.66	49.45	68.251	96.62	64.39	49.22	35.17	34.03	26.57	138.77	65.36
Norflurazon	L	-212.86	50.09	32.008	75.57	55.21	51.2	35.81	26.76	17.51	230	60
Prometon	L	-218.81	43.56	47.601	48.29	29.35	32.27	21.22	20.48	14.82	177.45	58.49
Simazine	L	-179.23	45.73	42.155	64.41	43.82	35.69	21.27	18.21	9.14	196.96	54.52
Tebuthiuron	L	-197.23	46.83	38.957	39.57	29.51	25.51	19.79	18.16	14.63	97.96	57.56
1,3-Dicloropropene	NL	-83.82	49.59	0	30.48	25.43	13.48	11.13	3.92	2.76	145.07	46.48
Alachlor	NL	-197.08	48.14	19.855	48.29	34.32	22.18	14.37	14.49	10.91	91.38	61.93
Carbaryl	NL	-203.45	46.75	30.436	51.73	40.84	23.65	16.83	18.09	13.65	219.6	56.46
Carbofuran	NL	-209.87	46.53	38.694	57.24	33.79	34.24	17.66	23.91	12.88	199.5	57.73
Chlorthal-dimethyl	NL	-152.82	53.52	42.199	59.47	52.23	25.88	23.56	13.91	13.91	112.38	59.28
Cyanazine	NL	-181.25	46.94	55.364	98.76	71.25	52.32	36.4	26.69	19.65	227.18	57.54
Dimethoate	NL	-187.59	41.62	39.448	67.79	53.32	29.28	21.88	13.64	10.92	134.63	55.24
Disulfoton	NL	-156.27	39.63	12.731	68.51	49.04	31.24	25.16	11.29	10.75	109.71	60.03
Ethoprop	NL	-189.16	42.5	24.738	36.4	23.47	21.25	13.59	13.47	9.77	76.82	58.41
Fenamiphos	NL	-206.55	39.49	43.689	49.51	29.01	26.47	17.7	17.35	12.44	201.18	64.03
Fonofos	NL	-154.58	40	7.33	47.81	32.55	24.77	16.14	10.83	7.78	91.83	58.92
Iprodione	NL	-163.74	51.58	48.769	72.09	54.74	41.58	34.64	15.38	13.34	148.97	62.49
Linuron	NL	-159.79	48.95	29.428	68.13	49.03	31.76	22.01	13.74	9.51	226.19	56.89
Methomyl	NL	-251.51	41.87	40.183	46.83	33.6	28.57	18.43	23.89	16.28	216.63	51.74
Methyl bromide	NL	-80.11	45.09	0	22.05	22.05	3.09	3.09	0	0	108.72	42.78
Napropamide	NL	-199.95	46.29	17.86	40.44	26.78	15.48	9.47	12.66	8.56	105.86	63.95
Oryzalin	NL	-134.58	50.78	135.485	88.78	68.3	50.82	41.14	20.97	17.29	264.53	64.61
Oxamyl	NL	-224.187	42.46207	54.539	70.48	48.72	41.36	28.58	33.62	25.02	210.5	56.08
Oxyfluorfen	NL	-179.81	49.47	50.343	89.23	61.47	54.45	39.29	30.56	23.85	148.77	63.97
Pendimethalin	NL	-143.74	47.86	81.147	74.35	54.56	32.4	22.88	6.54	5.5	117.81	62.39
Phorate	NL	-125.42	40.23	15.669	47.64	37.15	13.21	11.46	0.01	0.01	94.24	58.71
Propyzamide	NL	-169.18	47.58	21.584	35.5	23.98	25.86	18.36	17.18	12.6	218.85	58.85
Thiobencarb	NL	-173.33	42.02	13.609	34.1	24.47	12.6	8.46	9.92	7.54	97.38	59.99
Trifluralin	NL	-119.02	53.23	74.177	51.5	33.16	22.7	14.39	6.51	3.66	156.33	63.68

References

Empirical Variables in Table II-1

Bayless, E.R., P.D. Capel, J.E. Barbash, R.M.T. Webb, T.C. Hancock, and D.C. Lampe. 2008. Simulated fate and transport of metolachlor in the unsaturated zone, Maryland, USA. J. Environ. Qual. 37:1064-1072.

Cohen, S.Z., M. Creeger, R.F. Carsel and C.G. Enfield. 1983. Potential pesticide contamination of groundwater from agricultural uses. American Chemical Society Symposium Series, Treatment and Disposal of Pesticide Waste. Washington D.C., Aug 28 – Sept 2.

DPR. 1995. Soil dissipation study: determination of residues of DCPA (dimethyl tetrachloroterephthalate), its degradation products, and HCB in soil (CA)) – 1992 to 1994 (AB2021) (405P). Study ID # 135722, Registration Branch, California Department of Pesticide Regulation, Sacramento, CA, USA.

DPR. 2019. Pesticide Chemistry Database. Environmental Monitoring Branch, California Department of Pesticide Regulation, Sacramento, CA, USA.

Hammer, C. and K. White. 2009. Risks of DCPA use to federally threatened California red-legged frog (Rana aurora draytonii). Environmental Fate and Effects Division, United States Environmental Protection Agency, Washington, DC, USA.

Hance, R.J. 1973. Soil organic matter and the adsorption and decomposition of the herbicides atrazine and linuron. Soil Biol. Biochem., 6: 39-42.

Jury, W.A., D.D. Focht and W.J. Farmer. 1987. Evaluation of pesticide groundwater pollution potential from standard indices of soil-chemical adsorption and biodegradation. J. Environ. Qual., 16(4): 422-428.

Kookana, R.S., R.G. Gerritse and L.A.G. Aylmore. 1990. Effect of organic cosolvent on adsorption and desorption of linuron and simazine in soil. Aust. J. Soil Res., 28: 717-725.

Kruger, E.L. 1992. Degradation and movement of atrazine and deisopropylatrazine in soil. A Thesis Submitted to the Graduate Faculty in Partial Fulfillment of the Requirements for the Degree of Master of Science. Iowa State University, Ames, Iowa.

Krutz, L.J., D.L. Shaner and R.M. Zablotowicz. 2010. Enhanced degradation and soil depth effects on the fate of atrazine and major metabolites in Colorado and Mississippi Soils. Journ. Environ. Qual. 39:1369-1377.

Lewis, K.A., J. Tzilivakis, D. Warner and A. Green. 2016. An international database for pesticide risk assessments and management. Human and Ecological Risk Assessment: An International Journal, 22(4): 1050-1064.

Rao, P.S.C. and J.M. Davidson. 1982. Retention and transformation of selected pesticides and phosphorus in soil-water systems: a critical review. Environmental Research Laboratory. United States Environmental Protection Agency, Washington, DC, USA.

Sanchez-Camazona, M., M.J. Sanchez-Martin and R. Delagado-Pascual. 2000. Adsorption and mobility of linuron in soils as influenced by soil properties, organic amendments, and surfactants. J. Agric. Food Chem, 48: 3018-3026.

Spurlock, F. 2008. Distribution and variance/covariance structure of pesticide environmental fate data. Environ. Tox. and Chem., 27(8): 1683-1690.

USDA-ARS DB. 2019. The ARS pesticides properties database. United States Department of Agriculture, Agricultural Research Service. Beltsville, Maryland, USA.

U.S. EPA. 1994. Reregistration eligibility decision (RED), bentazon. Office of Prevention, Pesticides and Toxic Substances. United States Environmental Protection Agency, Washington, DC, USA.

U.S. EPA. 1997. Reregistration eligibility decision (RED), pendimethalin. Office of Prevention, Pesticides and Toxic Substances. United States Environmental Protection Agency, Washington, DC, USA.

U.S. EPA. 2012. Estimation Programs Interface Suite. United States Environmental Protection Agency, Washington, DC, USA.

Calculated Variables in Table II-2

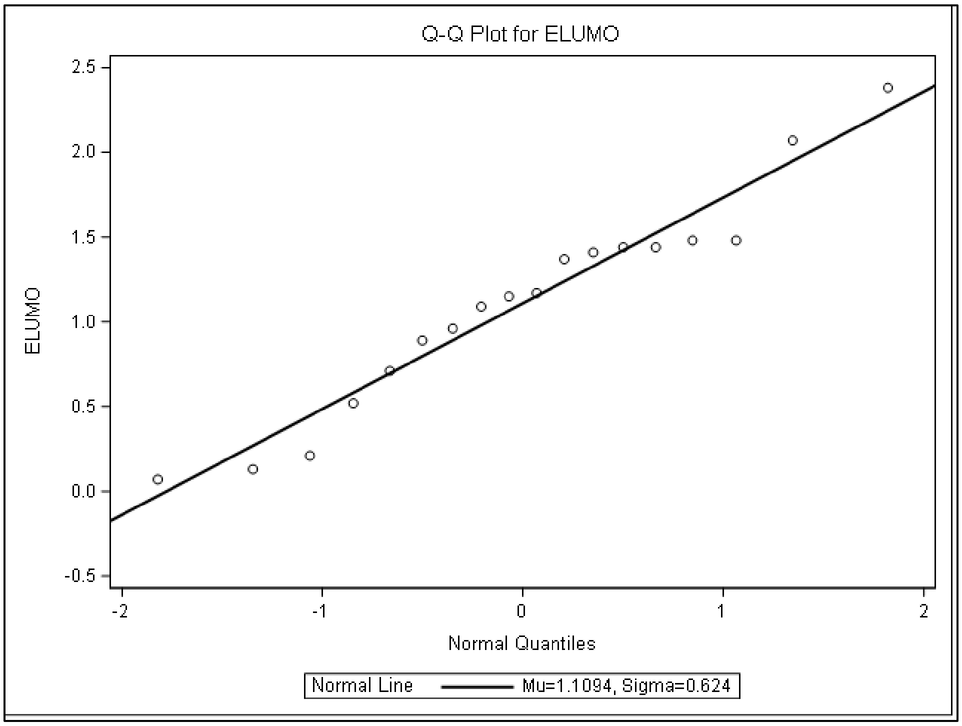
Spartan '20 software available at: http://wavefun.com (verified 17 March 2021).

U.S. EPA. TEST software available at: https://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test (verified 24 July 2024).

Appendix III — Comparison of Quantile-Quantile Plots for Chemical Properties with Mixed Results for Test of Normal Distribution

Table III-1. Q-Q plots of raw data for leacher and non-leacher chemical groups for E_{LUMO} property. Shapiro-Wilk *P*-value at 0.48 for leacher and 0.02 for non-leacher group.

Leacher



Non-Leacher

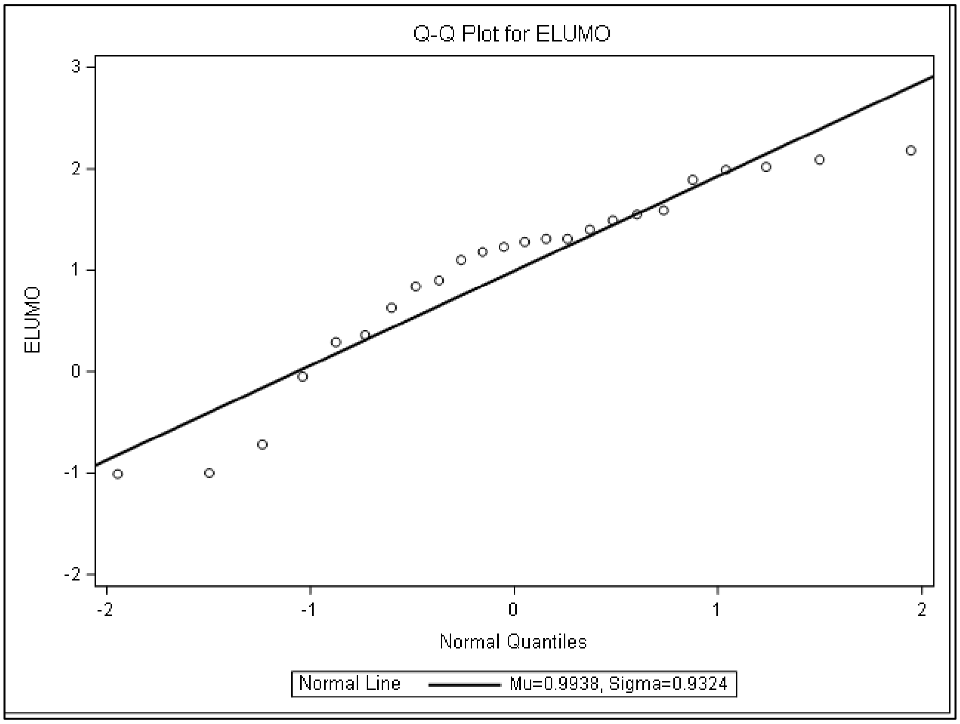
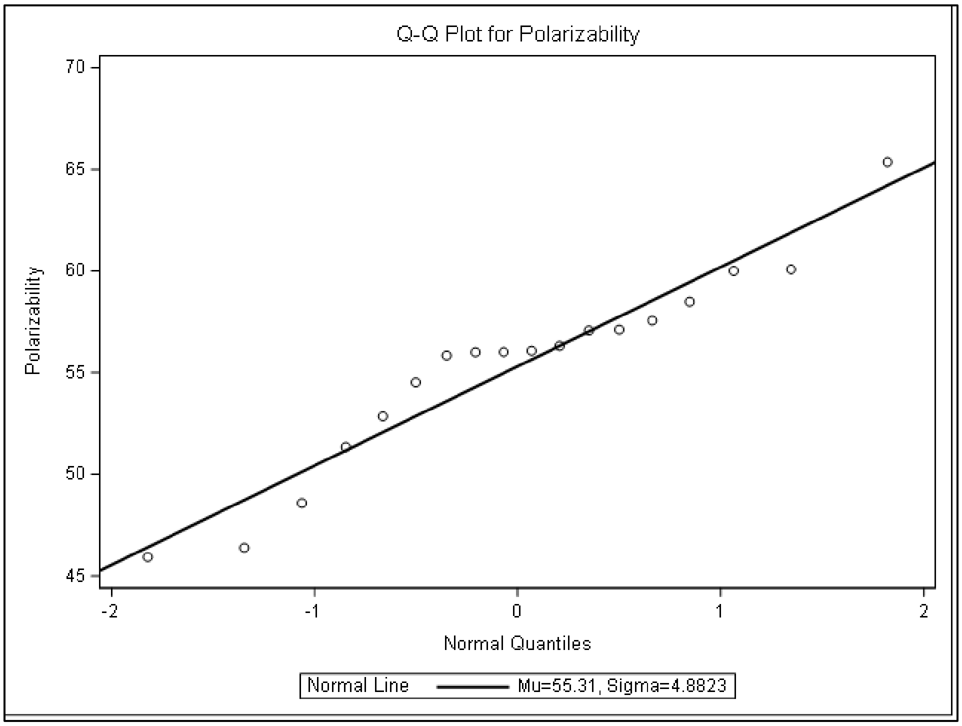


Table III-2. Q-Q plots of raw data for leacher and non-leacher chemical groups for Polarizability property. Shapiro-Wilk *P*-value at 0.25 for leacher and 0.004 for non-leacher group.

Leacher



Non-Leacher

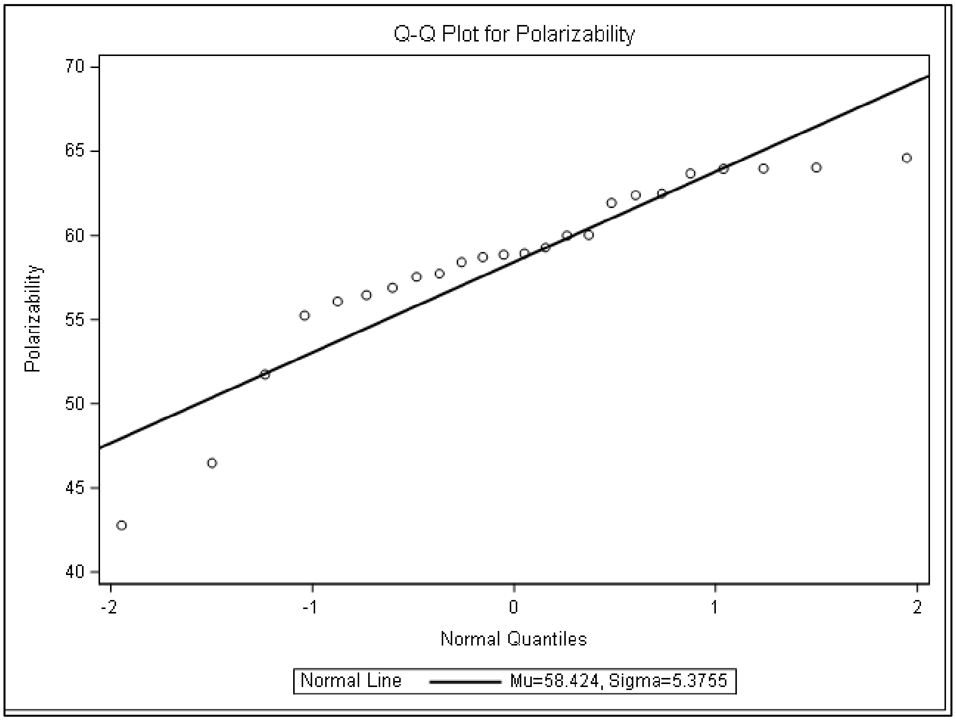
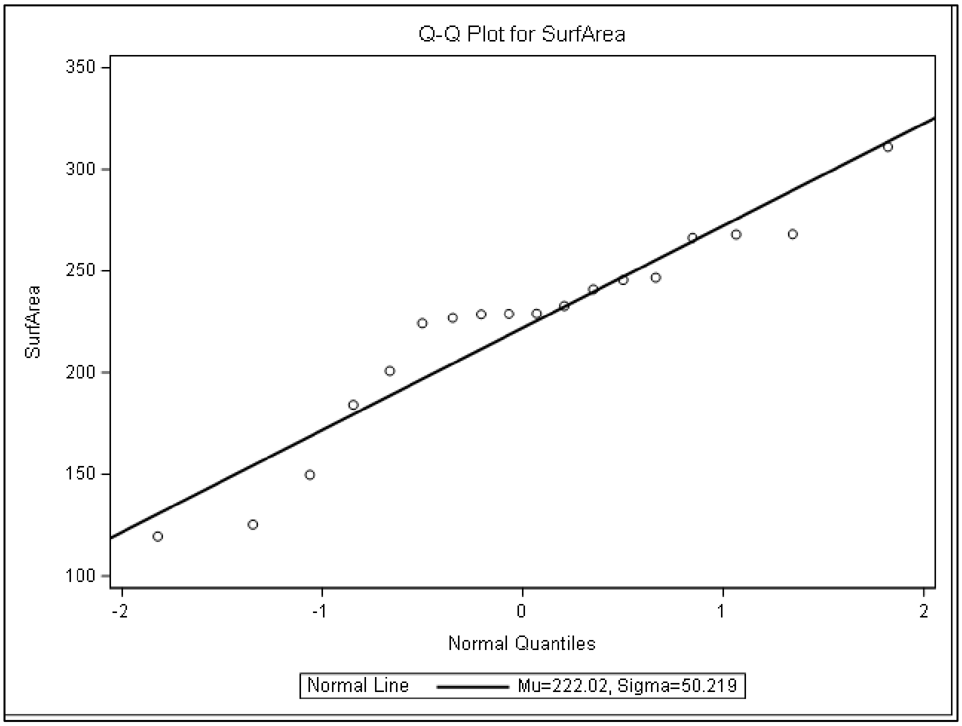


Table III-3. Q-Q plots of raw data for leacher and non-leacher chemical groups for SurfArea property. Shapiro-Wilk *P*-value at 0.10 for leacher and 0.006 for non-leacher group.

Leacher



Non-Leacher

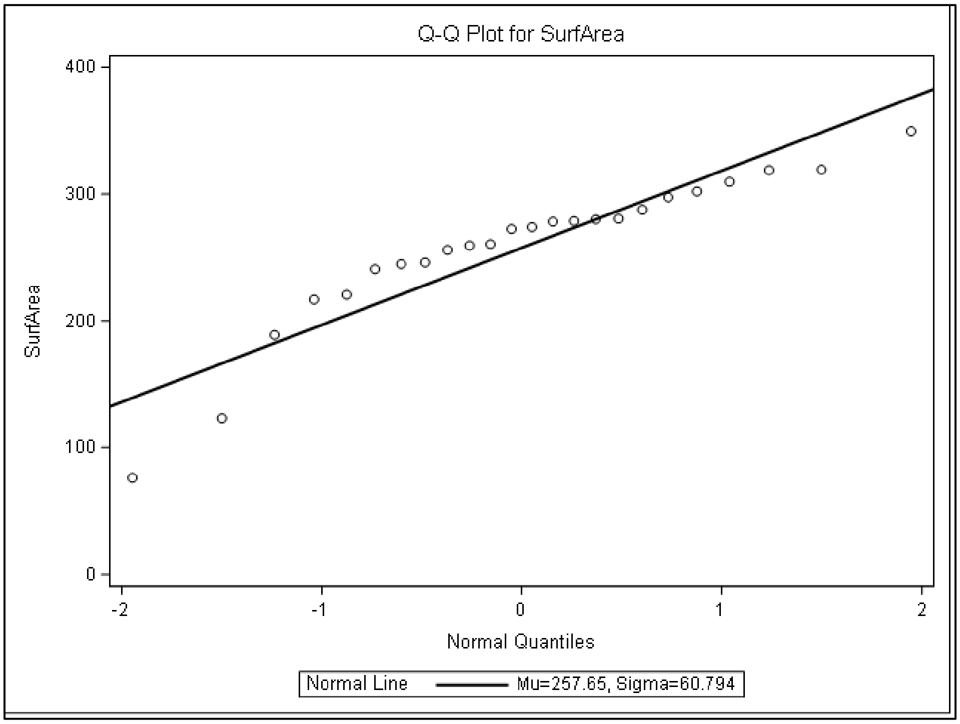
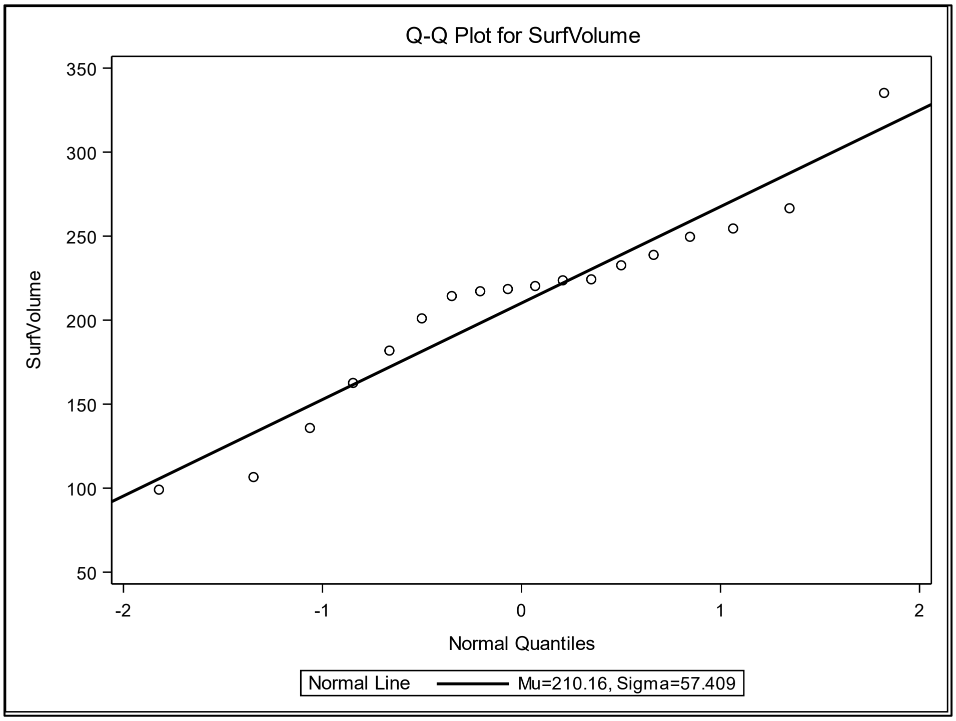


Table III-4. Q-Q plots of raw data for leacher and non-leacher chemical groups for SurfVolume property. Shapiro-Wilk *P*-value at 0.27 for leacher and 0.004 for non-leacher group.

Leacher



Non-Leacher

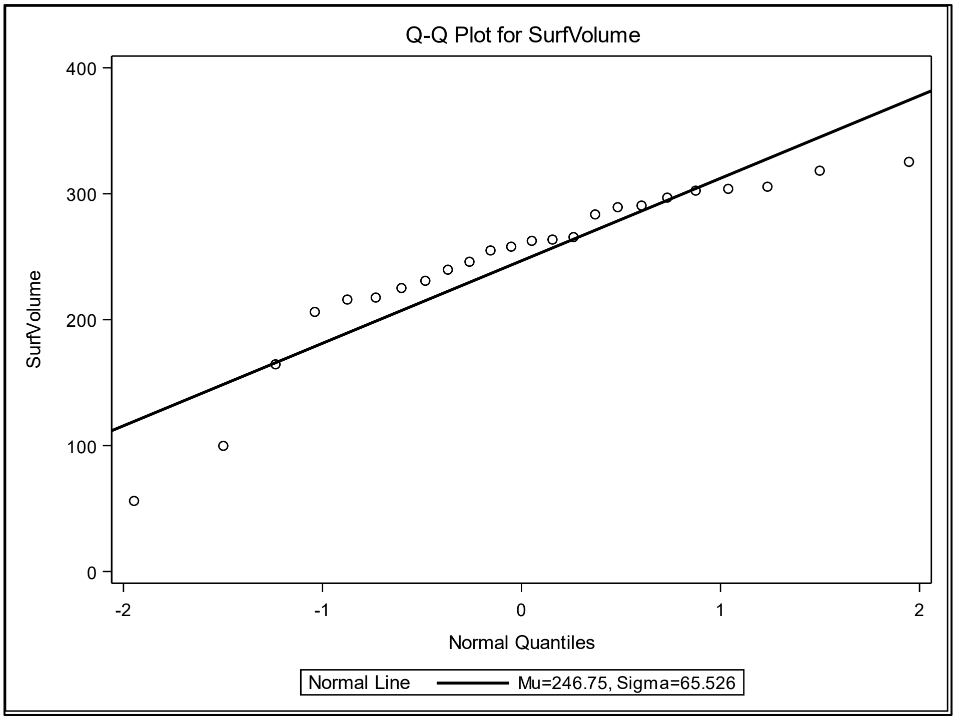
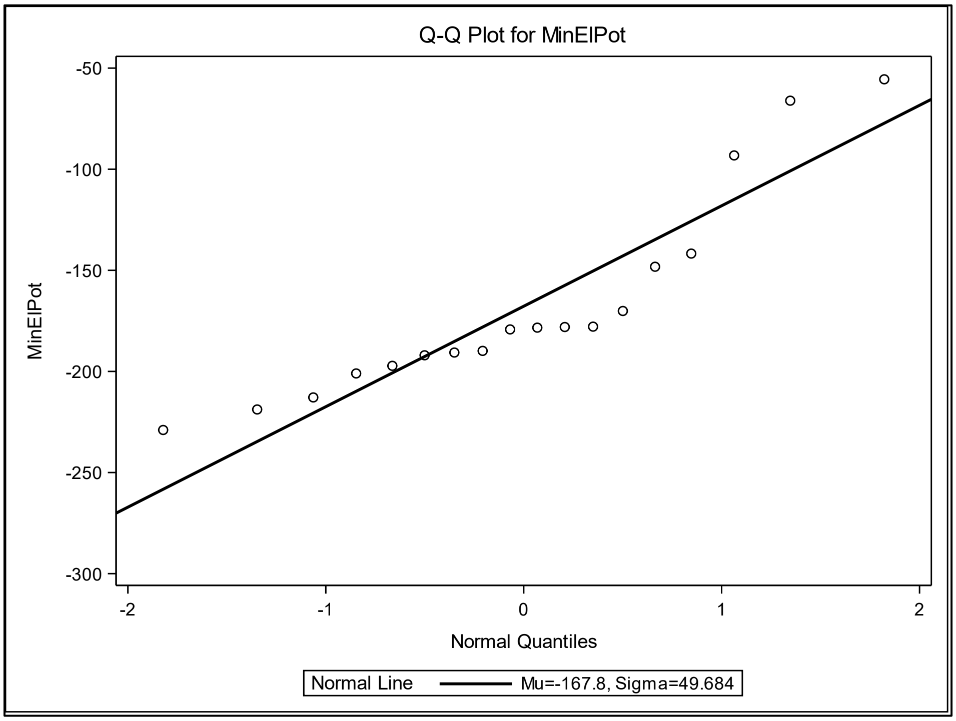


Table III-5. Q-Q plots of raw data for leacher and non-leacher chemical groups for MinElPot property. Shapiro-Wilk *P*-value at 0.009 for leacher and 0.82 for non-leacher group.

Leacher



Non-Leacher

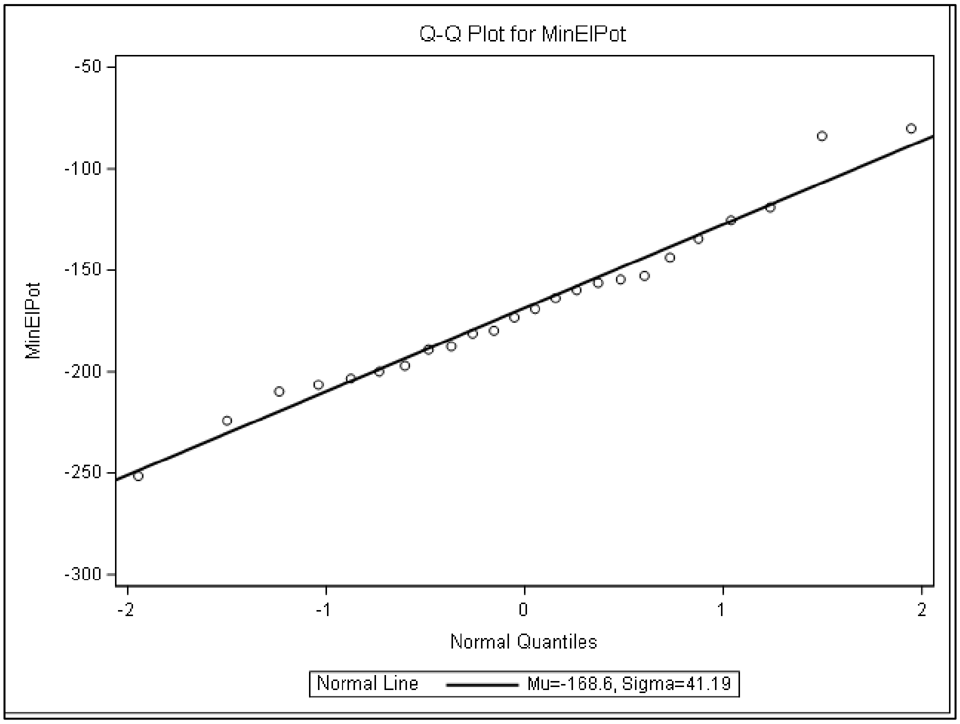
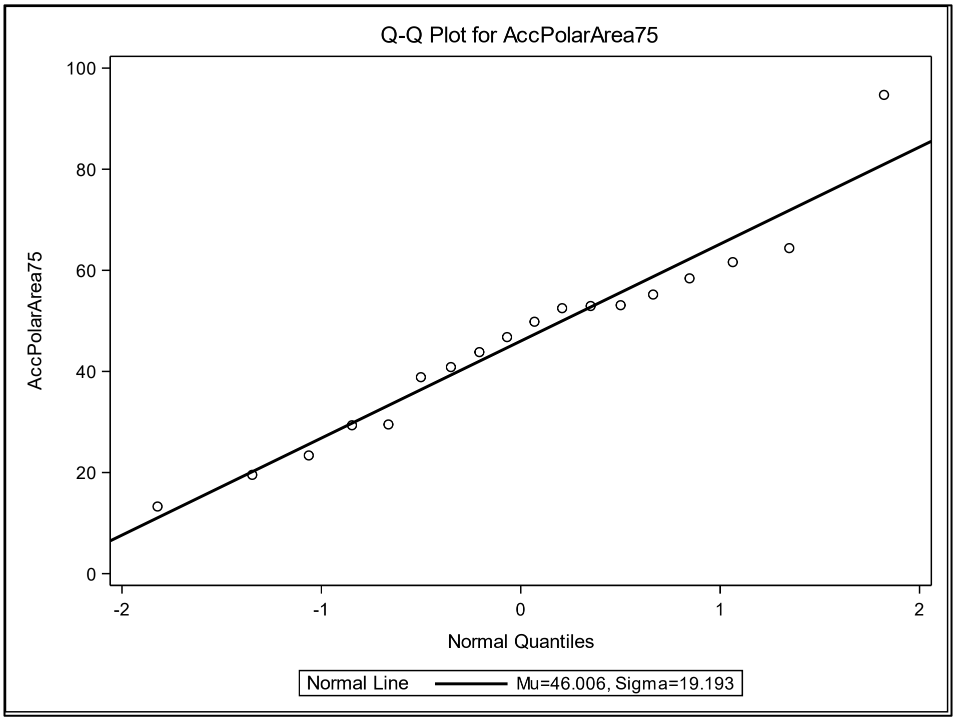


Table III-6. Q-Q plots of raw data for leacher and non-leacher chemical groups for AccPolarArea75 property. Shapiro-Wilk *P*-value at 0.49 for leacher and 0.07 for non-leacher group.

Leacher



Non-Leacher

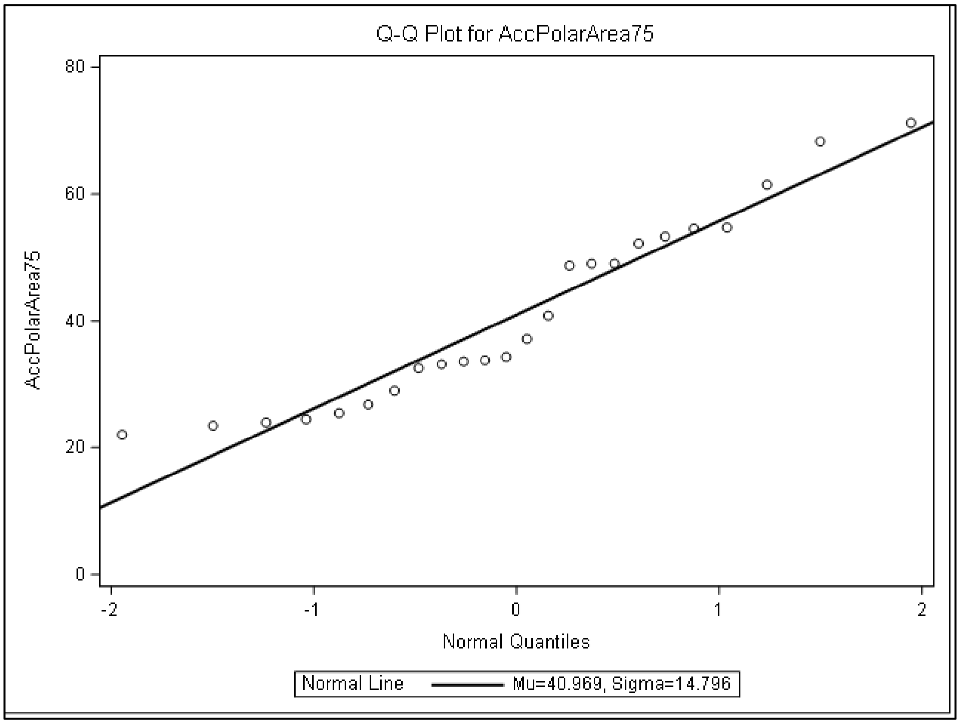
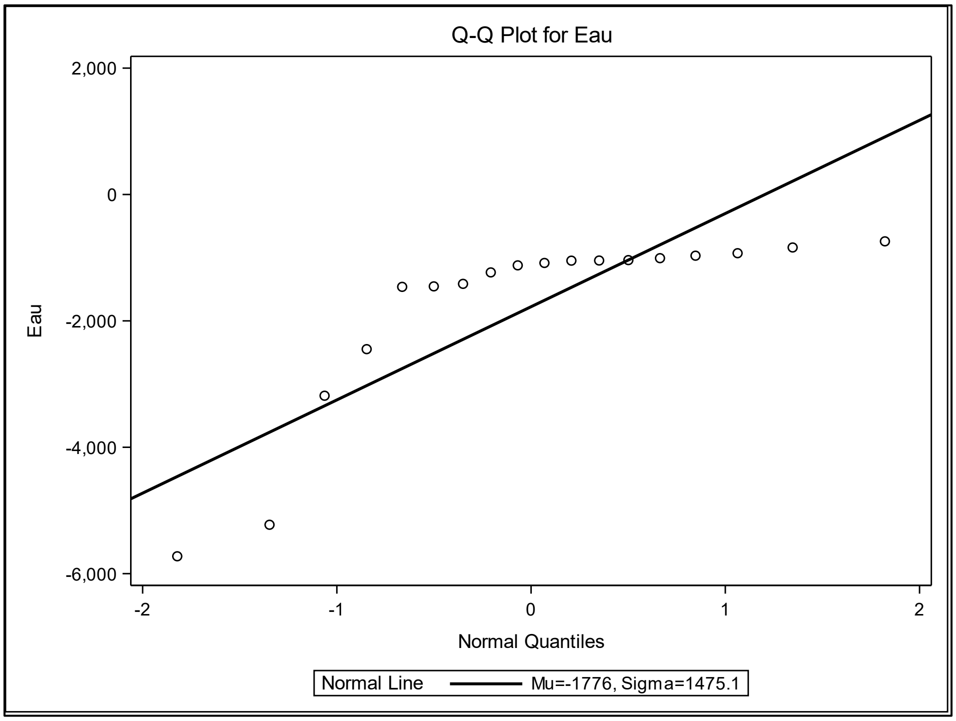


Table III-7. Q-Q plots between raw data for leacher and non-leacher chemical groups for EAU property. Shapiro-Wilk *P*-value at 0.001 for leacher and 0.27 for non-leacher group.

Leacher



Non-Leacher

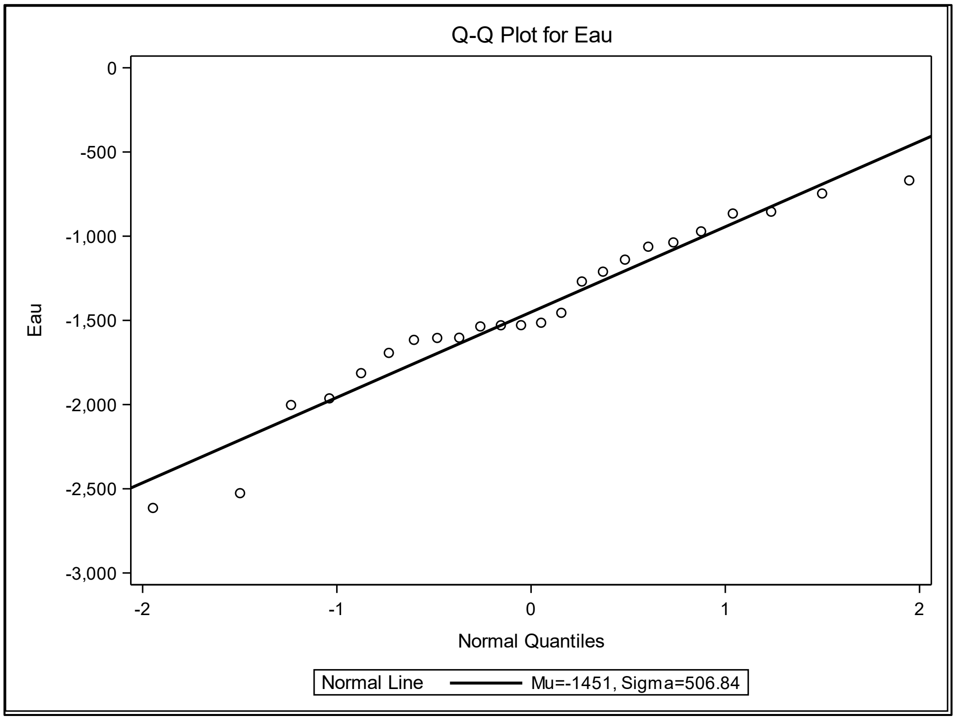
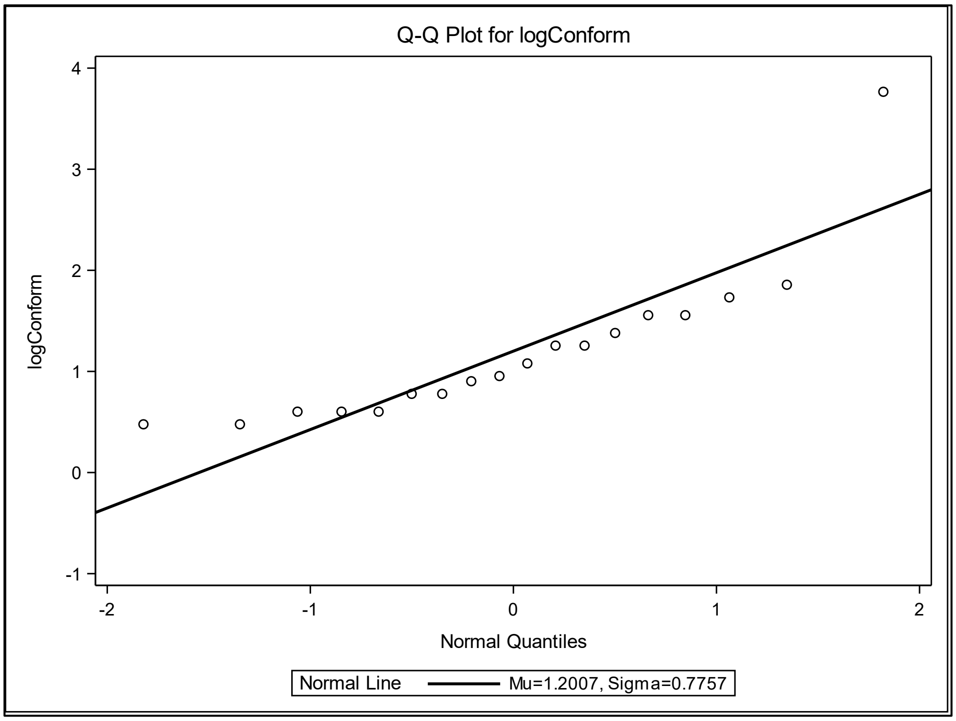


Table III-8. Q-Q plots of data for leacher and non-leacher chemical groups for LogConformers property. Shapiro-Wilk *P*-value at 0.001 for leacher and 0.64 for non-leacher group.

Leacher



Non-Leacher

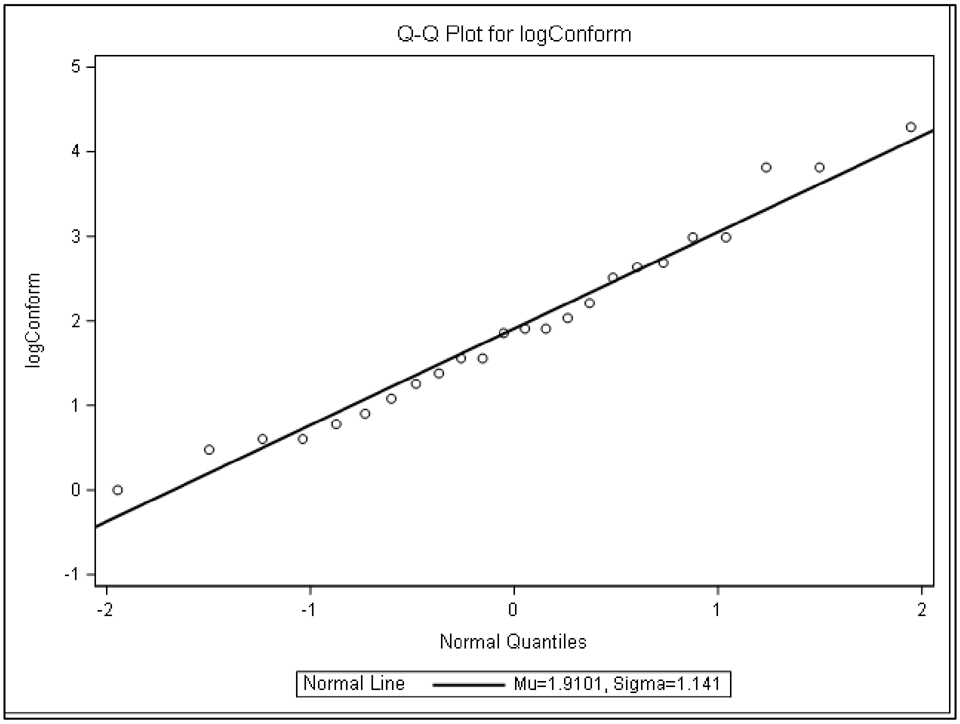
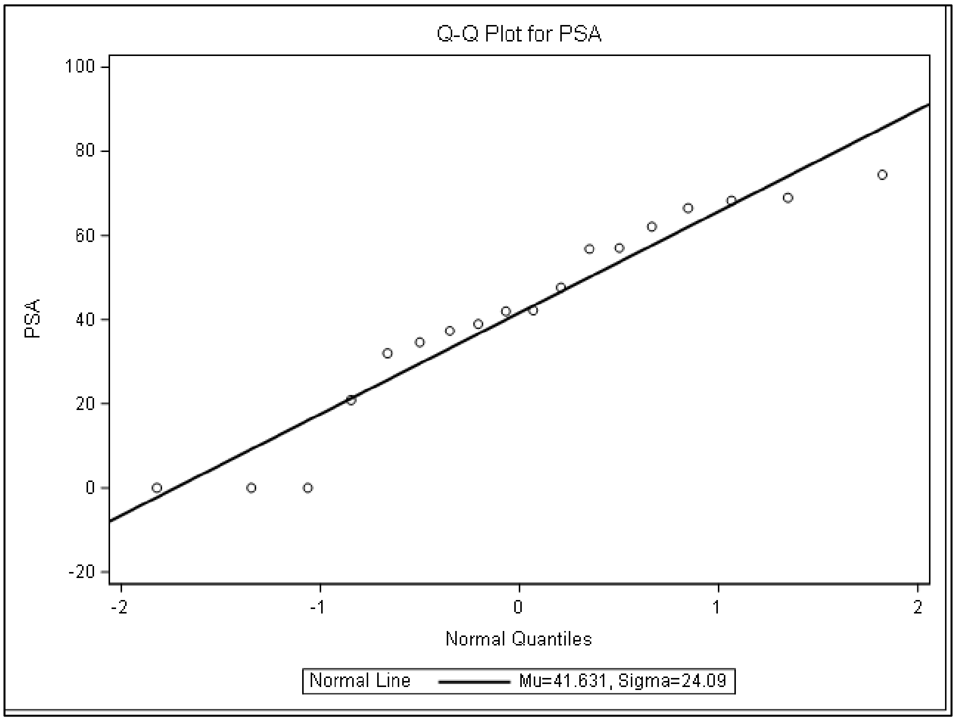
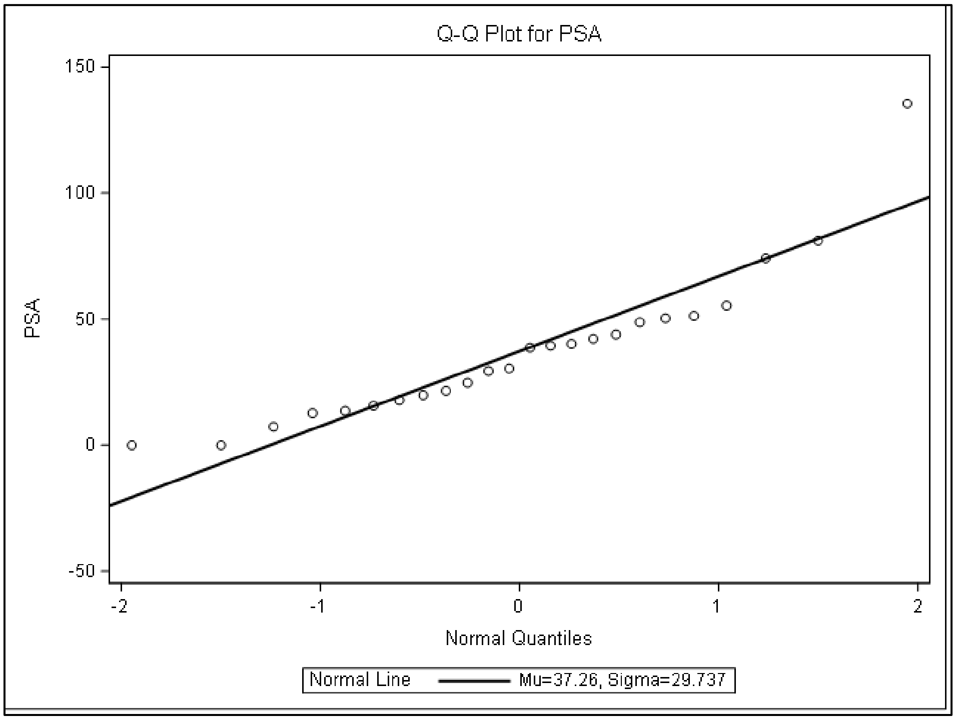


Table III-9. Q-Q plots of raw data for leacher and non-leacher chemical groups for PSA property. Shapiro-Wilk value *P*-value at 0.10 for leacher and 0.007 for non-leacher group.

Leacher

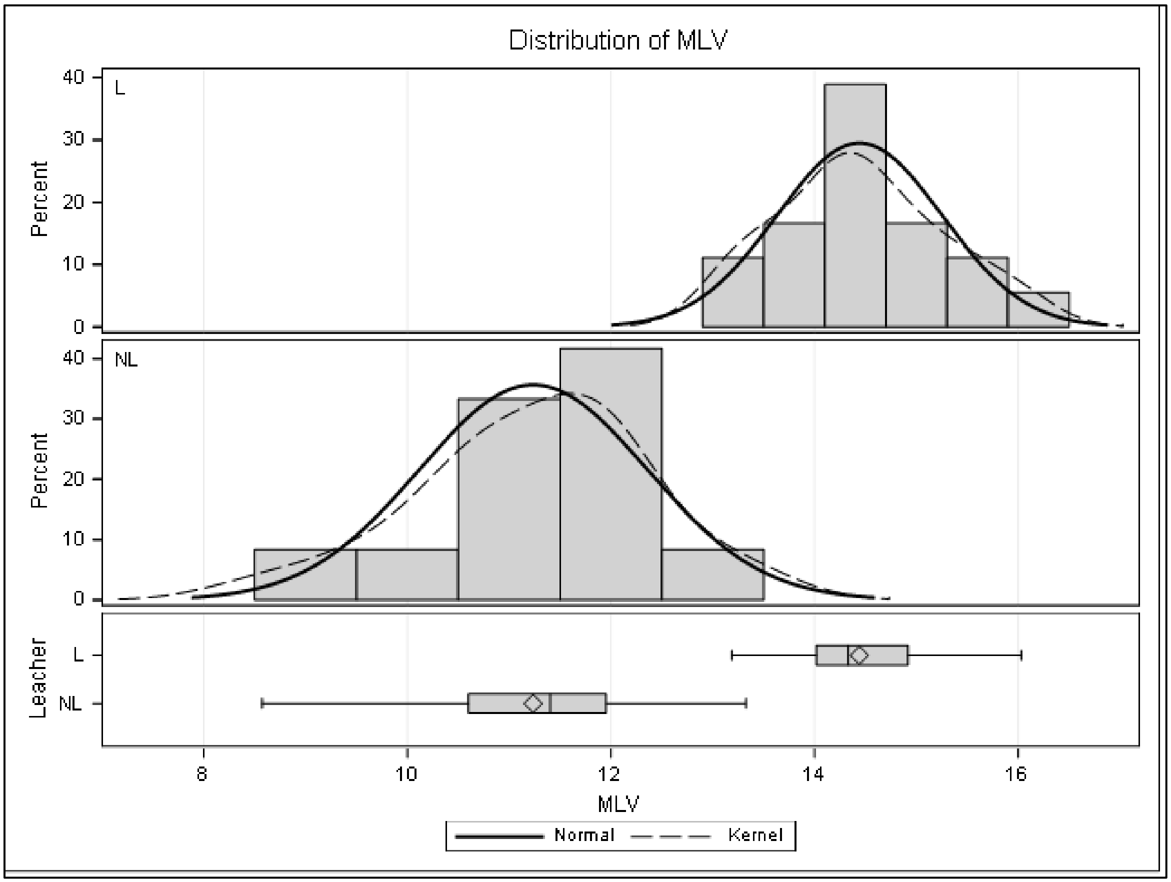


Non-Leacher



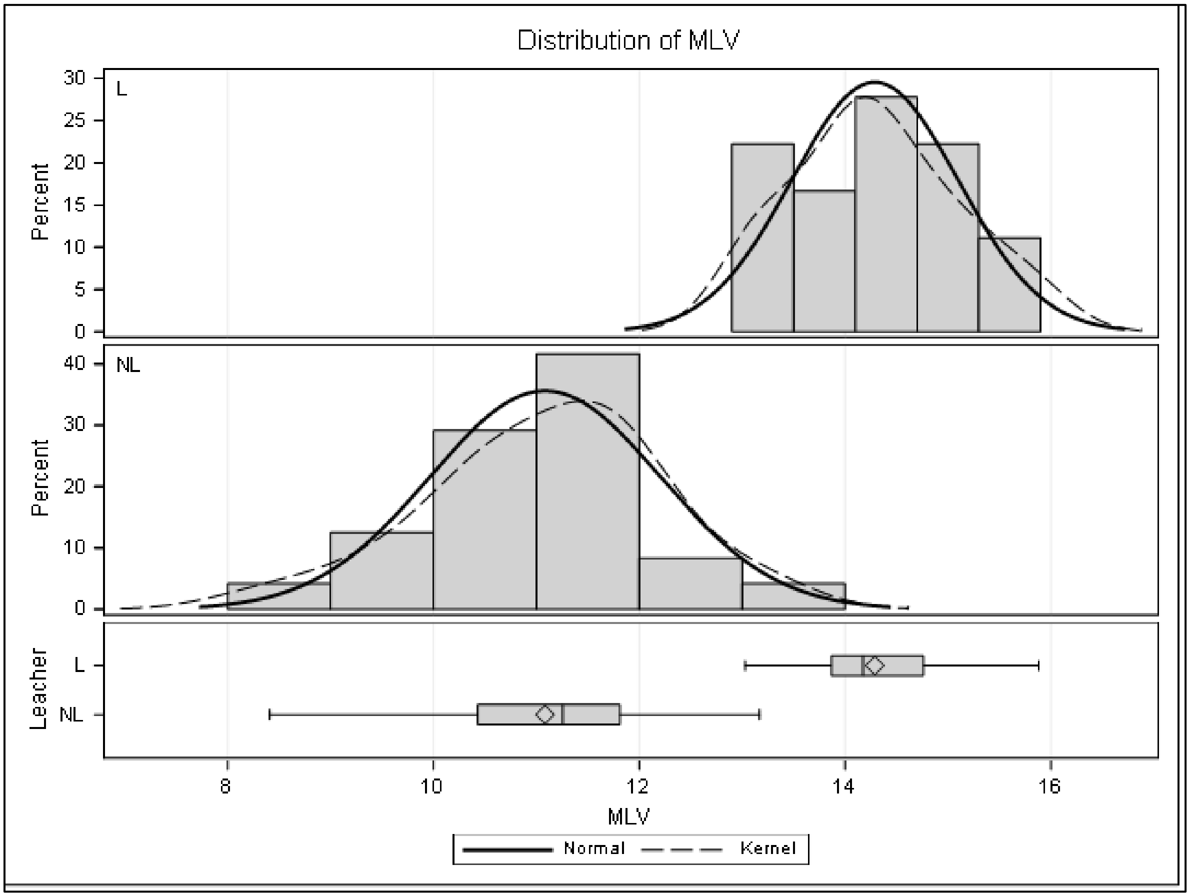
Appendix IV — Expanded Results and Statistics for the Full 11-Parameter Model and Subsequent Reduced-Parameter Models

Table IV-1. Distributions of multivariate leaching values (MLVs) with associated fits of the normal and kernel distribution for the leacher (L) and non-leacher (NL) chemical groups of the full 11-parameter canonical discriminant analysis model. Box plots indicate the mean, median, interquartile range, and minimum and maximum MLVs of each chemical group. Numerical statistics indicate standardized and raw canonical coefficients. An asterisk (*) denotes the smallest absolute standardized coefficient identified in the model to be removed during the next model-parameter reduction process.



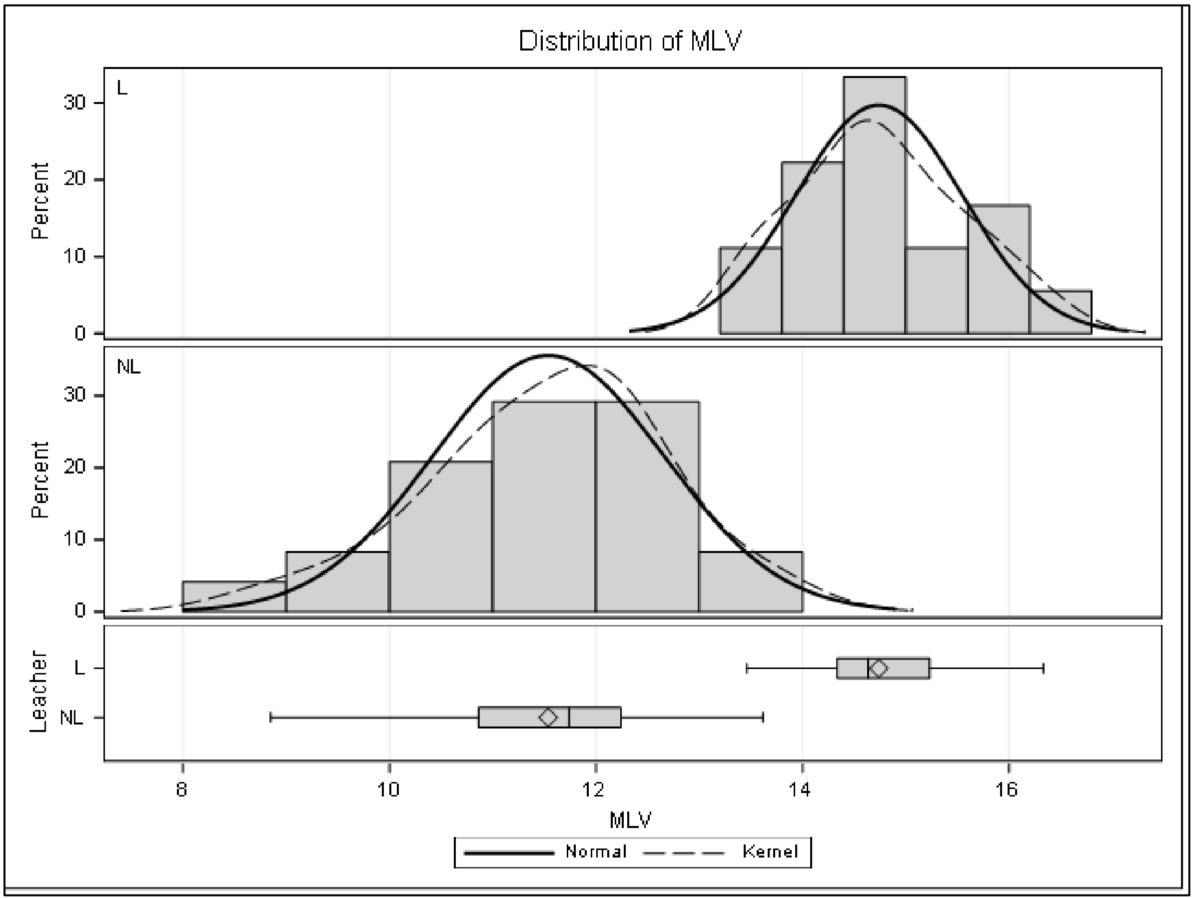
11-Parameter Model		
Variable	**Standardized Canonical Coefficients**	**Raw Canonical Coefficients**
$LogK_{oc}$	-0.9331	-1.2561
LogAero	1.2815	2.0045
$LogK_h$	0.2201	0.0690
PolarArea100	0.0635	0.0037
Dipole	0.4254	0.1842
E_{HOMO}	-0.4040	-0.6320
LogMaxElPot	0.3781	2.3066
MinLocIonPot	0.0709	0.0189
Hardness	0.0755	0.1475
SurfArea	-0.1664	-0.0028
LogConformers*	0.0257	0.02446

Table IV-2. Distributions of multivariate leaching values (MLVs) with associated fits of the normal and kernel distribution for the leacher (L) and non-leacher (NL) chemical groups of the ten-parameter canonical discriminant analysis model. Box plots indicate the mean, median, interquartile range, and minimum and maximum MLVs of each chemical group. Numerical statistics indicate standardized and raw canonical coefficients. An asterisk (*) denotes the smallest absolute standardized coefficient identified in the model to be removed during the next model-parameter reduction process.



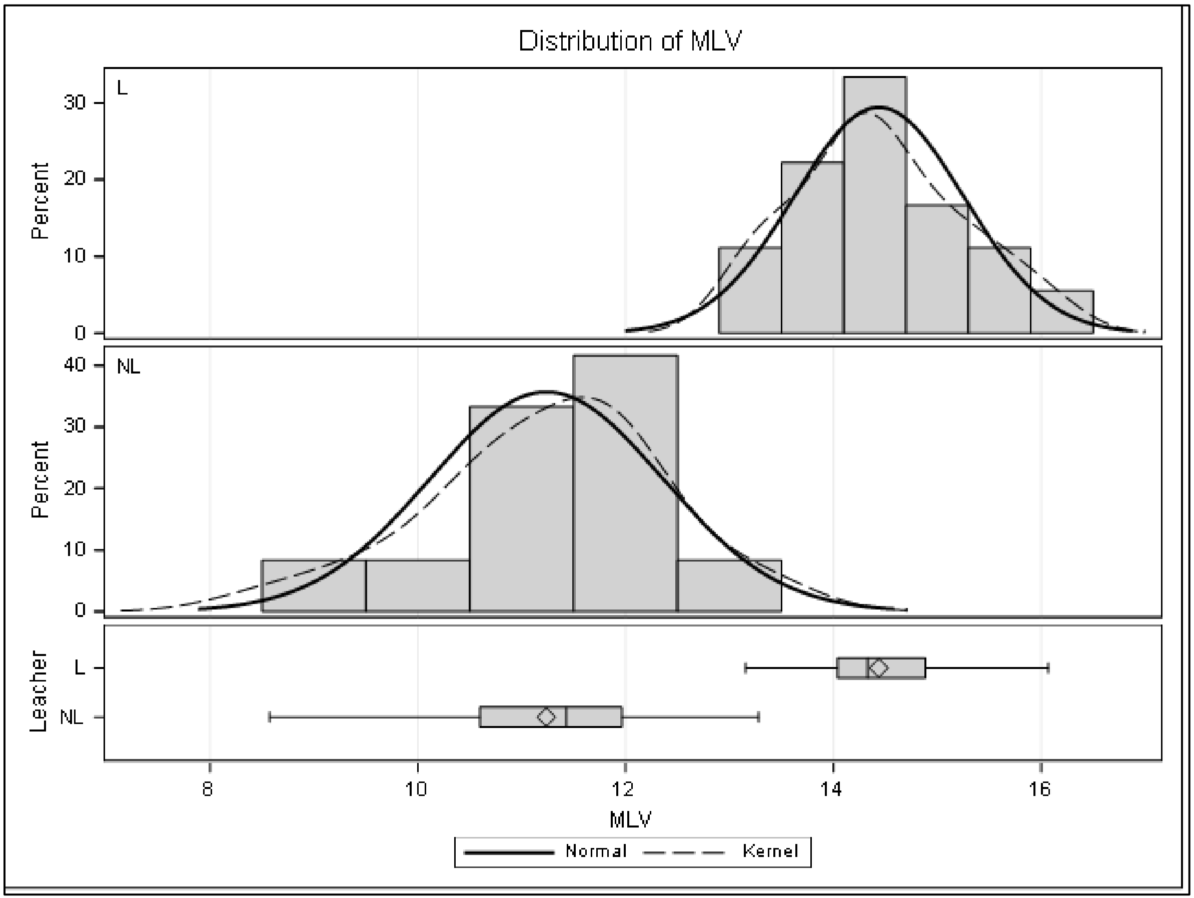
Ten-Parameter Model		
Variable	**Standardized Canonical Coefficients**	**Raw Canonical Coefficients**
LogK_{oc}	-0.9410	-1.2669
LogAero	1.2767	1.9971
LogK_{h}	0.2226	0.0698
PolarArea100*	0.0611	0.0035
Dipole	0.4299	0.1862
E_{HOMO}	-0.4104	-0.6419
LogMaxElPot	0.3718	2.2687
MinLocIonPot	0.0618	0.0164
Hardness	0.0709	0.1385
SurfArea	-0.1463	-0.0025

Table IV-3. Distributions of multivariate leaching values (MLVs) with associated fits of the normal and kernel distribution for the leacher (L) and non-leacher (NL) chemical groups of the nine-parameter canonical discriminant analysis model. Box plots indicate the mean, median, interquartile range, and minimum and maximum MLVs of each chemical group. Numerical statistics indicate standardized and raw canonical coefficients. An asterisk (*) denotes the smallest absolute standardized coefficient identified in the model to be removed during the next model-parameter reduction process.



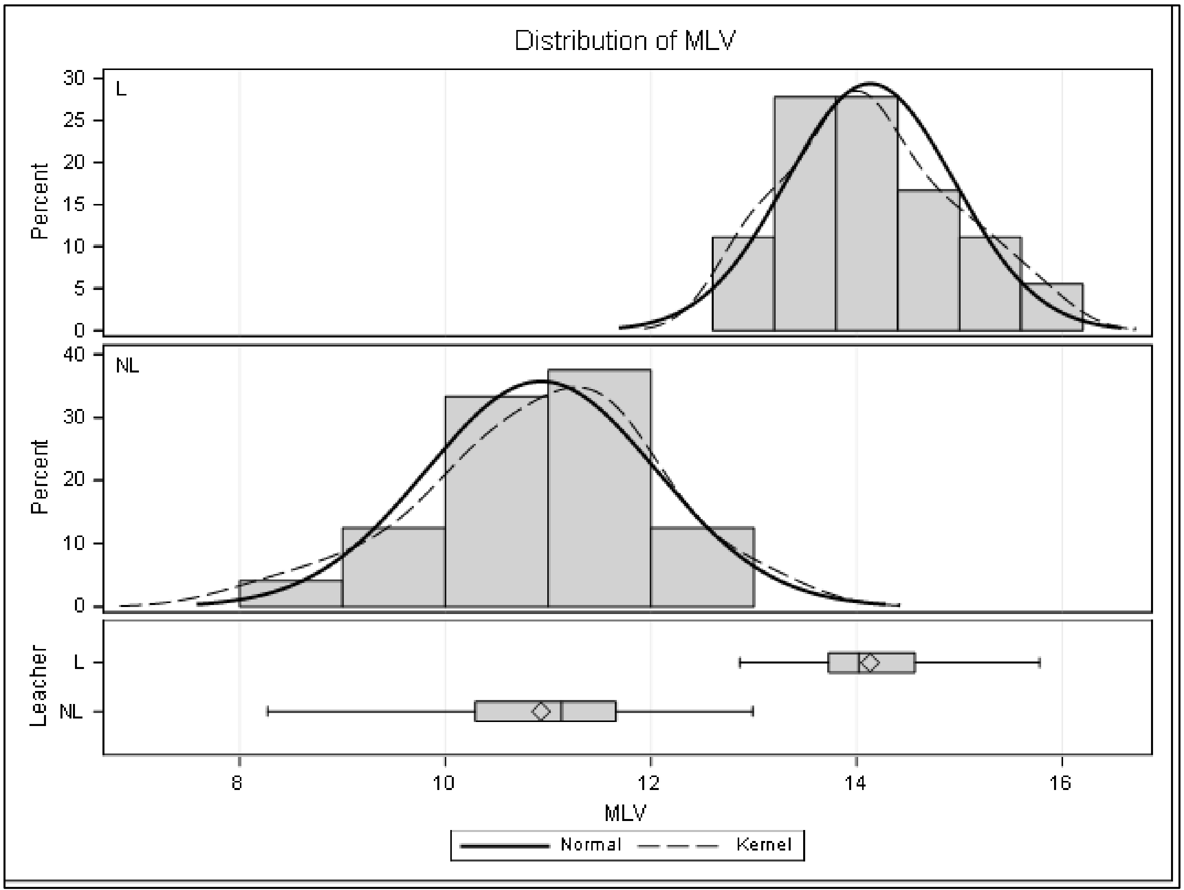
Nine-Parameter Model		
Variable	**Standardized Canonical Coefficients**	**Raw Canonical Coefficients**
$\mathrm{LogK_{oc}}$	-0.9360	-1.2600
LogAero	1.2789	2.0004
$\mathrm{LogK_h}$	0.2014	0.0631
Dipole	0.4547	0.1969
$\mathrm{E_{HOMO}}$	-0.4327	-0.6768
LogMaxElPot	0.3934	2.4004
MinLocIonPot*	0.0553	0.0147
Hardness	0.0600	0.1171
SurfArea	-0.1359	-0.0023

Table IV-4. Distributions of multivariate leaching values (MLVs) with associated fits of the normal and kernel distribution for the leacher (L) and non-leacher (NL) chemical groups of the eight-parameter CDA model. Box plots indicate the mean, median, interquartile range, and minimum and maximum MLVs of each chemical group. Numerical statistics indicate standardized and raw canonical coefficients. An asterisk (*) denotes the smallest absolute standardized coefficient identified in model to be removed during the next model-parameter reduction process.



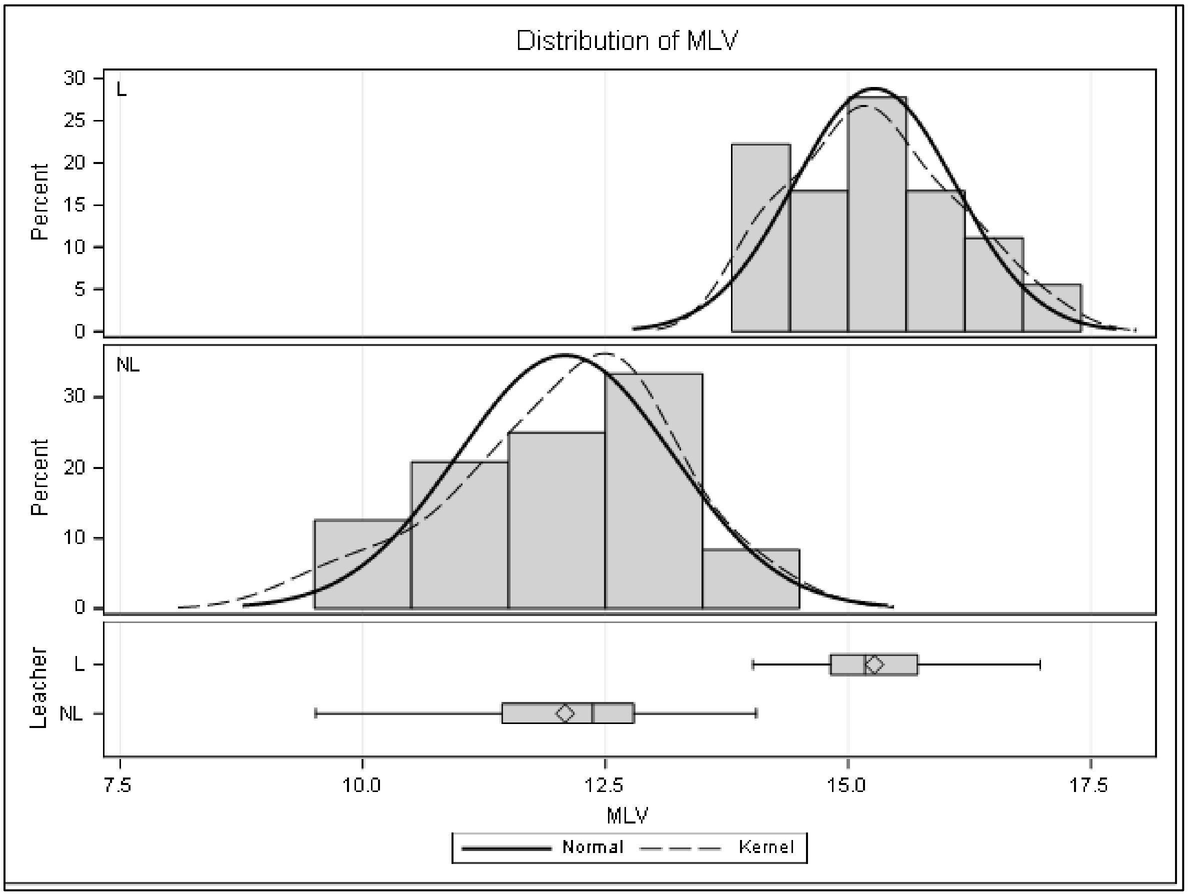
Eight-Parameter Model		
Variable	**Standardized Canonical Coefficients**	**Raw Canonical Coefficients**
LogK$_{oc}$	-0.9347	-1.2583
LogAero	1.2881	2.0149
LogK$_{h}$	0.1984	0.0622
Dipole	0.4509	0.1953
E$_{HOMO}$	-0.4726	-0.7392
LogMaxElPot	0.4057	2.4754
Hardness*	0.0240	0.0469
SurfArea	-0.1385	-0.0024

Table IV-5. Distributions of multivariate leaching values (MLVs) with associated fits of the normal and kernel distribution for the leacher (L) and non-leacher (NL) chemical groups of the seven-parameter CDA model. Box plots indicate the mean, median, interquartile range, and minimum and maximum MLVs of each chemical group. Numerical statistics indicate standardized and raw canonical coefficients. An asterisk (*) denotes the smallest absolute standardized coefficient identified in model to be removed during the next model-parameter reduction process.



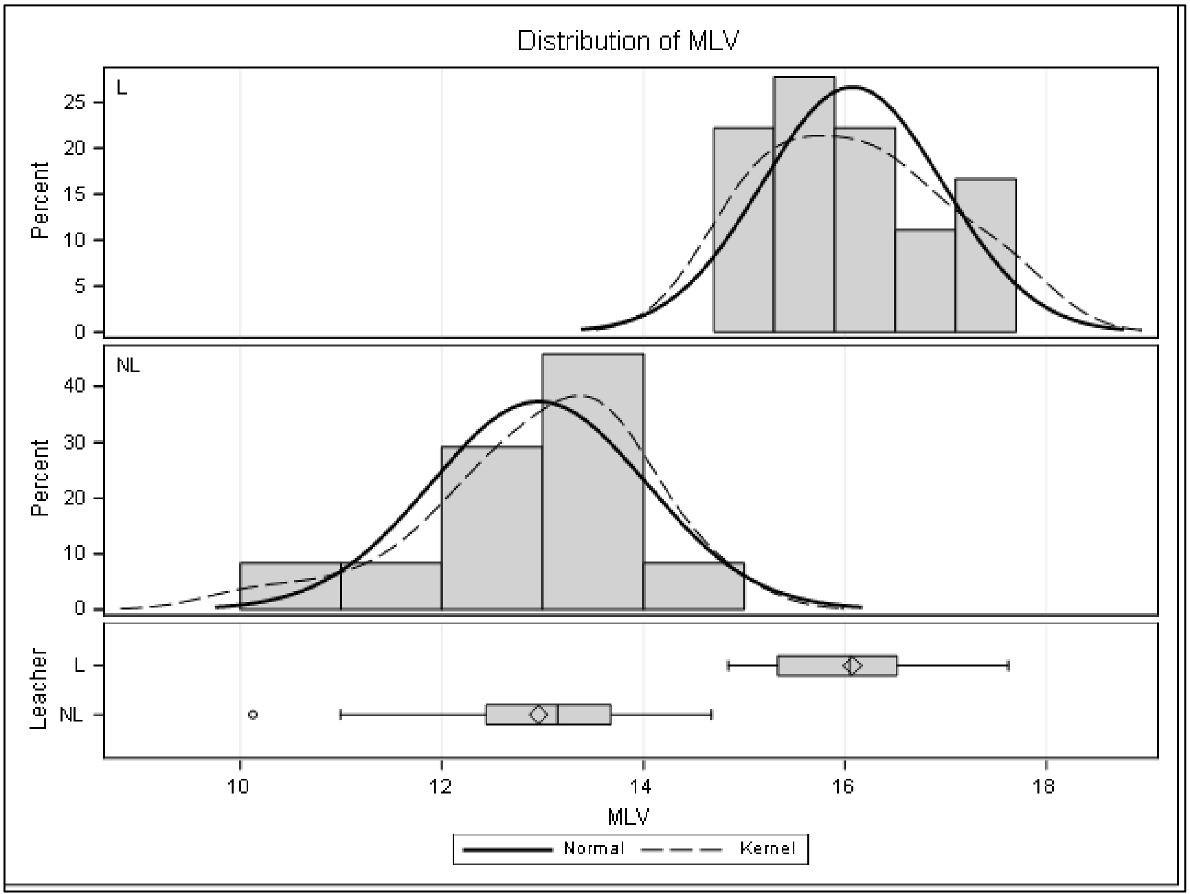
Seven-Parameter Model		
Variable	**Standardized Canonical Coefficients**	**Raw Canonical Coefficients**
LogK_{oc}	-0.9465	-1.2742
LogAero	1.2864	2.0123
LogK_{h}	0.2026	0.0635
Dipole	0.4545	0.1968
E_{HOMO}	-0.4800	-0.7508
LogMaxElPot	0.3975	2.4254
SurfArea*	-0.1441	-0.0025

Table IV-6. Distributions of multivariate leaching values (MLVs) with associated fits of the normal and kernel distribution for the leacher (L) and non-leacher (NL) chemical groups of the six-parameter CDA model. Box plots indicate the mean, median, interquartile range, and minimum and maximum MLVs of each chemical group. Numerical statistics indicate standardized and raw canonical coefficients. An asterisk (*) denotes the smallest absolute standardized coefficient identified in model to be removed during the next model-parameter reduction process.



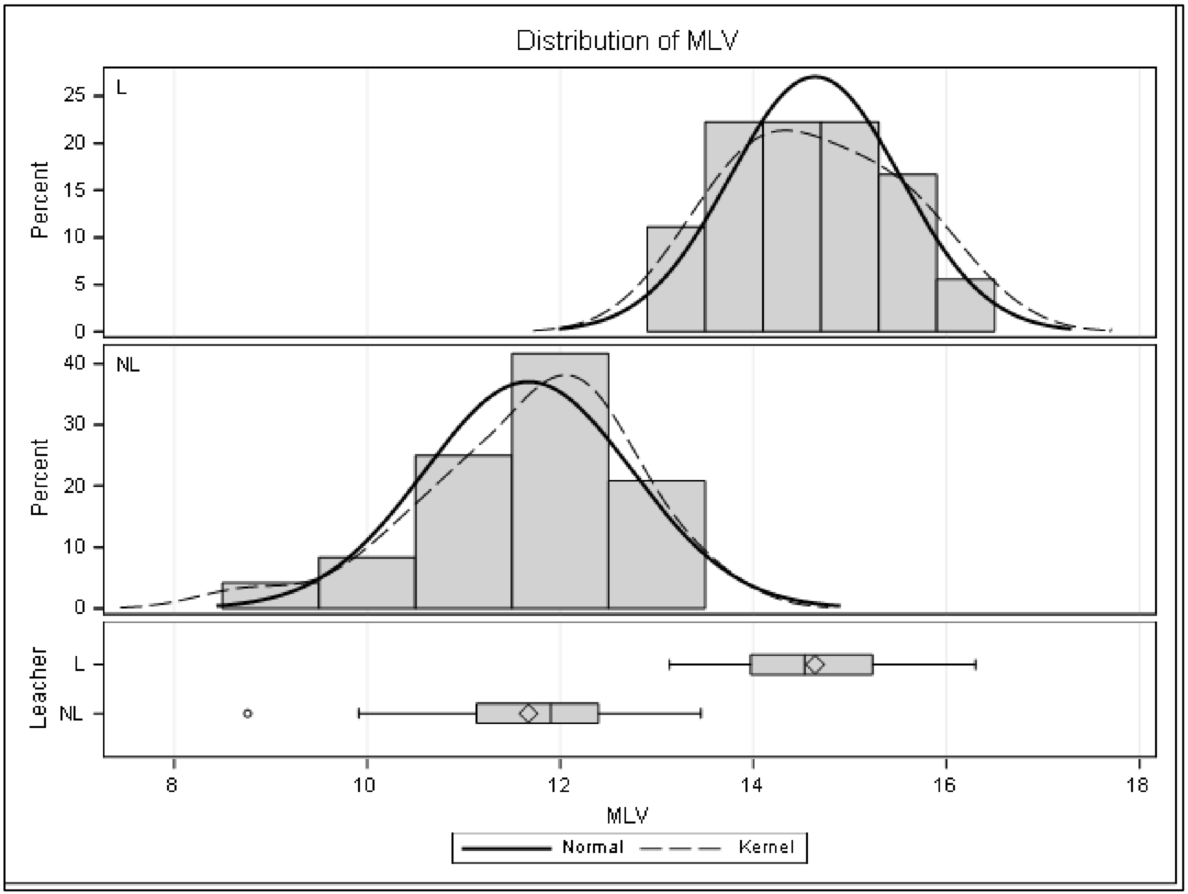
Six-Parameter Model		
Variable	**Standardized Canonical Coefficients**	**Raw Canonical Coefficients**
$\log K_{oc}$	-1.0168	-1.3688
LogAero	1.2962	2.0276
$\log K_h$*	0.2854	0.0894
Dipole	0.4563	0.1976
E_{HOMO}	-0.5246	-0.8206
LogMaxElPot	0.4223	2.5767

Table IV-7. Distributions of multivariate leaching values (MLVs) with associated fits of the normal and kernel distribution for the leacher (L) and non-leacher (NL) chemical groups of the five-parameter CDA model. Box plots indicate the mean, median, interquartile range, and minimum and maximum MLVs of each chemical group. Numerical statistics indicate standardized and raw canonical coefficients. An asterisk (*) denotes the smallest absolute standardized coefficient identified in model to be removed during the next model-parameter reduction process.



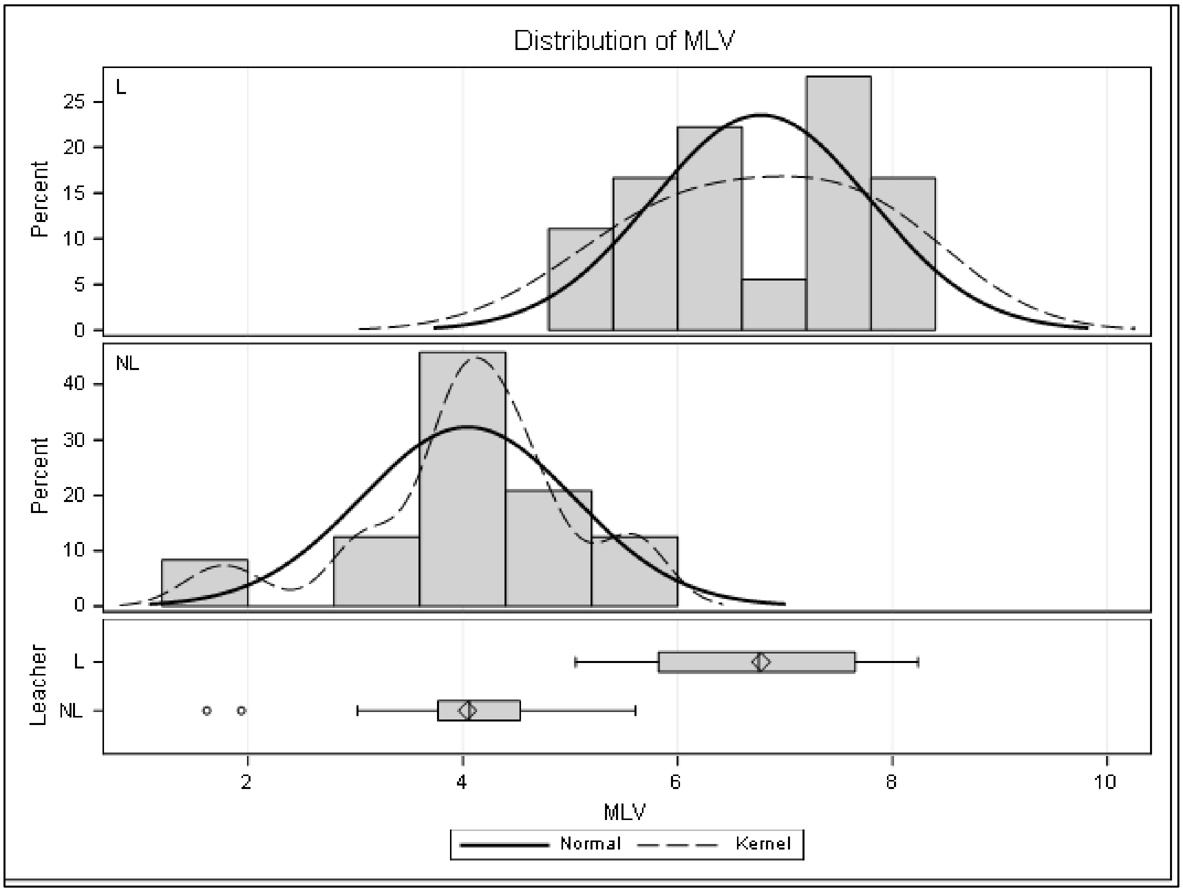
Five-Parameter Model		
Variable	**Standardized Canonical Coefficients**	**Raw Canonical Coefficients**
$\mathrm{LogK_{oc}}$	-0.9060	-1.2197
LogAero	1.2370	1.9349
Dipole*	0.3607	0.1562
$\mathrm{E_{HOMO}}$	-0.5843	-0.9140
LogMaxElPot	0.3725	2.2728

Table IV-8. Distributions of multivariate leaching values (MLVs) with associated fits of the normal and kernel distribution for the leacher (L) and non-leacher (NL) chemical groups of the four-parameter CDA model. Box plots indicate the mean, median, interquartile range, and minimum and maximum MLVs of each chemical group. Numerical statistics indicate standardized and raw canonical coefficients. An asterisk (*) denotes the smallest absolute standardized coefficient identified in model to be removed during the next model-parameter reduction process.



Four-Parameter Model		
Variable	**Standardized Canonical Coefficients**	**Raw Canonical Coefficients**
$LogK_{oc}$	-0.9186	-1.2366
LogAero	1.3138	2.0551
E_{HOMO}	-0.4531	-0.7087
LogMaxElPot*	0.4375	2.6691

Table IV-9. Distributions of multivariate leaching values (MLVs) with associated fits of the normal and kernel distribution for the leacher (L) and non-leacher (NL) chemical groups of the three-parameter CDA model. Box plots indicate the mean, median, interquartile range, and minimum and maximum MLVs of each chemical group. Numerical statistics indicate standardized and raw canonical coefficients.



Three-Parameter Model		
Variable	**Standardized Canonical Coefficients**	**Raw Canonical Coefficients**
LogK$_{oc}$	-1.0137	-1.3647
LogAero	1.3374	2.0920
E_{HOMO}	-0.3343	-0.5229

Appendix V — Theoretical Basis for Relating Quantum Variables to Reflect the Relative Potential of a Chemical's Soil Mobility

Overview

When pesticides enter the soil environment, their persistence and *mobility* depend strongly on the physicochemical properties of the *pesticide*. Biotic (microbial) and abiotic (chemical hydrolysis, photolysis, etc.) degradation, solubility, sorption, and volatility are all major property-based mechanisms which contribute significantly to their overall residence times and leaching potential (Pal et al., 2006). If a pesticide is soluble and chemically stable it may move with water percolating downward through the soil, eventually reaching groundwater. If, however, the pesticide is insoluble and/or tightly bound to soil particles, then it is more likely to be retained in the upper soil layers and small amounts may be lost to surface waters through runoff or erosion. The physicochemical properties of pesticides that affect mobility are significant contributors to their leaching potential in the soil. Sorption to soil minerals (clay minerals, iron oxides, etc.), humic substances, and organic matter impedes pesticide movement in the soil and increases the time available for degradation by microorganisms (Kah et al., 2007). If a pesticide is not degraded, volatilized, or retained by the solid phase (solid soil particles having varied composition and sizes) or suspended soil components, it will likely be mobile and a potential groundwater contaminant. In California, the leaching potential of new pesticide active ingredients is evaluated by the California Department of Pesticide Regulation (DPR) prior to registration. To better appraise the leaching risks associated with pesticides, a leaching model has been developed based on the notion that (1) the physicochemical properties of pesticides largely determine their mobility and persistence in soil, (2) their molecular structure must encode the features responsible for these properties, and (3) it is possible to represent these features mathematically:

$$\text{Leaching} = f\,(\text{Structure}) \qquad (1)$$

Equation 1 represents the central axiom of Quantitative Structure-Activity Relationship (QSAR) modeling, which asserts that the physical and chemical properties of molecules are a function of their molecular structures (Organization for Economic Cooperation and Development, 2013). However, while QSAR formalism was utilized in this current modeling effort, the objective was not to develop a classical QSAR model in which structural descriptors (i.e., independent or predictor variables) are utilized to directly estimate a particular endpoint (e.g., toxicity or half-life). Rather, this work was directed toward generating a binary classification model for discriminating between pesticides that have a propensity to leach to groundwater and those that do not. Such a model has the form:

$$\Phi = f\,(\alpha_1\chi_1, \alpha_2\,\chi_2, \alpha_3\,\chi_3 \ldots \alpha_n\chi_n) \qquad (2)$$

where χ_n are descriptors representing particular properties, obtained empirically or calculated from the electronic wave function of the molecule, and α_n are coefficients acquired from Canonical Discriminant Analysis (CDA). The discriminant function Φ is a chemical-specific value derived from the application of the CDA coefficients. It was anticipated that leaching and non-leaching chemicals might be classified based on a combination of quantum chemical properties derived from their molecular wave functions and experimentally-measured empirical properties. Quantum chemical properties were particularly of interest as, unlike experimentally-measured quantities, there is no inherent measurement error associated with such calculations (Karelson et al., 1996). Quantum chemical descriptors include such properties as molecular surface areas, orbital energies, electrostatic charges, electron densities, dipole moments, and ionization potentials which were calculated using Spartan computational software (Hehre and Ohlinger, 2016).

QSAR property calculations were conducted for each of the 42 chemicals used for model development. Each chemical's molecular structure was represented as a PubChem *.sdx* file and subsequently imported into the Spartan program for computational analysis. An initial pool of 42 topological, 22 quantum chemical, and 7 empirical chemical properties were considered for the development of a leaching classification model. Topological, QSAR, and empirical properties with raw or transformed values which did not reflect normal distributions, generally because they generated constant or zero values or were overtly extraneous were removed. For each of the remaining chemical properties, Student's *t*-tests performed between the leacher and non-leacher chemical groups that failed to produce a significant result, indicating no potential for CDA discrimination, were also removed. Finally, Pearson correlation analysis was employed on the remaining chemical properties to identify groups of highly correlated properties (i.e., r > 70%). One representative property within each correlated group was retained for CDA while the remaining properties were considered redundant and removed. Statistical discrimination between the model development set of known leacher and non-leacher chemicals (42 in total) was subsequently conducted using CDA. In CDA, a multivariate analysis of variance was conducted to determine if the retained chemical properties resulted in a statistically significant separation of the leacher and non-leacher groups. QSAR and empirical properties in the initial model were those that indicated distribution normality, potential for discrimination between leacher and non-leacher chemicals as indicated by the *t*-test, and that were not highly correlated with one another. The initial 11 chemical properties selected for the first CDA analysis provided a model with a low misclassification rate of approximately 5%. From the initial 11-parameter model, chemical properties were sequentially removed from the analysis whereby the property with the lowest absolute standardized coefficient was omitted. This procedure generated a definitive five-variable model having a misclassification rate also of 4.8%. The model was validated using leave-one-out (LOO) and leave-pair-out (LPO) cross-validation methodologies. The overall misclassification rate from this procedure was the same as for the full 11-parameter model. The five-parameter model included two empirical and three quantum chemical properties. An endpoint value or Multivariate Leaching Value (MLV) was calculated for each member of the model development set of chemicals from application of raw canonical coefficients (α_1, α_2, α_3, …n) to the chemical properties (X_1, X_2, X_3, ...X_n) according to Equation 2:

$$\begin{aligned}\text{MLV} = \;& 1.9349\,(\text{Log soil aerobic half-life}) - 1.2197\,(\text{Log } K_{OC}) + 0.1562\,(\text{dipole moment}) \\ & - 0.9140\,(E_{HOMO}) + 2.2728\,(\text{Log MaxElPot})\end{aligned} \qquad (3)$$

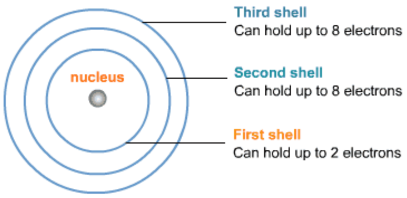
Applying Equation 3, an MLV for a new chemical with unknown leaching potential can be generated. The lower prediction limit at the 95 percent confidence level of the distribution of MLV scores for the leacher group of chemicals was used to establish a Multivariate Leaching Threshold (MLT) value of 14.4706. If the resulting MLV is equal to or above the MLT (14.4706), it is assigned membership into the leacher group of chemicals, otherwise the chemical is designated a member of the non-leacher group [i.e., leacher if MLV ≥ MLT (14.4706), otherwise non-leacher].

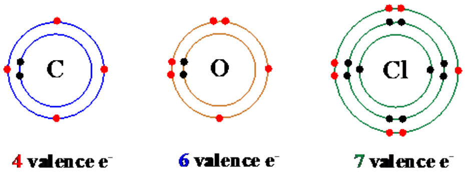
I. Leaching: A Chemistry Perspective

Intermolecular Forces

The physicochemical mechanisms that drive molecules to leach are the manifestation of the forces that act between them, i.e., they are a result of the interplay between various *intramolecular* and *intermolecular* forces (Figure V-1). These forces govern how molecules interact with their environment and each other and are inexorably associated. These are *real* forces that can be substantiated and enumerated experimentally and computationally. That the intermolecular forces are the fundamental source of the mechanisms that mediate leaching may not be immediately apparent, as such forces are generally imperceptible. However, the physicochemical properties that propagate from them, such as solubility or volatility, are conspicuous and familiar. Understanding and appreciating the role that these forces play in the leaching process requires that they be examined at the molecular level.

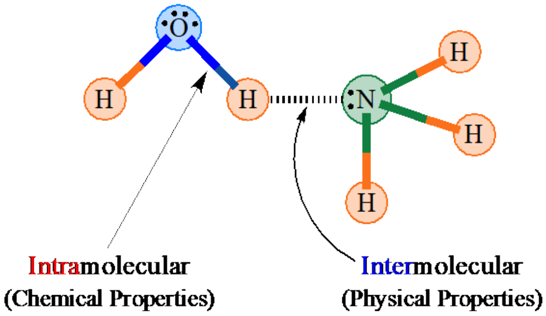
The electron density surface of a molecule is the outermost boundary that the molecule occupies in space. It defines its shape, volume, and size, but equally important it encapsulates all the physical and chemical characteristics that makes it a distinctive structure. All bonding and nonbonding interactions occur with the *valence* (outer shell) electrons at these outlying surface regions. The valence electrons in atoms can be envisioned using atomic orbital diagrams, as shown below for carbon, oxygen, and chlorine:





In molecules, valence electrons are not confined to the vicinity of one or two atomic nuclei, but are *delocalized* (i.e., spread over the entire molecule). Chemical (bonding) interactions originate from regional reaction centers having distinct structural characteristics (e.g., partial charges, lone electron pairs, conjugation, etc.), but the comparatively weak interactions that are the source of the physical (nonbonding) properties of molecules may emanate either locally (e.g., dipole-dipole interactions such as hydrogen-bonding at a donor/acceptor site) or globally (e.g., dispersion interactions anywhere on the molecular surface). Intramolecular forces, the strong internal forces that act *within* a molecule to hold it together, are characterized by three types of bonds: covalent, ionic, and metallic (Figure V-2). Covalent bonds are those in which a pair, or pairs, of electrons is shared by two atoms. Ionic bonds are formed as a result of the electrostatic attraction between ions of opposing charge. Ions are formed when the electrons involved in ionic bonding are transferred from the less electronegative to the more electronegative atom. Metallic bonds result from the attraction between positive metal ions and surrounding delocalized electrons. Intermolecular interactions are attractive or repulsive forces acting *between* molecules and other particles (e.g., atoms or ions) and are weak relative to the intramolecular forces. Intermolecular forces are categorized into a number of types, all of which emanate from dipolar interactions (Figure V-3).

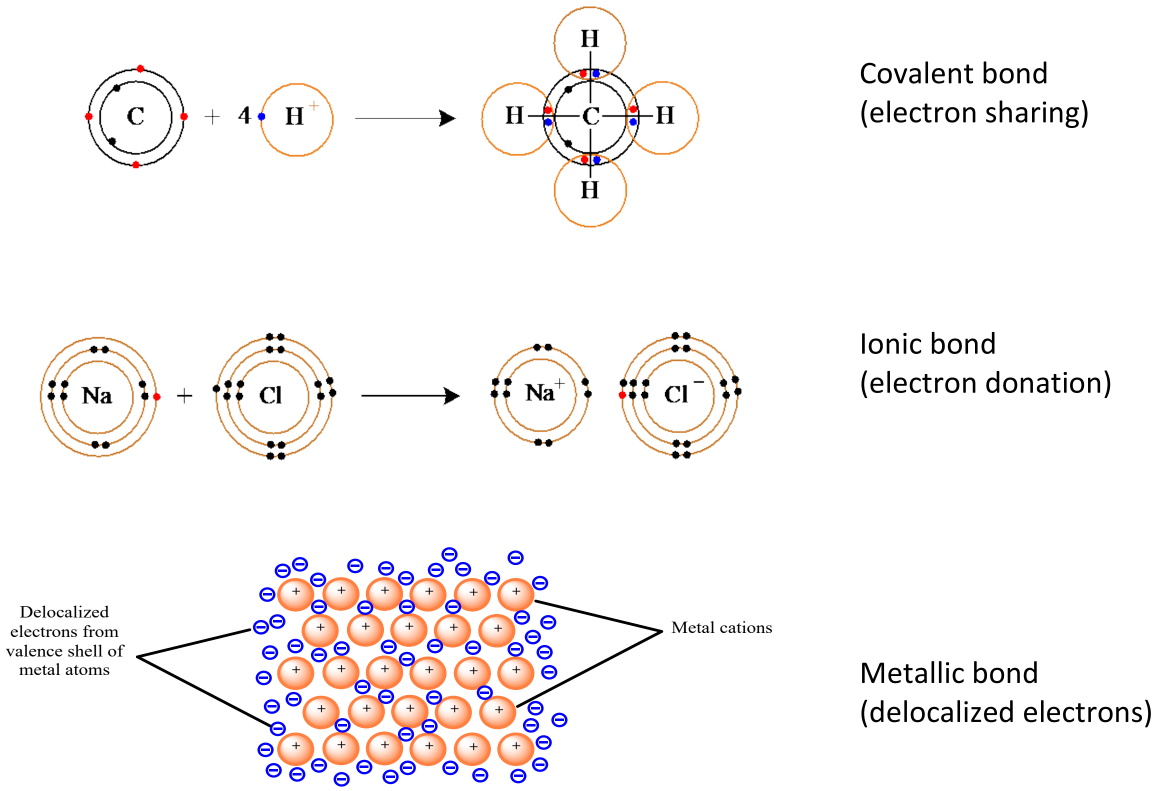
Figure V-1. Chemical and physical properties associated with intramolecular and intermolecular forces that are key influences on soil mobility.



STRONG Interactions:	WEAK Interactions:
Covalent bonding	Physisorption
Ionic bonding	(reversible binding)
Metallic bonding	Solubility
Chemisorption	Vapor pressure/Volatility
(irreversible binding)	Henry's Constant
	Absorption

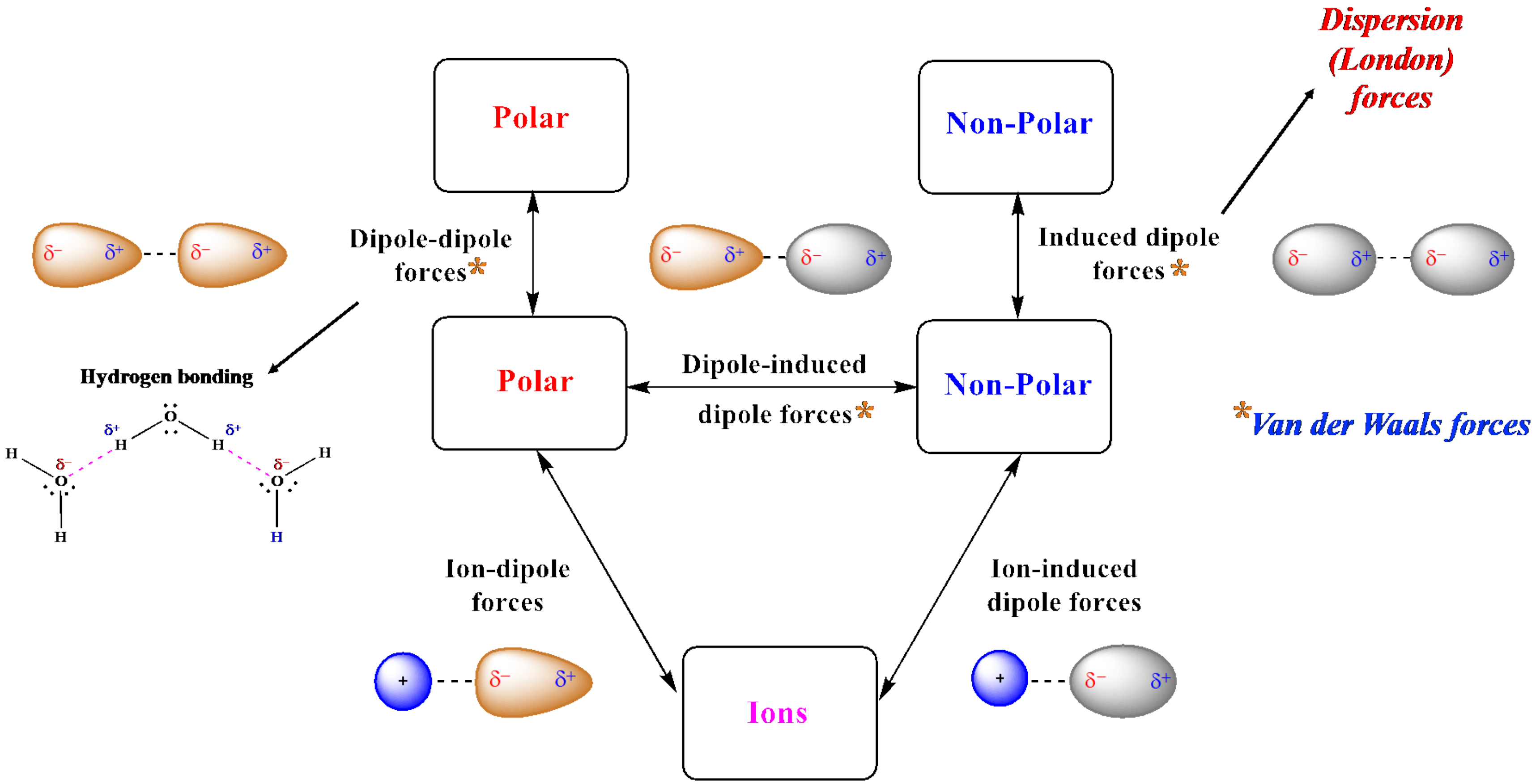
These types of interactions include dipole-dipole attractions in polar molecules (most notably the aforementioned hydrogen-bonding, the strongest of the intermolecular forces), induced dipole-dipole attractions, and induced dipole attractions such as London dispersion forces (the weakest of intermolecular forces). Although comparatively weak, these forces are pervasive in nature and are responsible for a variety of familiar bulk molecular properties, including the non-ideal behavior of gases (in contrast to that predicted by the ideal gas law), viscosity, diffusion, and the surface tension of liquids, vapor pressure and the boiling point of gases, and the melting point and sublimation of solids. Dipole-dipole interactions, induced dipole-dipole interactions, and dispersion forces are attractive forces known collectively as *van der Waals forces*, defined as the weak, short-range electrostatic attractive forces between uncharged molecules arising from the interaction of permanent or transient electric dipole moments. A fourth van der Waals force, known as the *hydrophobic effect*, is repulsive rather than attractive in character. The hydrophobic effect is observed with nonpolar molecules in aqueous solutions.

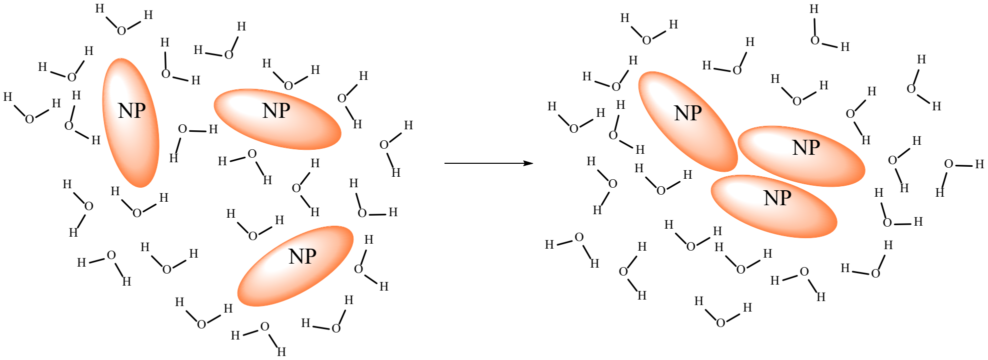
Figure V-2. Types of intramolecular bonding forces: covalent [sharing of electrons(s)], ionic [donation of electron(s)], and metallic [delocalized electron(s)].



In such solutions, water is excluded to the point that water molecules are forced to bond with each other rather than with a nonpolar molecule, or to the nonpolar portion of a molecule, leading to the formation of aggregates or, in *amphiphilic* (molecules containing both a nonpolar hydrophobic region and a polar hydrophilic region) compounds, micelles.

Figure V-3. Intermolecular forces: the weak attractive forces between molecules that drive the leaching process. Forces denoted by an asterisk (*) are categorized as *van der Waals forces*. Similarly, induced dipole interactions are known as *dispersion* or *London forces*.





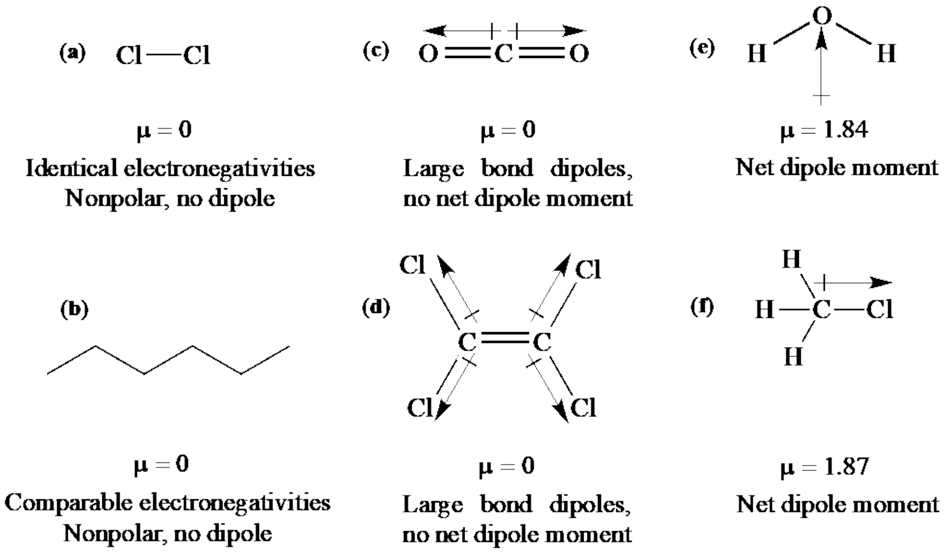
The hydrophobic effect is not a true interaction *per se* as the effect simply inhibits a true dipole-dipole interaction (i.e., hydrogen bonding by water) from taking place. The driving force behind the hydrophobic effect is entropy. When water solvates (dissolves) molecules, it is ordering itself (i.e., forming solvation cages around the molecules in solution). If the molecules are nonpolar or amphiphilic, they will cluster and exclude water. In doing so, less water will have to order itself, thus creating greater entropy (randomness), a state highly favored by nature. While not particularly strong, the hydrophobic effect can sometimes be as important a force as hydrogen bonding in the development of certain properties, e.g., *the adsorption* of nonpolar pesticides onto *hydrophobic* soil surfaces or in the formation of the tertiary structures (folding behavior) of proteins.

The physical properties of pesticides that contribute to soil mobility are all highly dependent on the various intermolecular forces discussed above. The degree to which such properties as solubility, sorption, and volatility drive the leaching process depends in large part on the type and magnitude of surface interactions that are established between a given pesticide and the soil environment. These surface interactions—the intermolecular and intramolecular forces—are in turn determined by molecular structure. As molecular structure varies, so do the attendant molecular surfaces and the properties originating from them. Consequently, the interrelationships between molecular structure, surfaces, and properties are inexorably linked. To a large degree the observable endpoints—the properties themselves—appear to emanate from polarity, orbital energies, and electrostatic charge distribution—all drawn from and unique to each pesticide chemical structure. These phenomena thus appear to be useful soil mobility indicators.

Physicochemical Properties: The Driving Mechanisms of Leaching

The leaching process is complex and involves a number of physical, chemical, and environmental mechanisms. Soil properties such as texture, pH, and organic matter content affect pesticide mobility, as do management practices (e.g., the rate, timing and method of pesticide application), rainfall and irrigation, and site conditions such as depth to groundwater. However, the physicochemical properties of individual pesticide active ingredients perhaps influence leaching potential more than any other factor. Despite the complicated network of interdependent and interconnected processes influencing pesticide mobility in soils, four physicochemical properties of pesticides consistently emerge as the dominant factors affecting their propensity to leach: solubility, volatility, sorption, and persistence. These properties determine the likelihood that a pesticide will persist in the environment, bind to soil and organic matter (sorption), dissolve in water (solubility), and become airborne (volatility). They are intrinsic molecular properties strongly dependent on chemical structure and, although mutable under inconsistent environmental conditions, pervasive. The solubility of pesticides in water, for example, depends entirely on the presence or absence of charge separation within the molecule, which produces favorable bond dipole orientations. When there are no polar bonds in a molecule, there is no separation of charge and the molecule is nonpolar. Chlorine gas (Cl_2) has no polar bonds because the electron charge is identical on both atoms, making it a nonpolar molecule (Figure V-4a). None of the bonds in hydrocarbon molecules, such as hexane, C_6H_{14}, are significantly polar, so hydrocarbons are nonpolar (Figure V-4b). Often, polar bonds may be present in a molecule, but the symmetrical arrangement of these bonds results in no net dipole as the dipoles cancel each other. This situation occurs in carbon dioxide (Figure V-4c) and perchloroethylene (Figure V-4d). Polarity occurs in a molecule as a result of a difference in electronegativity between the bonded atoms comprising its molecular structure. This arrangement results in a dipole moment, i.e., the vector sum of all its constituent bond dipoles, and an overall separation of charge in the molecule. In addition, the molecule must have a geometry which is asymmetric in at least one direction, so that the bond dipoles do not cancel or diminish each other. Water (Figure V-4e) and chloromethane (Figure V-4f) are examples of elementary polar molecules having dipole moments with favorable geometric symmetry.

Figure V-4. Dipoles and net dipole moments of (**a**) chlorine gas, (**b**) hexane, (**c**) carbon dioxide, (**d**) perchloroethylene, (**e**) water, and **(f)** chloromethane.



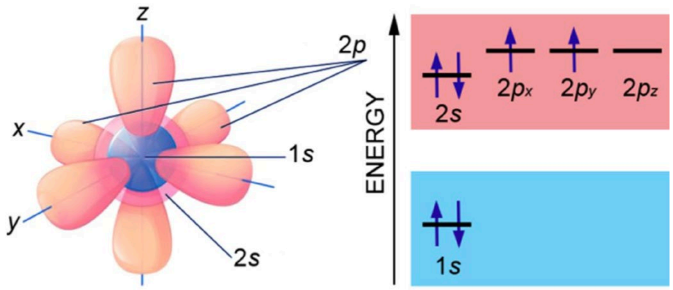
Molecular structure can strongly influence the susceptibility of pesticides to bind to soil particles, soil organic matter, humic substances, and the mineral and other hydrophobic fractions of soils. The adsorption process is generally classified as either physical adsorption (physisorption) or chemical adsorption (chemisorption). Weak interactions between substrate and adsorbate, characteristic of hydrogen bonding, induced dipolar interactions, and other van der Waals forces, lead to physisorption. Strong interactions, characteristic of covalent, ionic, and sometimes metallic bonding, lead to chemisorption. Adsorbed molecules remain structurally unchanged when physisorbed but may be cleaved into two or more fragments when chemisorbed. Chemisorption involves more energy than physisorption. The difference between the two processes is loosely based on the binding energy of the interaction. The presence of charge or a separation of charge in the molecule are key factors in determining the type and strength of its interaction with the soil surface. If no charge, or no favorable separation of charge, is present in the molecule, then it is neutral (or perhaps weakly polar) and will tend to be hydrophobic and associate with lipophilic structures, e.g., the hydrophobic components of humic macromolecules. Conversely, increasing polarity enhances the aqueous solubility of the organic molecule so that adsorption is reduced. However, polar molecules can adsorb to soils, usually via a hydrogen-bonding mechanism onto clay minerals (physisorption) or through the formation of *coordination* complexes (chemisorption) (MacKay and Canterbury, 2005). Molecular size, shape, and surface charge have strong influences on adsorption processes, particularly when the adsorptive pesticide must penetrate through restricted openings to reach the binding sites in adsorbents. Large molecules tend to be more strongly held than smaller ones with similar shapes, polarities, and charge characteristics. Increased adsorption with increased molecular size is a well-established phenomenon described by Traub's rule, which states that adsorption onto nonpolar surfaces increases with molecular size and decreasing solubility (Karanfil, 2006). For pesticides having a molecular structure consisting of both hydrophilic (polar) and hydrophobic (non-polar) groups, the hydrophobic fragments will be adsorbed at the surface and the hydrophilic fragments will tend to stay in the water phase.

The volatility of a pesticide depends on its molecular structure and the strength of the intermolecular forces with which it interacts with its environment, as these must be overcome in order for it to escape to a gaseous state. For low molecular weight pesticides, the dispersion forces are very weak and can be easily overcome. Hence, nonpolar pesticides like methyl bromide are gases at room temperature. As the molecular weight increases, so do the dispersion forces. With pesticides having large relative molecular mass, the van der Waal's forces are sufficiently large to keep the compound from volatilizing even at higher temperatures. Since virtually all pesticides are covalent molecules, they are attracted to one another by either van der Waal's (induced dipole) forces or dipole-dipole interactions, or both. In addition, there are two types of dipole-dipole interactions, permanent dipole-dipole attractions and hydrogen bonding. The van der Waal's forces are always present between covalent molecules but are the weakest forces when the molecules are small. Permanent dipole-dipole forces are stronger than van der Waal's attractions but weaker than hydrogen bonding. London dispersion forces act between all molecules regardless of their structure. The London dispersion force is the weakest intermolecular force. It is a temporary attractive force that results when the electrons in two adjacent atoms occupy positions that result in induced dipole-induced dipole attractions. The strength of the dispersion force depends on the mass, surface area, and shape of the molecule. When comparing two compounds that have similar molecular structure, the shape of the molecule becomes important. Branched molecules are more spherical in shape and have a lower surface area. This means that there is less surface for the dispersion force to act, and the overall force between the molecules is smaller. Hence, greater branching reduces the intermolecular force and consequently the volatility. Pesticides with hydrogen-bonding capabilities have lower volatility because the molecule can actively interact and form weak bonds with the solvent (water), the soil itself, and other soil constituents. Pesticides with hydrogen bonding capabilities include those containing alcohol, carboxylic acid, amine, and amide functional groups. Hydrogen bonding in carboxylic acids is particularly strong, as hydrogen bonds can form with both the acidic hydrogen and the carboxylic oxygen. Hydrogen-bonding occurs when a molecule contains hydrogen attached to oxygen or nitrogen. These two elements are highly electronegative and draw the electrons away from the hydrogen atom. Because hydrogen has only a single electron, this inductive effect exposes the hydrogen nucleus, causing a high partial charge density, making the N-H and O-H bonds very polar. Accordingly, hydrogen-bonding is a relatively strong force (about one tenth of a normal covalent bond).

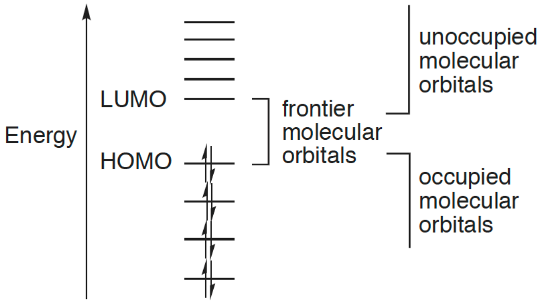
II. Quantum Chemical Descriptors: Fundamental Supporting Theory and Characterization

Fundamental particles such as electrons are defined as *matter wave*s (i.e., having wave-particle duality) and are described using a wave function, ψ. The surface of a molecule is delineated by its total electronic wave function. The value of ψ at a given three-dimensional point in space is proportional to the amplitude of the electron matter wave at that point. But because most wave functions are complex functions containing $i = \sqrt{-1}$, the amplitude of the matter wave has no real physical meaning. It is not measurable. However, the square of the wavefunction, ψ^2, is proportional to the *probability* of finding an electron in a particular volume of space within an atom or molecule. The function ψ^2 is called the *electron probability density* or, more informally, the *electron cloud*. Molecules are surrounded by the negatively charged electron cloud containing its electrons, and it is the size and shape of the cloud and *not* that of the nuclear skeleton that defines the size and shape of the molecule. The molecular surface can be defined as an *equal probability boundary surface* given by the wave function ψ^2, where there is a ≥ 95% chance of an electron occupying a point on the cloud at a given time (Gao, 2013).

An electron in an atom has a fixed energy and a fixed radial and angular spatial distribution called an *orbital*. The radial distribution defines the distance an electron is from the nucleus, the angular distribution describes the shape of the orbital, i.e., the shape of the electron cloud. An *s* orbital has a spherical shape because its angular distribution is uniform at every angle. A *p* orbital has a "dumbbell" shape because its angular distribution is not uniform at every angle. Similarly, a *d* orbital has an even more complex angular distribution and a cloverleaf shape. Collectively, the number and type of orbitals in an atom define its electronic structure. When atoms combine to form a molecule, the electronic structure of the resultant molecule is different than that of its individual constituent atoms. The electrons in atoms are *localized*—the electron density is confined to the atomic orbitals (e.g., *s, p, d*) to which they belong, as shown for a carbon atom in the diagram below. The two valence electrons in carbon are confined to the $2p_x$ and $2p_z$ orbitals, respectively:

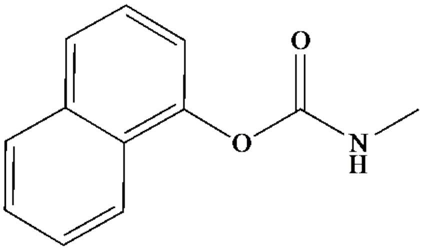


When a molecule is formed, however, the electrons become *delocalized*, i.e., the electron density is allowed to spread out so that the *molecular orbital* electrons extend over several adjacent atoms, or even the entire molecule. To exemplify this concept, consider two molecular orbitals that are of particular interest in a molecule, the *highest occupied molecular orbital* (HOMO) and the *lowest unoccupied molecular orbital* (LUMO). The HOMO is the highest energy *molecular orbital* containing electrons; therefore, energetically it is the easiest to remove electrons from this orbital, e.g., donating electron density to form a bond (act as a Lewis base), undergo oxidation, etc. The LUMO is the lowest lying orbital that is empty, so energetically it is the easiest to add more electrons into this orbital, e.g., accepting electron density to form a bond (act as a Lewis acid), undergo reduction, etc.



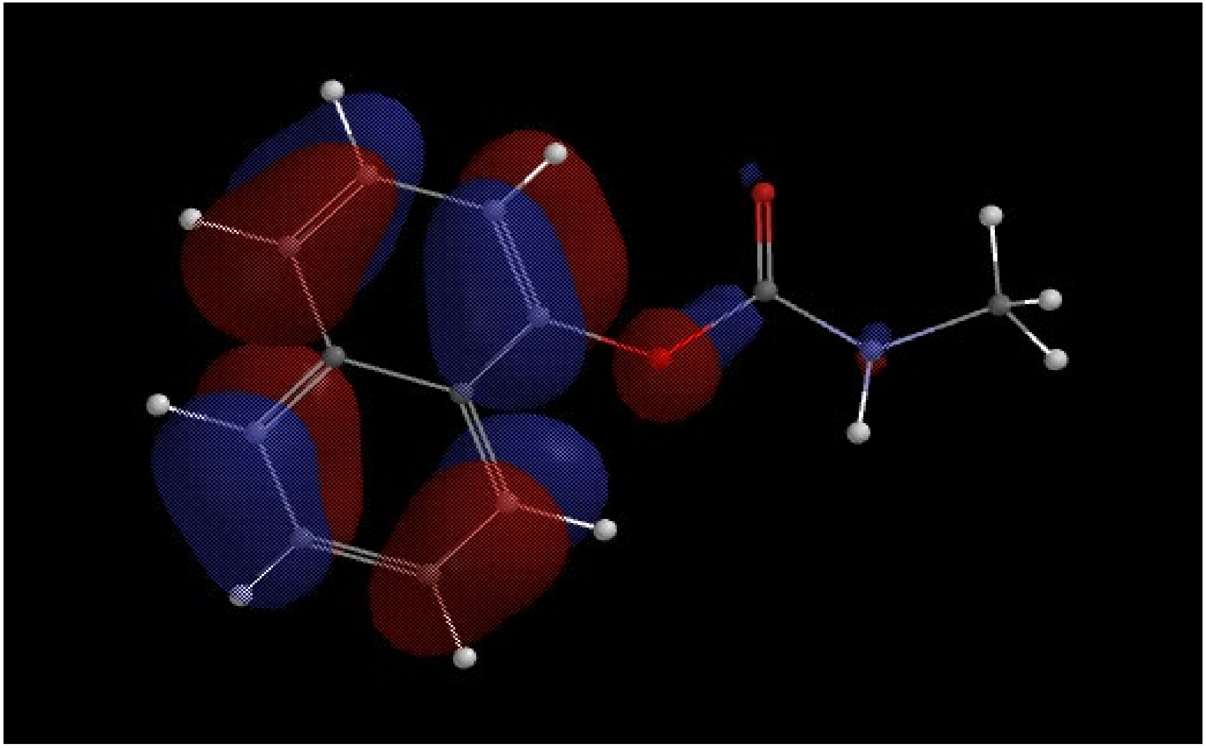
The HOMO and LUMO concept originate from Frontier Molecular Orbital (FMO) theory, which focuses on the orbitals at the outer edge (the frontier) of a molecule rather than all the orbitals. Outer-edge orbitals tend to be the most spatially delocalized and have the highest and lowest energies, whether they are occupied or unoccupied. The rationale for concentrating attention on these two orbitals is that they will usually be the closest in energy of the interacting orbitals. A basic postulate of FMO theory is that interactions are strongest between orbitals that are close in energy. Accordingly, reactions between two molecules typically involve the HOMO of one and the LUMO of the other. FMO theory proposes that these strong initial interactions can then guide the course of a reaction or interaction. Electronic properties based on frontier orbital energies are closely related to chemical and biological reactivity.

Since envisioning molecular orbitals, in particular the HOMO, can be conceptually challenging, it's instructive to have a visual depiction. Consider the carbamate insecticide carbaryl, a member of the model development set and a non-leacher:



Carbaryl (1-naphthyl methylcarbamate)

The diagram below shows the HOMO for carbaryl derived from its wavefunction:

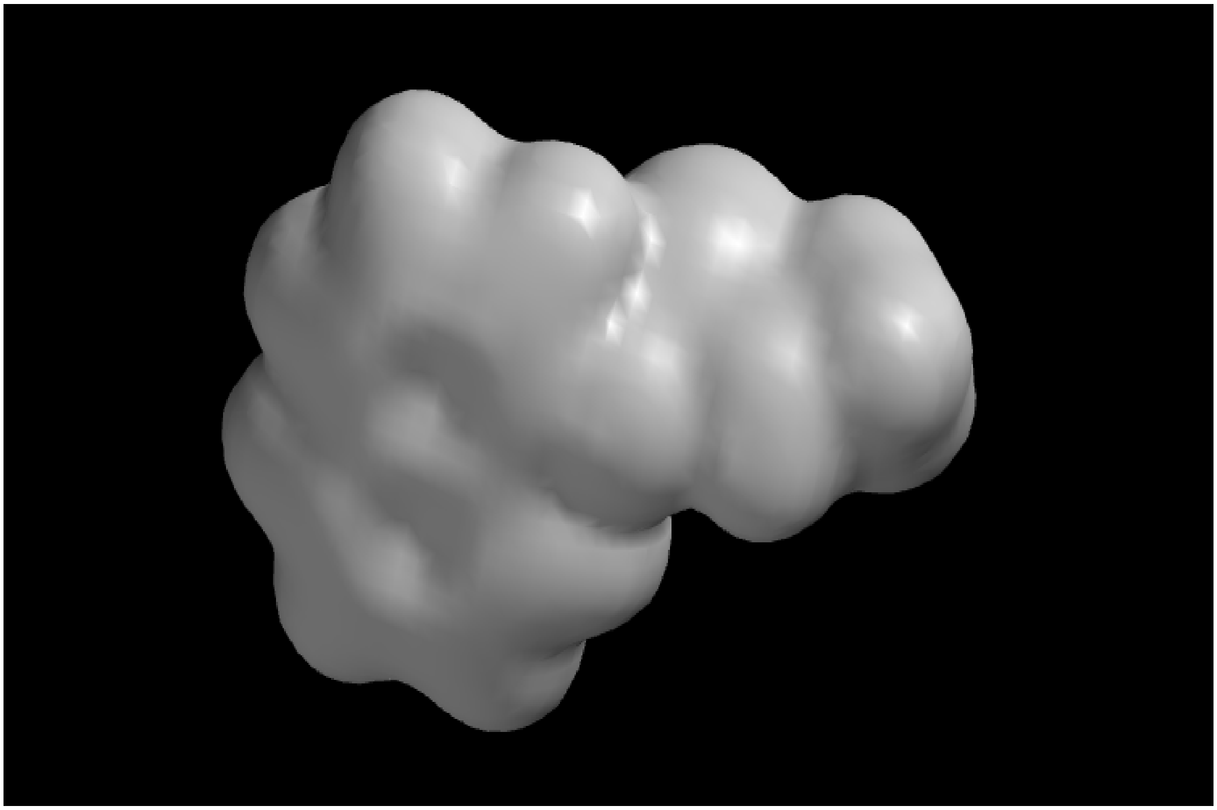


The red (negative) and blue (positive) lobes represent the electron clouds indicating where the electrons are likely to be found. Note that the highest energy electrons in the carbaryl system are those that are spread over the naphthalene ring system and extending to the carbamate oxygen moiety. The HOMO of carbamate is thus a molecular orbital which puts significant electron density on the ring and on the oxygen bonded to it. These are electron rich or *nucleophilic* regions of the molecule, where a nucleophile is defined as a chemical species that donates an electron pair to form a chemical bond in a reaction (which is also the definition of a Lewis base). Hence, high-energy HOMO electrons are the most available electrons for interacting in chemical reactions or in other interactions involving electrons (e.g., ionic bonding, H-bonding, and other intermolecular forces).

HOMO energies have been shown to be related to some of the key mechanisms that characterize and influence the mobility of chemicals in the soil environment. Kamachi et al. (2019) have reported that HOMO energies and dipole moment are the first and second most important properties, respectively, for rationalizing and predicting the adsorption energies of small molecules on titanium dioxide (TiO_2) surfaces. HOMO energies have also been shown to have quantitative relationships with the Freundlich adsorption exponent 1/*n* (Oskouie et al., 2002) and organic carbon normalized adsorption coefficients, K_{OC} (Doucette, 2003). The National Research Council (2014) has noted that many physicochemical and environmental fate properties of molecules can be described from first principle based on HOMO and LUMO frontier orbital energies. In a comprehensive review by Mamy et al. (2015), 790 quantitative structure-activity relationship (QSAR) models for estimating the fate of organic compounds in the environment from their molecular properties were evaluated. All of the models were developed using only structural molecular descriptors. The most significant equations were found for pK_A, K_{OW}, adsorption to soils, certain biodegradation processes, and abiotic degradation in soils, water, and sediments. Five molecular descriptors—E_{HOMO}, E_{LUMO}, polarizability, dipole moment, and molecular weight—were in particular used in the 790 equations. The utility of quantum chemical descriptors for estimating biological, chemical, physicochemical, and solvent properties was also explored by Karelson et al. (1996). E_{HOMO} was one of the key molecular properties used in the numerous modeling studies summarized in this article, including models for estimating biological activities, chemical reactivities, partition coefficients, and physicochemical properties.

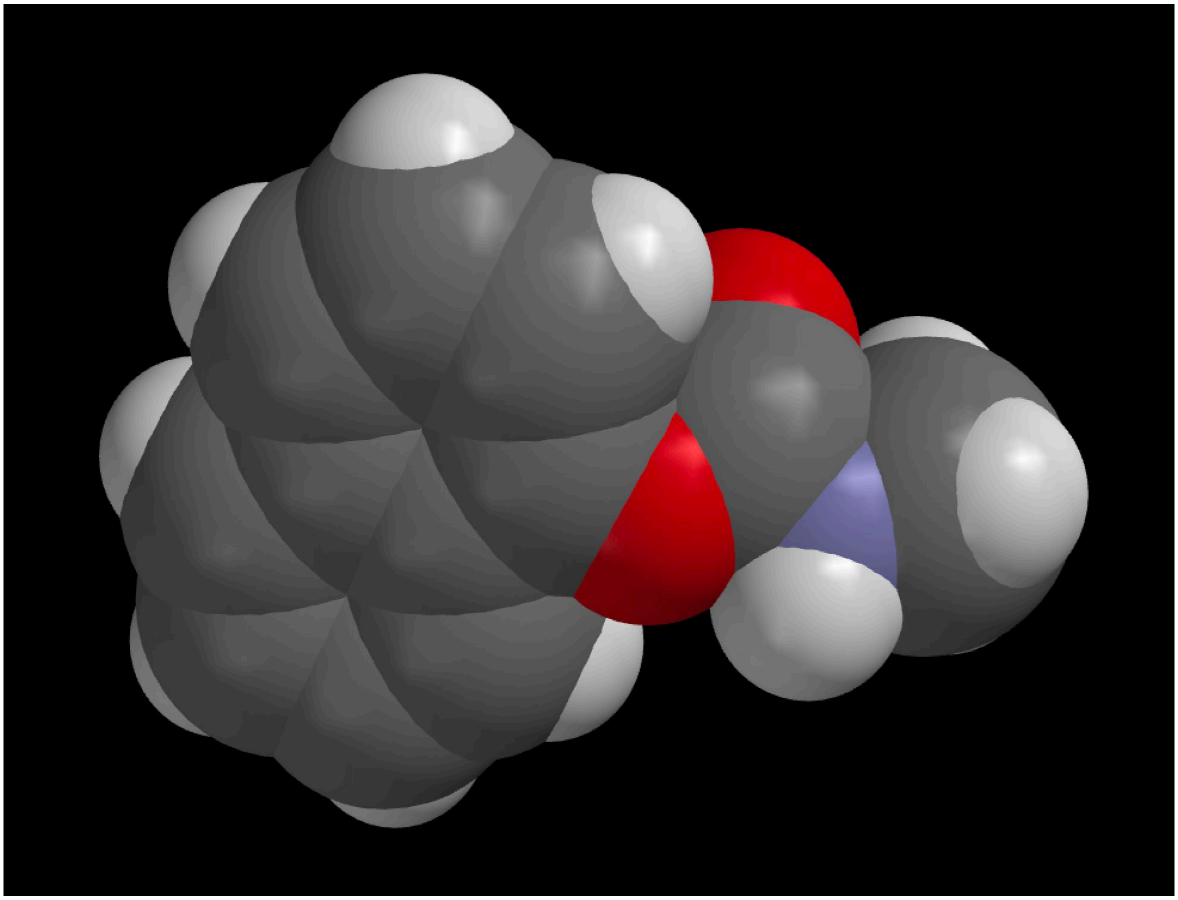
The size and shape of an electron cloud is described by the electron density, i.e., the number of electrons per unit volume. The Spartan program renders the surface area of a molecule from an *isodensity surface* (*iso* = equivalent, i.e., an isodensity surface represents points of a constant electron density), where the electron densities are obtained directly from quantum chemical calculations. The size and shape of the surface is determined by the percentage of total electrons enclosed, typically 0.002 electrons per cubic atomic unit. This represents a very low level of electron density and such points are typically found near the outermost fringe of the molecule's electron cloud. Therefore, these points approximate the maximum spatial boundary of the molecule, i.e., its size and shape. Interestingly, when electron density is plotted as a function of distance from the nucleus, it never falls to zero. Thus, when atoms and molecules brush up against each other, their electron clouds overlap and merge to a small extent. The image shown in Figure V-5 illustrates such an electron cloud for the molecular orbitals comprising the boundary surface of carbaryl.

Figure V-5. The molecular boundary surface (electron cloud) of the carbamate insecticide carbaryl. The molecular surface area and volume are calculated from this surface.



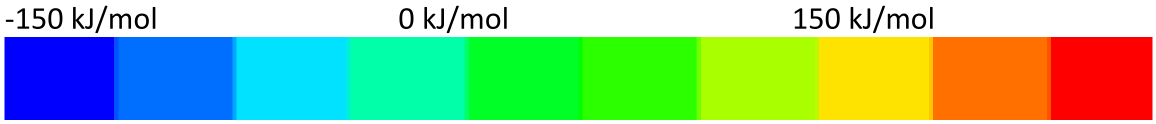
Molecular surface area is a measure of the total area of this isodensity surface in units of Å^2. Upon determining the isodensity points for a particular molecule, they are then mapped onto its geometry-optimized line angle structure (i.e., the molecular skeleton) to yield a 3-D molecular surface. Note that although the density-based surface shown in Figure V-5 does not provide discernible boundaries between atoms like the space-filling model does (Figure V-6), it is a more accurate representation of the molecule because in reality electrons are associated with the molecule as a whole and not with individual atoms (Spartan ’20 Tutorial and User's Guide).

Figure V-6. The space-filling model of carbaryl, where individual atoms are represented by spheres whose radii are proportional to the actual radii of the atoms and whose center-to-center distances are proportional to the distances between the atomic nuclei.



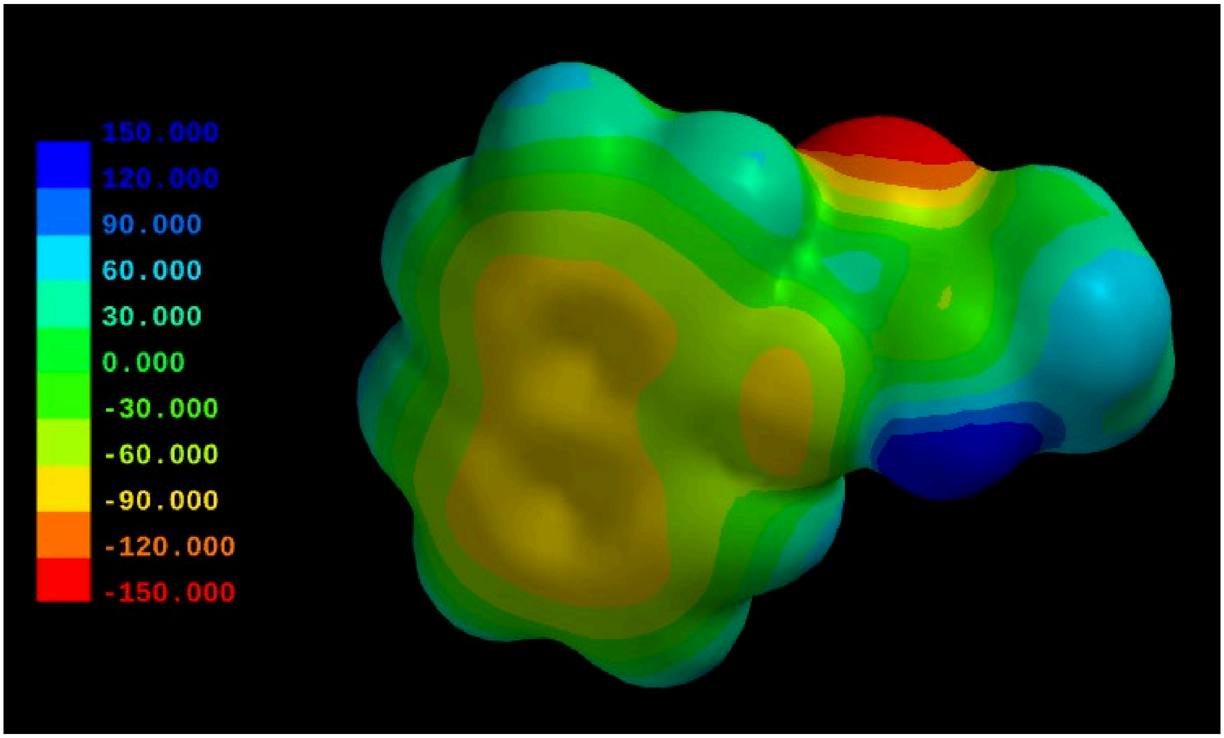
Once the isodensity surface of a molecule is known, an electrostatic potential map can be generated for it. These maps are useful diagrams for envisioning molecular charge distributions and for identifying polar and non-polar regions. The electrostatic potential at any point (x, y, z) on the 3-dimensional surface of the molecule is specified by the electrostatic potential energy between an imaginary positively charged (+1) ion located at that point and the molecule. If the ion is attracted to the molecule, then the potential is negative. If the ion is repelled by the molecule, then the potential is positive. Thus, the charged ion will be attracted to electron-rich regions (negative potentials) of the molecule and repelled by electron-poor regions (positive potentials). Spartan calculates the electrostatic potential at selected points on the 0.002 isodensity surface and maps the surface by colors representing different potentials. The highest negative potential is assigned a bright red color and the highest positive potential, a bright blue color.

Intermediate potentials are assigned colors according to the color spectrum shown below (using a probe sensitivity of ± 150 kJ):



The red and blue regions of an electrostatic potential map specify the most electron-rich (partially negative charged) and the most electron-poor (partially positive charged) regions of the molecule, respectively (Hehre, 2003). The intermediate colors orange, yellow, green, and light blue areas represent a range of nonpolar regions of the molecule (with green being the most nonpolar). The electrostatic potential map thus identifies and makes observable the charged and polar regions of the molecule and reveals the overall molecular charge distribution (Figure V-7).

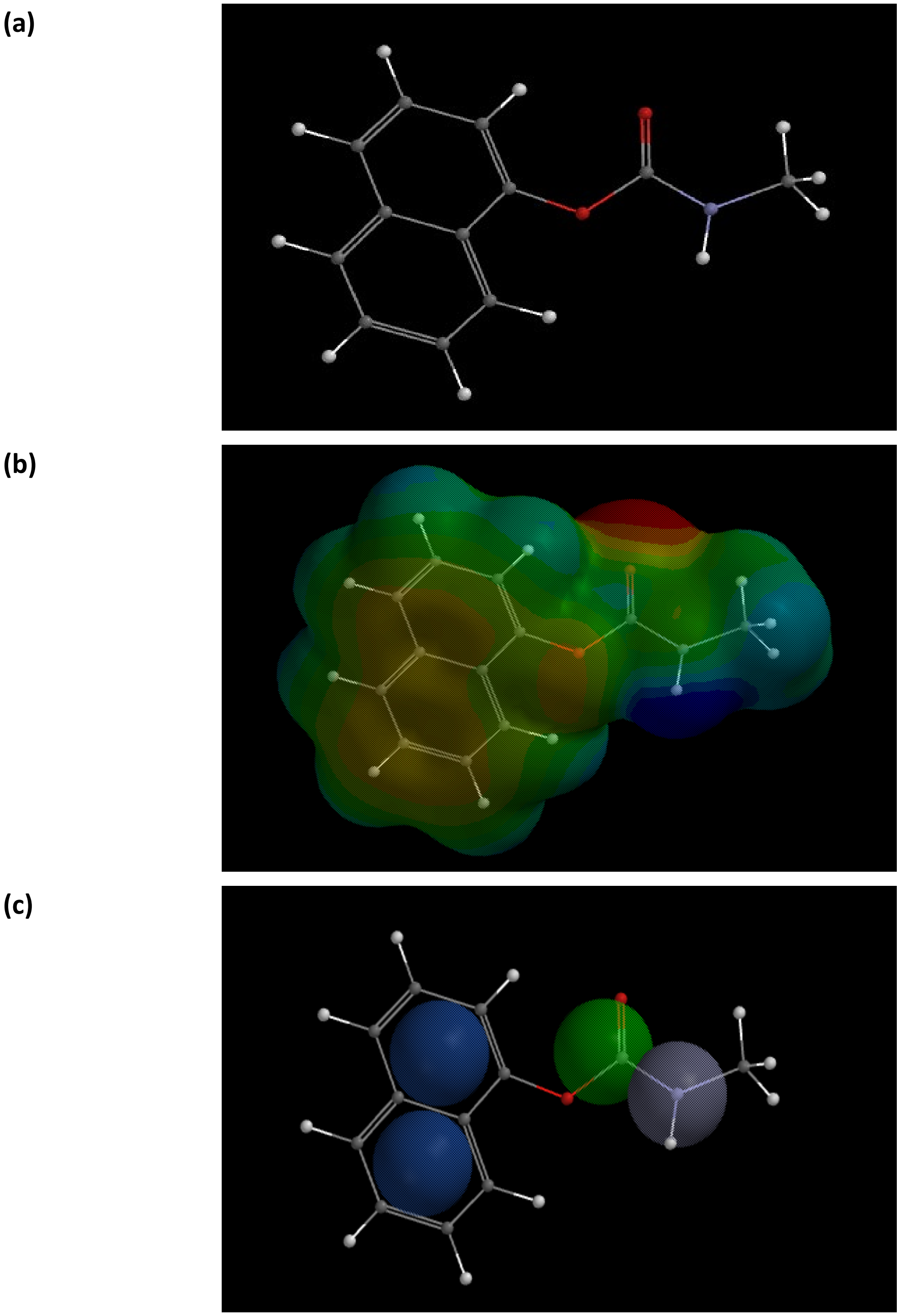
Figure V-7. The electrostatic potential mapped over the electron density of carbaryl. The vivid blue and red regions of the molecule represent the areas where the electrostatic potential is < -150 kJ/mol and > 150 kJ/mol, revealing areas of the molecule having the highest partial negative (blue) and positive (red) charges. These regions are also potentially good hydrogen bond acceptor and donor sites, respectively.



The Maximum Electrostatic Potential (MaxElPot) is defined as the point for which the absolute value of the electrostatic potential is ≥ maximum sensitivity (in this case, ± 150 kJ) and corresponds to the most acidic proton in the molecule. Hydrogen-bond acceptor and donor sites are also related to the electrostatic potential obtained from quantum chemical calculations. For example, a negative potential (red) associated with an oxygen center suggests that it might serve as a hydrogen-bond acceptor, while a positive potential (blue) associated with a hydrogen attached to an oxygen, nitrogen, or fluorine center suggests its role as a hydrogen-bond donor. Such information may be useful for predicting which acceptor or donor sites are likely to be strong and which are likely to be weak. If a molecule has a permanent dipole and subsequent dipole moment, it's a polar molecule. A dipole moment is the vector sum of all of the bond dipoles in the molecule. Since it is a vector quantity, a molecular dipole moment has both a magnitude and a direction. The accumulation of negative charge over an area in a molecule is measured in *debyes* (represented by the letter D), and its magnitude is dependent upon the size of the charge and the distance separating the point charges. Although a polar bond is a prerequisite for a molecule to have a dipole, not all molecules with polar bonds exhibit dipole moments due to unfavorable molecular symmetry or geometry.

In Figure V-8, the chemical structure of carbaryl is shown (**a**) relative to its electrostatic potential map (**b**) to identify regions of the molecule that may be suitable hydrogen bonding acceptor and/or donor sites, reactive sites, and hydrophilic/hydrophobic sites on the molecule. The bright blue area on the electrostatic potential map corresponds to the most acidic hydrogen (i.e., the N-methyl proton in carbaryl). But the most electron-deficient hydrogen is also a strong hydrogen bond donor site (and also an electrophilic center), shown as the purplish colored sphere (**c**). The green sphere identifies the carbonyl oxygen as an electron rich hydrogen bond acceptor site (and also a nucleophilic center). The blue spheres represent hydrophobic regions of the molecule, in this case the naphthalene ring system. The red and blue regions in the molecule shown in Figures V-7 and V-8(b) also represent *polar bonds*, where *polarity* is defined as an unequal sharing of electrons in a covalent bond. More precisely, polar bonds form as a consequence of *charge separation* between two atoms involved in a covalent bond. Charge separation occurs when the electronegativities of the two atoms differ substantially, e.g., C–O, H–O, C–F, and C–N. In such situations, the bond is said to be *dipolar*. While dipoles develop between atoms in bonds due to electronegativity differences, molecules form *dipole moments.*

Figure V-8. The chemical structure of the carbamate insecticide carbaryl (**a**). When the electrostatic potential is overlaid on the electron density map (**b**) the bright blue area corresponds to the most acidic hydrogen (i.e., the N-methyl proton). But the most electron-deficient hydrogen is also a strong hydrogen bond donor site (**c**), shown as the purplish colored sphere (the green sphere identifies a hydrogen acceptor site, and the blue spheres represent hydrophobic regions of the molecule).



The dipole moment descriptor in Equation 3 is indicative of molecular polarity. Polar molecules readily form hydrogen bonds, which is the intermolecular force necessary for aqueous solvation and physisorption to soil particles and/or the humic/fulvic carbon fractions present in natural waters and soils. In general, the larger the magnitude of the dipole moment the more polar the molecule. Those pesticides having a propensity to leach are more likely to have larger dipole moments, and vice versa, because polar molecules tend to be more water soluble and are thus more mobile in most soils. However, the statistical relationship between mobility and dipole moment may or may not be linear, as other factors can mitigate polarity effects. For example, the insecticide oxamyl [(EZ)-N,N-dimethyl-2-{[(methylcarbam-oyl)oxy]imino}-2-(methylthio)-acetamide] has a significant dipole moment ($\mu = 6.98$), but is not inclined to leach because it degrades rapidly under aerobic conditions in agricultural soils ($t_{1/2} < 10$ days). In contrast, the soil fumigant dibromochloropropane (DBCP) has a relatively low dipole moment ($\mu = 1.88$), but is a known leacher because it is persistent in the soil environment ($t_{1/2} > 300$ days). Given enough time and a recalcitrant nature, almost any chemical has the propensity to eventually leach to groundwater.

Discussion and Summary

That the molecular and electronic structure of molecules play a key role in determining their observed properties and behavior—indeed, that such properties are literally a *function* of structure—is clearly demonstrated by the widespread success of the numerous predictive models based on correlations with molecular structure descriptors generated from structural information. When considering quantum chemical descriptors, virtually all have at least some predictive power because they encode information about the fundamental nature of matter, from the size, shape, and volume of molecules to their electron densities, charge distribution, and orbital energies. Together with empirical descriptors representing macroscale properties, theoretical calculations can provide insight into the intermolecular mechanisms generating the properties under consideration. The objective of CDA was to find linear combinations from the initial pool of theoretical and empirical descriptors that provided maximal separation between the Leaching and Non-Leaching groups of chemicals. From these multivariate analyses, CDA has selected two empirical and three quantum indices for Equation 3. The model describes pesticide persistence directly in terms of experimentally-measured aerobic soil metabolism half-lives, where chemicals having half-lives of less than 60 days are considered transient, and chemicals having half-lives greater than 180 days are considered recalcitrant (U.S. EPA, 2013). Mobility is accounted for empirically by the organic carbon-normalized adsorption coefficient, K_{OC}. While these terms are familiar and often robust indicators of persistence and mobility in the soil environment, their information content and discriminating power has proven to be less than optimal. When evaluated independently, the empirical properties Log_{10} soil aerobic half-life and K_{OC} each account for approximately 86% (6 misclassifications) and 76% (10 misclassifications) of the estimated discriminating power of Equation 3, respectively. But when the quantum chemical predictor variables are included, the model yields just 2 misclassifications, or 95.2% of the total *discriminatory power* of the model (Figure V-9). The quantum chemical descriptors appear to function as statistical weighting factors or fine-tuning terms, amplifying the contributions of the empirical terms and sharpening the overall discriminating power of the model. They encode information describing the structural features of pesticides that link key intermolecular forces to the physicochemical properties influencing their mobility in soils.

Figure V-9. When evaluated independently, the empirical descriptors Log_{10} soil aerobic half-life and K_{OC} each account for approximately 86% and 76% of the estimated discriminating power of the model, respectively. When the quantum chemical predictor variables EHOMO, dipole moment, and Log MaxElPot are included, the separation power increases to over 95%.

$$\mathrm{MLV} = \underbrace{1.9349\,\overset{\textit{Persistence}}{\underbrace{(\text{Log Aerobic } t_{1/2})}_{85.7\%}} - 1.2197\,\overset{\textit{Mobility}}{\underbrace{(\text{Log } K_{oc})}_{76.2\%}}\overbrace{+\,0.1562\,(\text{Dipole Moment}) - 0.9140\,(E_{HOMO}) + 2.2728\,\text{Log}\,(\text{MaxElPot})}^{\textit{Composite Descriptors}}}_{95.2\%}$$

Molecular descriptor values for the 42-chemical development set acquired from experimental measurements (Log_{10} soil aerobic half-life, log K_{OC}) and molecular modeling calculations (dipole moment, E_{HOMO}, log MaxElPot) are summarized in Table V-1. The mean value of each descriptor for the known leacher group and known non-leacher group was determined, and the results are compared below. For those descriptors that were log-transformed, the antilogs are shown in parentheses.

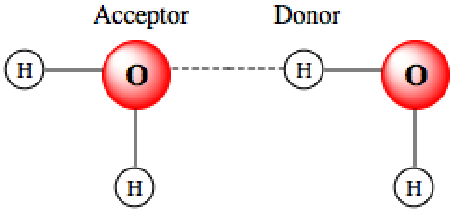
Status	**Mean Log (Soil Aerobic Half-life)**	**Mean Log (K_{OC})**	**Mean Dipole**	**Mean EHOMO**	**Mean Log (MaxElPot)**
Leachers	2.2465 (176.4)	1.8816 (76.1)	4.8039	-8.8789	2.2694 (185.9)
Non-leachers	1.4809 (30.3)	2.5669 (368.9)	3.4246	-8.5075	2.1626 (145.4)

Table V-1. Leacher (L) and non-leacher (NL) model development set of chemicals and the empirical and quantum chemical descriptors used in the Pesticide Leaching Model (Equation 3). The term *MaxElPot* = Maximum Electrostatic Potential.

Leacher Status	Pesticide	Log Aerobic Half-life	Log K_{OC}	Dipole Moment	E_{HOMO}	Log MaxElPot
L	1,2-Dichloropropane (1,2-D)	2.1461	1.6990	0.6500	-10.4000	2.0395
L	2,3,5,6-Tetrachloroterephthalic acid (TPA)	2.1875	1.5798	0.0000	-9.2700	2.4860
L	Aldicarb sulfone	1.3222	1.0000	6.6100	-9.3100	2.3722
L	Atrazine	2.1644	1.9370	5.1900	-8.5200	2.2946
L	Bentazon	1.4914	1.8603	4.4900	-8.7300	2.4481
L	Bromacil	2.5366	1.1492	5.3300	-8.4700	2.4376
L	Deethylatrazine	2.2304	2.0414	5.2100	-8.5900	2.3495
L	Deisopropylatrazine	2.0107	2.1139	5.2800	-8.6100	2.3438
L	Dibromochloropropane (DBCP)	2.5563	2.0000	1.8800	-10.0900	2.1407
L	Diuron	2.5705	2.7326	7.2700	-7.9700	2.4059
L	Ethylene dibromide (EDB)	1.8451	1.9395	0.0000	-10.0100	2.0648
L	Hexazinone	2.3531	1.6551	7.0500	-8.7500	2.1181
L	Imidacloprid	2.9987	2.4609	8.1200	-9.0600	2.3097
L	Metolachlor ethanesulfonic acid (ESA)	2.1206	0.9542	8.9400	-8.9000	2.1423
L	Norflurazon	2.1139	2.6628	7.5900	-8.1500	2.3617
L	Prometon	2.6618	2.0000	3.1800	-8.0800	2.2491
L	Simazine	2.0414	2.1810	5.2600	-8.5400	2.2944
L	Tebuthiuron	3.0864	1.9020	4.4200	-8.3700	1.9910
NL	1,3-dichloropropene	1.5145	1.8195	2.3000	-9.3800	2.1616
NL	Alachlor	1.1761	2.1055	4.5500	-8.6000	1.9609
NL	Carbaryl	0.7404	2.6298	3.1400	-7.6100	2.3416
NL	Carbofuran	1.3444	1.4099	2.1500	-7.6100	2.2999
NL	Chlorthal-dimethyl (DCPA)	1.4116	3.4091	0.0000	-9.1600	2.0507
NL	Cyanazine	1.1875	2.3744	5.1200	-8.6900	2.3564
NL	Dimethoate	0.3802	1.0000	1.7700	-8.7100	2.1291
NL	Disulfoton	1.1931	2.6906	6.0600	-8.2600	2.0402
NL	Ethoprop	1.5328	2.2636	0.3800	-8.8500	1.8855
NL	Fenamiphos	1.3838	2.3508	3.7400	-8.1400	2.3036
NL	Fonofos	1.7980	2.9515	2.6400	-8.2400	1.9630
NL	Iprodione	1.7482	2.8451	2.8700	-9.1300	2.1731
NL	Linuron	1.6893	2.8149	6.4200	-8.0700	2.3545
NL	Methomyl	1.6646	1.6021	4.1000	-8.1600	2.3357
NL	Methyl bromide	0.9823	2.1021	2.3100	-9.7400	2.0363
NL	Napropamide	2.6580	2.8247	4.1000	-7.4600	2.0247
NL	Oryzalin	1.8014	2.9478	3.6500	-8.5500	2.4225
NL	Oxamyl	1.0294	1.4997	3.5500	-8.4400	2.3233
NL	Oxyfluorfen	2.4639	3.8196	6.9000	-8.6000	2.1725
NL	Pendimethalin	2.1004	4.1761	4.6700	-8.2700	2.0712
NL	Phorate	0.4771	2.7311	2.8800	-8.3100	1.9742
NL	Propyzamide	1.4200	2.9165	4.1200	-9.1000	2.3401
NL	Thiobencarb	1.5682	2.7743	4.1600	-8.5500	1.9885
NL	Trifluralin	2.2765	3.5481	0.6100	-8.5500	2.1940

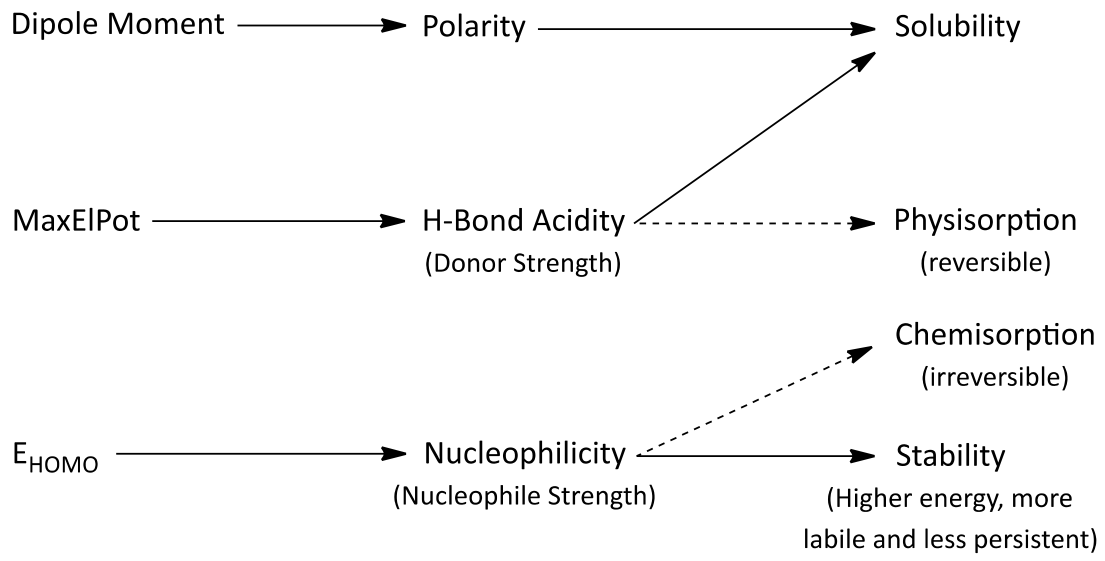
These values offer insight into the relative ability of each structural descriptor to differentiate a leacher class pesticide from a non-leacher class pesticide chemical based on the constituent information it encodes. Predictably, based on the known influence of persistence on leaching potential, the mean aerobic metabolism half-life is much longer for leachers ($\bar{x}$ = 176 days) than for non-leachers ($\bar{x}$ = 30 days). Likewise, pesticides having a propensity to bind to soil organic matter are known to be less mobile than those that do not; consequently, K_{OC} values are much smaller for leachers ($\bar{x}$ = 76 mL/g) than for non-leachers ($\bar{x}$ = 369 mL/g). As previously stated, the two empirical terms differentiate leacher from non-leacher quite well, but the addition of the quantum chemical terms increases the discriminating power of the model significantly (from a misclassification rate of around 15% to 5%).

The quantum chemical descriptors encode specific information about the atomic and molecular properties of molecular systems of interest—the fundamental structural features associated with the system at the atomic level. The dipole moment, for example, is indicative of molecular polarity, but it also addresses the relative hydrogen-bonding capacity of individual molecules. H-bonding is an important mechanism associated with such properties as water solubility and reversible binding to soil (physisorption). A pesticide having a large dipole moment is thus more likely to be water soluble or more readily physisorbed onto soil particles or organic matter than one with a smaller dipole moment, because it has a larger H-bonding capacity. The magnitude of the descriptor value can thus provide insight into which intermolecular forces might be operating. Similarly, the magnitude of certain energy levels of the molecular orbitals present in a compound may offer some predictive qualities. The HOMO electrons are the highest energy electrons in the molecular system and as such are the most readily donatable (i.e., nucleophilic) in a reaction (with the LUMO of the reactant). The higher the energy of the HOMO the more likely the compound will be nucleophilic, labile, and perhaps more susceptible to environmental degradation processes, including irreversible soil binding (chemisorption—adsorption which involves a chemical reaction between the soil surface and the adsorbate). That the higher *mean* HOMO energy is associated with pesticides that do not have significant leaching potential seems to suggest that, in some cases, the parent may break down precipitously or bind irreversibly in the soil environment, which effectively immobilizes it. The maximum electrostatic potential identifies the most acidic hydrogen in the molecular system. A hydrogen atom attached covalently to a relatively electronegative atom such as N, O, or F is not only acidic, but is also an effective hydrogen bond donor. A hydrogen bond donor site is also known as a *hydrogen bond acid* (similarly, a hydrogen bond acceptor site is called a *hydrogen bond base*), analogous to the Lewis definition of acids and bases as electron donors and acceptors. In a hydrogen bond, the electronegative atom not covalently attached to the hydrogen is designated the proton acceptor, whereas the hydrogen that is covalently bound to the hydrogen is termed the proton donor, as illustrated below.



A hydrogen bond donor site encourages solubility and irreversible binding to soils and discourages the volatility of polar, semi-polar, and amphiphilic compounds (where an amphiphile is a chemical compound possessing *both hydrophilic* and hydrophobic groups, e.g., fatty acids or soaps). The relationship between the quantum chemical descriptors discussed above, the intermolecular and orbital properties they encode, and the physicochemical property or leaching mechanism associated with them are shown in Figure V-10. These associations are also summarized in Table V-2.

Figure V-10. The connections between the quantum descriptors, the molecular property they encode, and the leaching mechanisms associated with them.



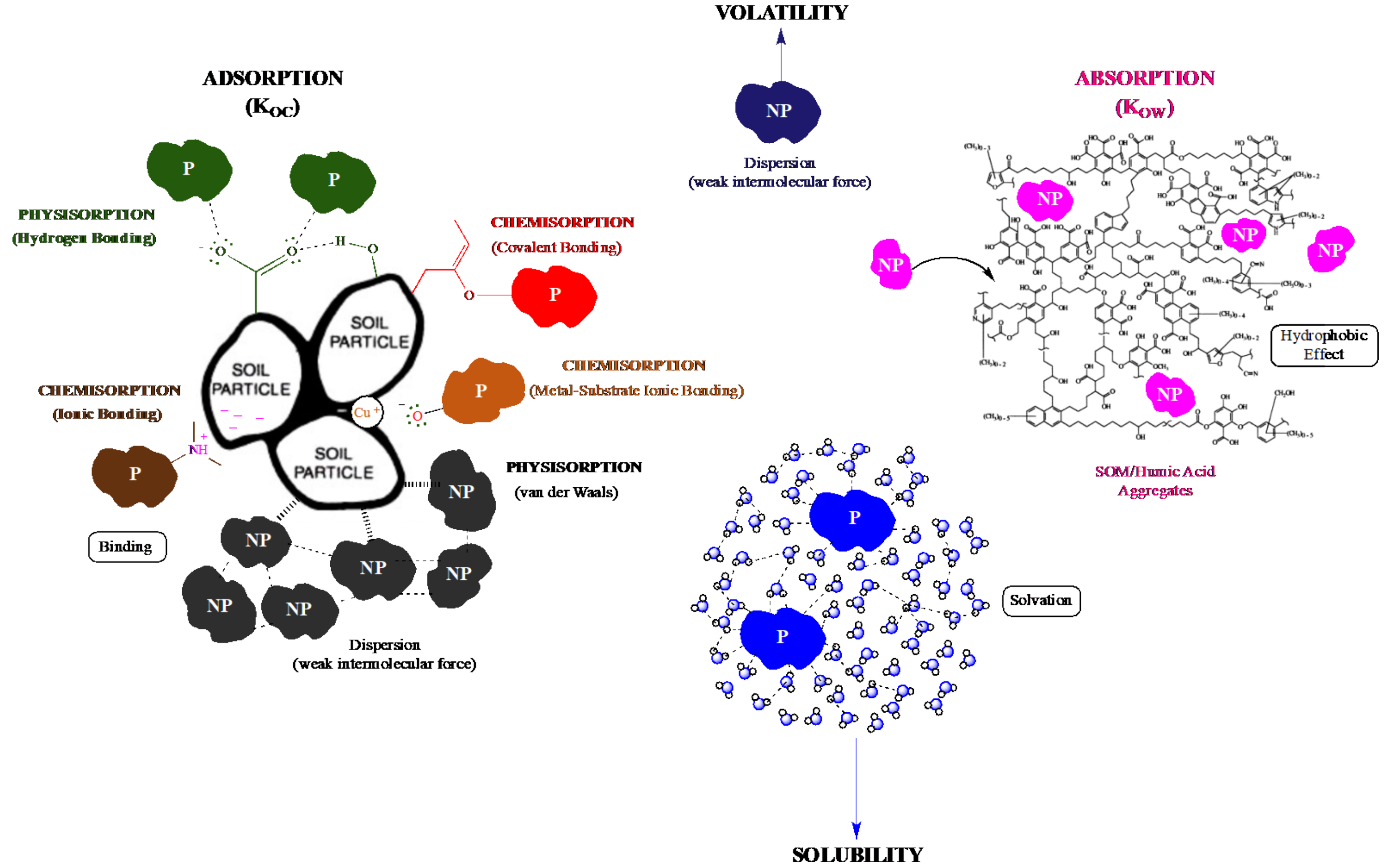
Although dipole moment, E_{HOMO}, and MaxElPot each encode information about specific molecular properties (e.g., dipole moment ≡ hydrogen bonding capacity and molecular polarity; see Figure V-7), these terms effectively form a *composite molecular descriptor* representing aggregate electronic structural properties that combine in unique but unknown proportions to describe individual molecules. They primarily act as ancillary terms which refine the strong but incomplete associations already established by the empirical descriptors aerobic soil half-life and K_{OC}. Dipole moment, E_{HOMO}, and Log MaxElPot were selected using a dimension-reduction procedure from a multi-dimensional space initially containing dozens of input variables; therefore, it is difficult (if not impossible) to discern in what manner or to what extent they combine to encode relevant supplemental information. However, it is possible to make inferences based on fundamental chemical principles, as previously discussed. For example, a pesticide having a large dipole moment will likely be polar and (depending on its other properties) water soluble or susceptible to physisorption, since mechanistically both are hydrogen-bond dependent. A pesticide having a small dipole moment, on the other hand, might be expected to be non-polar and prone to adsorption/absorption via dispersion forces or to volatility.

The various intramolecular and intermolecular forces producing the properties that drive pesticide mobility in agricultural soils are summarized graphically in Figure V-11. *Intermolecular forces*, much weaker than the *intramolecular* forces of attraction that hold molecules together, are particularly significant because they determine the *physical properties* of molecules (i.e., properties that do not change the identity of the substance). *These forces* represent the underlying mechanisms from which the observed and quantifiable properties of molecules develop. Accordingly, mobility is largely governed by the weak forces that originate in the dipolar interactions *between* molecules (see Figure V-3). The attractions and repulsions experienced by interacting molecules are coulombic in nature, i.e., they occur as a result of the interplay between neighboring charges and partial charges, resulting in the development of dipolar forces. How do these forces affect pesticide mobility? Soils are composed of a mixture of sand, silt, clay, and organic matter. Both the clay and organic matter particles have a net negative charge. Thus, these negatively charged soil particles will attract and hold positively charged pesticides, both via chemisorption (intramolecular covalent/ionic/metallic bonding) and physisorption (intermolecular hydrogen bonding, dispersion) mechanisms. The dispersion (i.e., induced dipole-induced dipole) forces characterizing uncharged, nonpolar pesticides is the principle driving force for the hydrophobic effect and the absorption process in soils, suspended and dissolved solids, and organic matter. Dispersion forces, if numerous enough (e.g., in a large nonpolar molecule), will generally inhibit volatilization, facilitate the hydrophobic effect (i.e., colloidization and aggregation), and ultimately promote absorption to nonionic surfaces. Polar and many amphiphilic compounds may (1) interact strongly with the aqueous soil solution via hydrogen bonding and are readily solvated and transported through the soil profile, and (2) adsorb to soil particles, organic matter, and other soil fractions containing polar functional groups or charged ionic or metallic species.

Table V-2. Overview of the model quantum descriptors and the intermolecular forces and physicochemical properties associated with them.

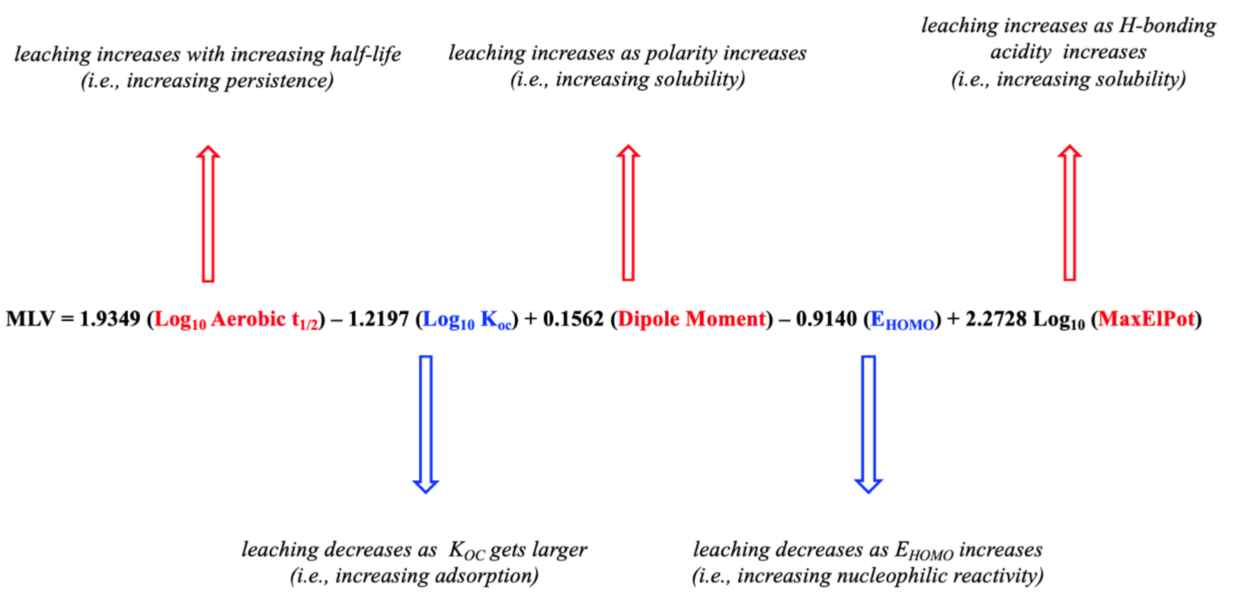
<table>
<tr><th colspan="3">Dipole Moment</th></tr>
<tr><td>Descriptor</td><td>Physicochemical Property</td><td>Associated Intermolecular Force</td></tr>
<tr><td>High values:</td><td>Solubility
Physisorption (H-bonding)
Physisorption (ionic bonding)</td><td>Van der Waals forces:
• Hydrogen bonding
• Dipole-dipole interactions
• Ion-dipole interactions</td></tr>
<tr><td>Low values:</td><td>Non-polar interactions:
Absorption (hydrophobic effect)
Volatility (weak attractions)
Physisorption (via dispersion forces)</td><td>Weak Intermolecular Forces:
• Induced dipole interactions
• Dipole-induced dipole interactions
• Ion-induced dipole interactions
• London dispersion forces</td></tr>
<tr><th colspan="3">Maximum Electrostatic Potential (MaxElPot)</th></tr>
<tr><td>Descriptor</td><td>Physicochemical Property</td><td>Associated Intermolecular Force</td></tr>
<tr><td>High values:</td><td>Physisorption (via-H-bonding)</td><td>Acidic hydrogens and associated hydrogen bond donor sites</td></tr>
<tr><td rowspan="2">Low values:</td><td>Solubility (via H-bonding)
Volatility (weak attractions)</td><td>Weak non-polar interactions (van der Waals forces); hydrophobicity</td></tr>
<tr><td>Absorption into soil organic matter (hydrophobic effect)</td><td>Non-polar, weak van der Waals interactions</td></tr>
<tr><th colspan="3">E_{HOMO}</th></tr>
<tr><td>Descriptor</td><td>Leaching Mechanism</td><td>Associated Intramolecular Force</td></tr>
<tr><td>High values:</td><td>Chemisorption (covalent, ionic or metallic/coordination complex bonding); moderate to rapid abiotic and/or biotic degradation</td><td>Orbital energy-facilitated nucleophilic reaction sites</td></tr>
<tr><td>Low values:</td><td>Chemical stability; Environmental persistence</td><td>Orbital stability, non-reactive.</td></tr>
</table>

Figure V-11. Overview of the intramolecular (covalent/ionic/metallic bonding) and intermolecular (van der Waals, dispersion) forces driving the mobility of pesticides in soils.



That the empirical terms Log Aerobic Half-life and Log K_{OC} are strongly associated with the environmental persistence of pesticides and pesticide mobility in soils, respectively, is unequivocal and well-established. A pesticide that is environmentally persistent has an increased potential to leach to groundwater based solely on its long-term availability in the soil. If it has a low K_{OC} it may eventually percolate through the soil profile and contaminate susceptible aquifers. If it has a high K_{OC} it may be prone to runoff via a number of potential transport mechanisms. The electronic and molecular orbital descriptors selected by CDA are aggregate indices that discriminate between leacher and non-leacher groups in a significant but inexplicable manner. Yet they represent real molecular properties. Collectively they encode information which broadly accounts for the intermolecular forces responsible for producing the various physicochemical properties that influence and drive the leaching process in the soil environment. As illustrated in Figure V-12, the leaching model indicates that leaching potential increases (i.e., MLV values > 14.4706) with increasing persistence (log soil aerobic half-life), polarity (dipole moment) and H-bond acidity (maximum electrostatic potential), but decreases (i.e., MLV values < 14.4706) with increasing adsorption (K_{OC}) and orbital energy (since bonding orbitals are always negative, E_{HOMO} energies get progressively less negative as they increase). Combined, these five parameters appear to uniquely capture the molecular properties of each pesticide active ingredient that are indicative of leaching. While the tendency to leach is strongly dependent on properties that can be accounted for directly by empirical relationships, these associations alone do not adequately discriminate between those chemicals that are potential leachers and those that are not. The quantum descriptors discriminate more surreptitiously than those drawn from empirical data, but the aggregate electronic structure information they encode provides key information about each chemical's unique physicochemical and molecular properties and helps compensate for the information absent in empirical data.

Figure V-12. Summary of how the model responds when each descriptor increases in magnitude.



$$\mathrm{MLV} = 1.9349\,(\mathrm{Log_{10}\ Aerobic\ t_{1/2}}) - 1.2197\,(\mathrm{Log_{10}\ K_{oc}}) + 0.1562\,(\mathrm{Dipole\ Moment}) - 0.9140\,(E_{HOMO}) + 2.2728\ \mathrm{Log_{10}}\,(\mathrm{MaxElPot})$$

References

Doucette, W.J. 2003. Quantitative Structure–Activity Relationships for Predicting Soil–Sediment Sorption Coefficients for Organic Chemicals, Environmental Toxicology and Chemistry, Vol. 22, No. 8, pp. 1771–1788.

Gao, S. 2013. Interpreting the wave function--what are electrons, and how do they move? Discusiones Filosóficas, 14(22), 13-23. Retrieved June 14, 2019, from http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0124-61272013000100002&lng=en&tlng=en (verified 24 July 2024).

Hehre, W. J. 2003. A Guide to Molecular Mechanics and Quantum Chemical Calculations, Wavefunction, Inc.

Hehre, W. and S. Ohlinger. 2016. Spartan’16 Tutorial and User's Guide, Appendix A, pp. 506-507. Wavefunction, Inc., 18401 Von Karman Avenue, Suite 370, Irvine, CA 92612.

Kah, M., Beulke, S., and Brown, C.D. 2007. Factors Influencing Degradation of Pesticides in Soil, *J. Agric. Food Chem.*, **55** (11), pp 4487–4492.

Kamachi, T., Tatsumi, T., Toyao, T., Hinuma, Y., Maeno, Z., Takakusagi, S., Furukawa, S., Takigawa, I., and Shimizu, K. 2019. Linear Correlations between Adsorption Energies and HOMO Levels for the Adsorption of Small Molecules on TiO2 Surfaces. J. Phys. Chem. 123, 34, 20988–20997.

Karanfil, T. 2006. Activated carbon adsorption in drinking water treatment, in: T. Bandosz (Ed.), Act. Carbon Surfaces Environ. Remediat., Elsevier, New York: pp. 345–373.

Karelson M., V.S. Lobanov, and A.R. Katritzky. 1996. Quantum-Chemical Descriptors in QSAR/QSPR Studies, Chem. Rev., 96 (3), pp 1027–1044.

MacKay, A.A. and Canterbury, B. 2005. Oxytetracycline sorption to organic matter by metal-bridging, J Environ Qual. 12;34 (6):1964-71.

Mamy, L., Patureau, D., Barriuso, E., Bedos, C., Bessac, F., Louchart, X., Martin-laurent, F., Miege, C., and Benoit, P. 2015. Prediction of the Fate of Organic Compounds in the Environment from Their Molecular Properties: A Review, Critical Reviews in Environmental Science and Technology, 45:1277–1377.

National Research Council. 2014. A Framework to Guide Selection of Chemical Alternatives. Chapter 5, Physicochemical Properties and Environmental Fate. P. 52. Washington, DC: The National Academies Press.

Organization for Economic Cooperation and Development (OECD). 2007. Guidance document on the validation of (Quantitative) Structure- Activity Relationship [(Q)SAR] models. OECD Series on Testing and Assessment, No. 69. Available at: https://www.oecd.org/env/guidance-document-on-the-validation-of-quantitative-structure-activity-relationship-q-sar-models-9789264085442-en.htm (verified 24 July 2024).

Oskouie, A.K., Miura, Y., Furuya, E.G., Noll, K.E. 2002. Relationship between the highest occupied molecular orbital (HOMO) electron density of adsorption sites on carbon and the Freundlich exponent, Carbon, Volume 40, Issue 8, July 2002, Pages 1199-1202.

Pal, R., Chakrabarti, K., Chakraborty, A., and Chowdhury, A. 2006. Degradation and Effects of Pesticides on Soil Microbiological Parameters-A Review. International Journal of Agricultural Research, 1: 240-258.

U. S. EPA. 2013. Interpretive Assistance Document for Assessment of Discrete Organic Chemicals. Available at: https://www.epa.gov/sites/production/files/2015-05/documents/05-iad_discretes_june2013.pdf (verified 24 July 2024).

Appendix VI —Procedure for Calculation of Molecular Properties of Chemicals Using Spartan ’20 Software

1. Search for chemical name at PubChem website (https://pubchem.ncbi.nlm.nih.gov/).
    a. Verify search result with the CAS number from DPR’s Pesticide Data Index Application and the reported molecular weight. Consider verifying search result of molecular diagram with structure diagram submitted by registrant.
    b. When considering a salt or ester that disassociates rapidly into the acid form, select structure of the acid form.
2. Following selection of the chemical in PubChem, select ‘Download’ tab, then ‘Save’ for ‘SDF’ file format under ‘2D Structure’ option.
3. In Spartan ’20 open the saved SDF file from the ‘File’ menu. Accept auto conversion from 2D to 3D format.
    a. If the SDF file is unavailable or the structure is incorrect compared to the structure diagram submitted by the registrant, manually create the molecular structure in Spartan ‘20. Alternatively, input the SMILES formula in Spartan ’20 to auto-generate the structure. Verify the resultant molecular structure with the structure diagram submitted by the registrant.
    b. If the molecule is chiral, change the chiral center(s) from R to S, as necessary, and freeze chirality.
4. Select ‘Setup’ from menu. Select ‘Calculations’ from drop down menu.
    a. Select ‘Equilibrium Conformer’ with ‘Molecular Mechanics’ ‘MMFF’ from drop down menu.
    b. Check ‘OPTIONS’ box and enter ‘SEARCHMETHOD=SYSTEMATIC’ into text field.
    c. Select ‘Submit’. Allow Spartan ’20 to complete execution, then select ‘OK’.
5. Select ‘Setup’ from menu. Select ‘Calculations’ from drop down menu.
    a. Select ‘Equilibrium Geometry’ at ‘Ground’ state in ‘Gas’ with ‘Density Functional’ ‘wB97X-D’ ‘6-31G’ from drop down menu.
    b. Check ‘QSAR’ box.
    c. Check ‘OPTIONS’ and enter ‘POSTSOLVENT=SM8:WATER’ into text field.

   d. Select ‘Submit’. Allow Spartan ’20 to complete execution. Select ‘Monitor’ icon in menu ribbon to display progress of execution and select ‘OK’ once completed. In the rare event when execution fails to complete, rerun the Equilibrium Geometry step without the SM8 option.

6. Select ‘Display’ menu and select ‘Properties’ in drop down menu.
   a. In ‘Properties’ box, select ‘Molecule’ tab, then click selection for ‘EHOMO’ and ‘Dipole Moment’.
   b. In ‘Properties’ box, select ‘QSAR’ tab, then click select for ‘MaxELPot’.
   c. Close ‘Properties’ box.
7. Select ‘Display’ menu, then ‘Spreadsheet’ from drop down menu. Values for ‘EHOMO’, ‘Dipole Moment’, and ‘MaxELPot’ will be displayed.
   a. If ‘Equilibrium Geometry’ was executed without the SM8 option, the calculated dipole moment value will be underestimated by 25-30%. Adjust dipole moment value accordingly.