Chapter 5

Application of SDD Index

5.1 Introduction

In this chapter, we analyze the relation between the SDD index with other physicochemical properties of PCBs, such as log water solubility, octanol-water partition ratio, with the aid of computers. We also compare it with other well-known vertex degree-based (VDB) indices. We show that the SDD-index has better correlation ability with a value of 0.96 for log-water solubility and 0.93 for octanol-water partition coefficient. We also identify the expected values for these two properties of PCBs whose experimental values are not available.

The organization of the chapter is as follows. Section 5.2, we gives application of physicochemical properties. In Section 5.3.1, we study correlation model and in Section 5.3.2, we study the correlation of the SDD index with the physiochemical properties of PCB and make a comparative analysis with other VDB indices.

5.2 Physicochemical Properties

One of the essential physicochemical properties in the modern drug discovery process is the solubility of organic molecules in water. Its significance is mainly due to its impact on ADME-related properties like drug uptake, distribution, and bioavailability. In addition, solubility testing can also be relevant for property-based computational screening methods in the drug discovery process. Hence, there is a significant contribution and interest in fast, reliable, structure-based methods for predicting solubility in water for promising drug candidates for a long time, see [123].

Lipophilicity is ranked as one of the most important physicochemical properties to screen lead compounds in the drug discovery process. It has applications in environmental risk assessment and agricultural fields. The correlation between lipophilicity and biological activity has provoked the extensive use of the octanol-water partition ratio, among others, as a descriptor in quantitative structure-activity/property relationships. Note that the octanol-water partition ratio is the most common way of expressing the lipophilicity of a compound and is the ratio of the concentration of a solute in a water-saturated octanol phase to the concentration in an octanol-saturated aqueous phase. For a detailed review of octanol-water partition ratio, see [124]. There are many more physicochemical properties that are relevant in drug discovery, such as melting point, enthalpy of formation, to name a few. We refer the readers to literature on applications of chemical graph theory to drug discovery for more details, see for instance, [123–125].

The industrial chemicals, polychlorinated biphenyls (PCBs) are congeners with various degrees of chlorination and substitution patterns, that have been used for various commercial applications, such as heat transfer fluids, organic dilutents, paint additives, etc.

5.3 VDB Indices vs Properties of PCB Congeners

To facilitate our study and to compare the physicochemical properties of PCBs, we have used the experimental data provided by the article [126]. The data given in Table B.1 as an appendix at the end of this thesis is for a quick reference. From the table B.1, we find that there are over 200 PCB compounds, each having distinct properties. Hence it becomes interesting to estimate if the VDB indices correlate with any of these properties.

5.3.1 Statistical Model

The model used for testing our assertion is given by the following linear regression equation $P = c + m \cdot TI$, where P is a physicochemical property, TI is a topological index (VDB index in our case), c is the intercept, and m is the slope of linear fit.

For instance, when we consider the physicochemical property of log water solubility of PCB congeners and compare it with the SDD-index, we obtain that

$$logSw = -0.6738 + 0.1885(SDD),$$

where the correlation between them is $r = 0.96$ with a standard error of 0.3002. The F-test value for the fit is 1800.556. The same is shown in the last row of Table 5.1.

The details of linear regression fit for each of the considered properties against the chosen VDB indices are given in Tables 5.1 to 5.5 along with their respective graphical representation in Figures 5.1 to 5.5. Each table provides the computational value of the intercept c and the slope m by fitting each VDB index over a specific physicochemical property. In this statistical fitting process, we also obtain the correlation coefficient r, the F test value for the standard error SE in the fit, and its corresponding test significance SF.

Note that the correlation coefficient tells us the strongness of the linear relationship, and its value lies between -1 to $+1$. If the correlation coefficient, r, is positive, it indicates that the value of the independent variable increases with the increase in the dependent variable. If r is negative, then the independent variable increases with the decrease in the dependent variable. If correlation is perfect then the value of r must be $+1$ or -1. In comparison, the standard error of the considered regression model gives the precision of the regression coefficient. At the same time, significance F helps analyze the reliability of the obtained results. Now we focus on analyzing the obtained results and interpreting the same.

5.3.2 Results and Data Analysis

As a first observation, we note that the SDD index has a higher correlation with the log water solubility parameter among all the physicochemical properties considered.

Log-Water Solubility: In the methodology, we have illustrated fitting the linear regression model for the log-water solubility of PCB congeners when compared with SDD-index. Similary, we next compare the experimental value of log water solubility of PCB congeners with the ZM_1-index and obtained the linear relation as

$$logSw = -1.4701 + 0.1025(ZM_1),$$

while the correlation between them is $r = 0.96$ with a standard error of 0.3137. We also show the comparative correlation plots between the log-water solubility (logSw) property with the VDB indices and SDD index in Fig 5.1. From the figure, we also observe that the data is aligned vertically due to the higher degree of degeneracy shown by the VDB indices on the PCB data.

Table 5.1 compare the log water solubility ($logSw$) of PCB with the VDB indices. That is, the highest value of correlation coefficient of $r = 0.96$ is observed between SDD and $logSw$ values. Similarly, the index ZM_1 also has a similar performance when compared to SDD index. While the other indices ZM_2, R, AZI, H, ISI has a correlation close to 0.92 to 0.95.

Thus, we find that the VDB indices have a higher correlation value with the log water solubility property of the PCBs, with SDD and ZM_1 having a slightly higher advantage than the others.

VDB index	c	m	r	SE	F	SF
ZM_1	-1.4701	0.1025	0.96	0.3137	1637.818	5.33E-77
ZM_2	-0.2071	0.0745	0.95	0.3379	1392.546	1.16E-72
R	-4.3232	1.4839	0.95	0.3209	1559.363	1.11E-75
AZI	-3.2953	0.0785	0.92	0.4334	794.336	3.55E-58
H	-6.0820	1.8015	0.95	0.3326	1442.167	1.35E-73
ISI	-2.3240	0.4828	0.95	0.3284	1482.105	2.53E-74
$\boldsymbol{SDD}$	**-0.6738**	**0.1885**	**0.96**	**0.3002**	**1800.556**	**1.47E-79**

TABLE 5.1: The statistical parameters of linear regression fit with correlation coefficient for VDB index and log water solubility ($logSw$) are given.

Octanol-water partition: Table 5.2 and Figure 5.2 compare the octanol-water partition coefficient, $logP$, of PCB with the VDB indices. Here we have obtained that the SDD index has the highest value of correlation coefficient of $r = 0.93$, while the indices ZM_1, ZM_2, R, H and ISI also follow closely with r-values above 0.9 and the lowest correlation in this category is displayed by AZI with $r = 0.86$. Thus, we conclude here again that the SDD index performs better than the others

VDB index	c	m	r	SE	F	SF
ZM_1	0.408	0.066	0.92	0.2995	765.4899	6.10E-58
ZM_2	1.253	0.048	0.90	0.3188	659.2568	3.28E-54
R	-1.430	0.961	0.91	0.3050	729.8739	9.66E-57
AZI	-0.628	0.049	0.86	0.3832	414.2623	2.99E-43
H	-2.548	1.164	0.91	0.3148	679.8153	5.71E-55
ISI	-0.125	0.312	0.91	0.3116	696.5609	1.42E-55
$\boldsymbol{SDD}$	**0.903**	**0.123**	**0.93**	**0.2860**	**852.3404**	**1.12E-60**

TABLE 5.2: The statistical parameters of linear regression fit with correlation coefficient for VDB index and octanol-water partition coefficient ($logP$) are given.

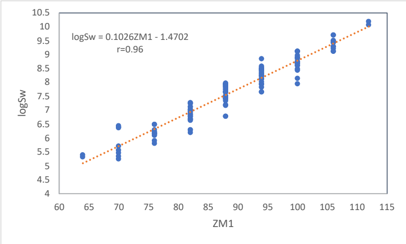
when studying the octanol-water partition coefficient of PCBs.

Melting Point: Table 5.3 and Figure 5.3 compares the melting point, MP, of PCBs with the VDB indices. Our first observation is that the highest correlation of $r = .78$ is displayed by two indices, namely ZM_1 and SDD. Further, we also note that the correlation is closer to 0.77 for the other

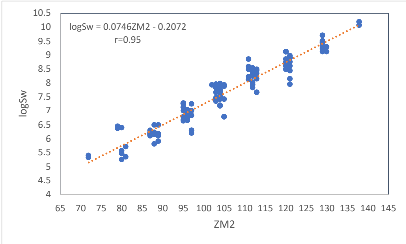
VDB index	c	m	r	SE	F	SF
ZM_1	-187.85	3.370	0.78	32.1001	121.3556	1.46E-17
ZM_2	-146.276	2.447	0.77	32.371	118.0252	2.84E-17
R	-281.594	48.756	0.77	32.171	120.469	1.74E-17
AZI	-244.9	2.551	0.75	33.8768	100.9932	1.01E-15
H	-339.207	59.153	0.77	32.305	118.8297	2.41E-17
ISI	-215.859	15.859	0.77	32.2561	119.4315	2.14E-17
$\boldsymbol{SDD}$	**-161.481**	**6.195**	**0.78**	**32.0059**	**122.5311**	**1.16E-17**

TABLE 5.3: The statistical parameters of linear regression fit with correlation coefficient for VDB index and Melting point (MP) are given.

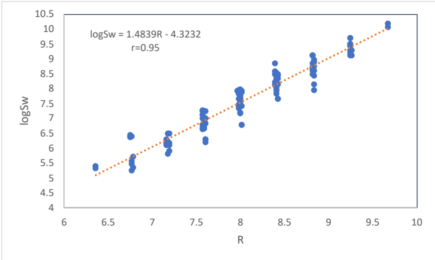
indices ZM_2, R, H, ISI, and it is 0.75 for AZI.



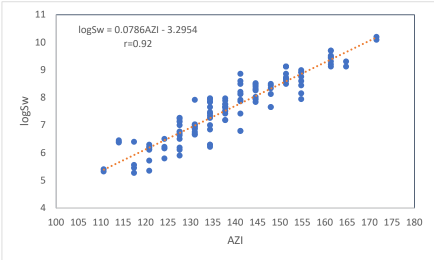
(a) ZM_1 index vs $logSw$



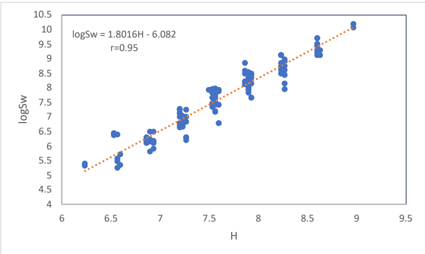
(b) ZM_2 index vs $logSw$



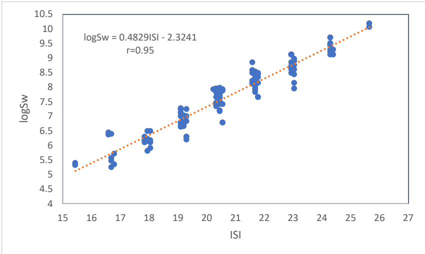
(c) R index vs $logSw$



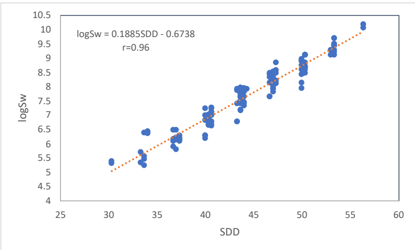
(d) AZI index vs $logSw$



(e) H index vs $logSw$



(f) ISI index vs $logSw$

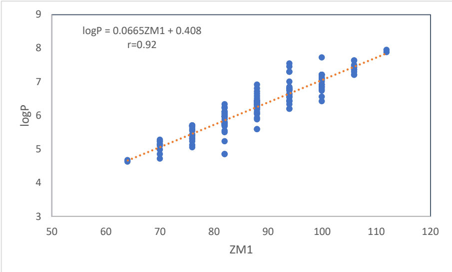


(g) SDD index vs $logSw$

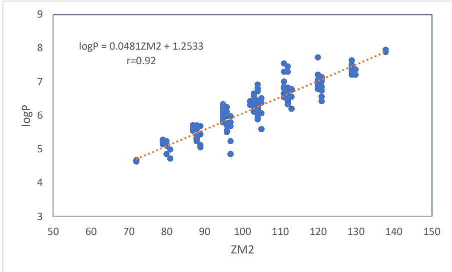
FIGURE 5.1: Correlation of vertex degree based indices with log water solubility ($logSw$).

<u>*Relative Enthalpy of Formation:*</u> Table 5.4 and Figure 5.4 compares the relative enthalpy of formation of PCBs with the VDB indices. The degree-based indices have similar behavior, with an identical correlation value of 0.67 (round-off to two decimal places).

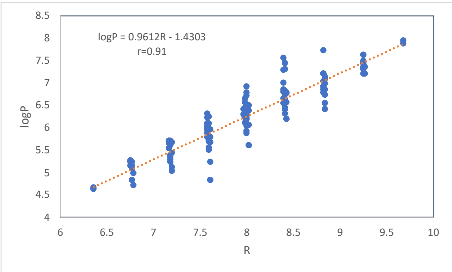
Moreover, we want to highlight here that one of the reasons could be the non-availability of sufficient experimental data for studying this property of PCBs (see column 6 of Table B.1). Even with the minimal data set, we see that the VDB indices have a better correlation.



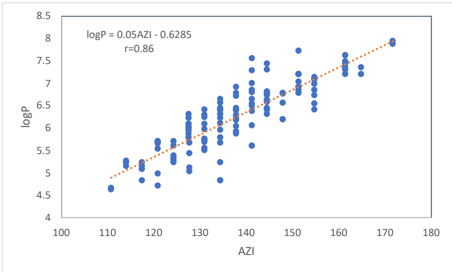
(a) ZM_1 index vs $logP$



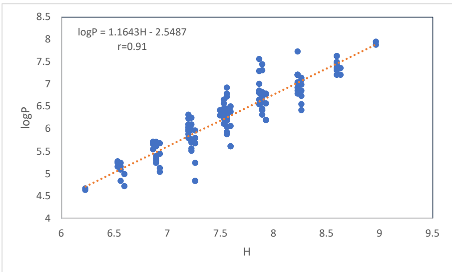
(b) ZM_2 index vs $logP$



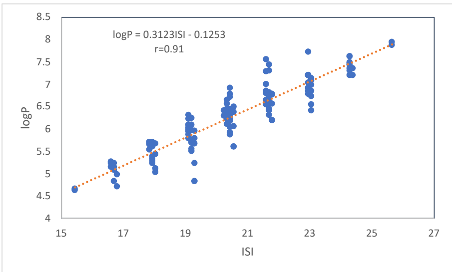
(c) R index vs $logP$



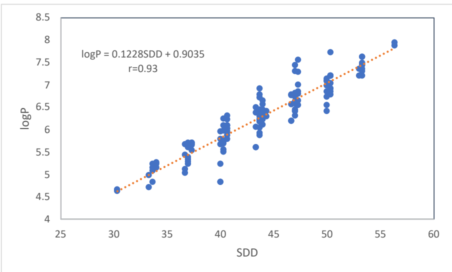
(d) AZI index vs $logP$



(e) H index vs $logP$



(f) ISI index vs $logP$

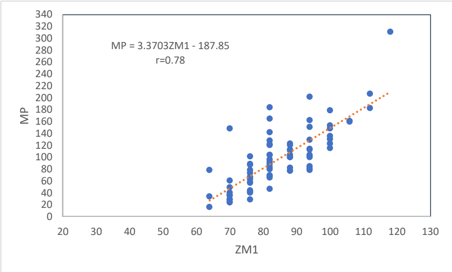


(g) SDD index vs $logP$

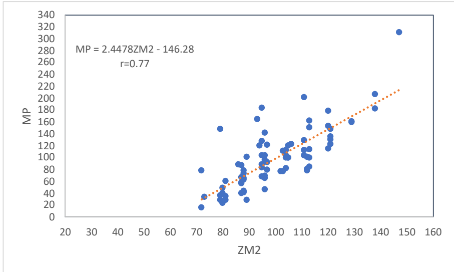
FIGURE 5.2: Correlation of vertex degree based indices with octanol-water partition coefficient ($logP$).

Thus, for the relative enthalpy of formation of PCB compounds, the VDB indices are positively correlated, with the correlation coefficient value being approximately the same.

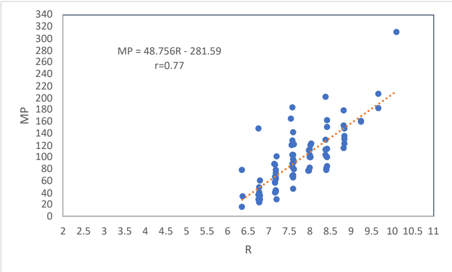
Log Henry constant: Table 5.5 and Figure 5.5 shows that the VDB indices and the log Henry constant, ($logH$), are not correlated and that the correlation coefficient is 0.11.



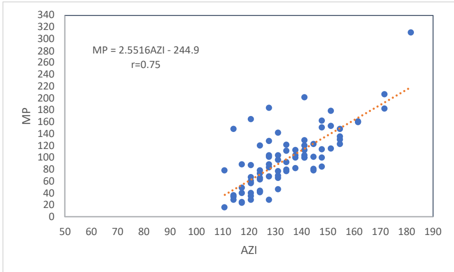
(a) ZM_1 index vs MP



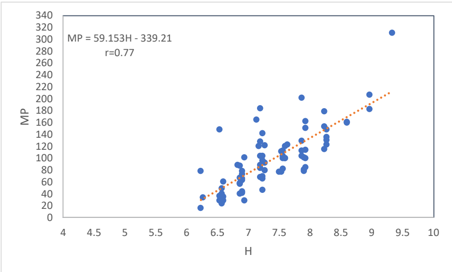
(b) ZM_2 index vs MP



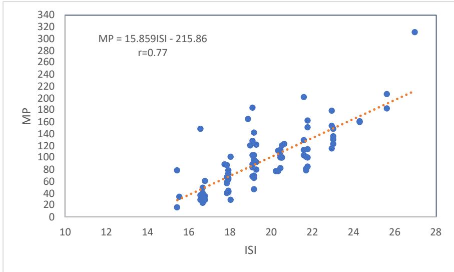
(c) R index vs MP



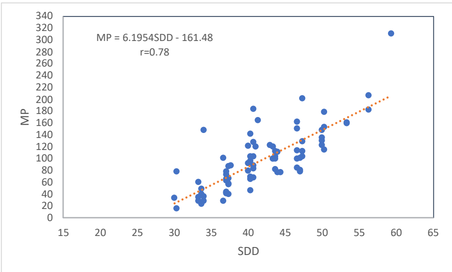
(d) AZI index vs MP



(e) H index vs MP



(f) ISI index vs MP



(g) SDD index vs MP

FIGURE 5.3: Correlation of vertex degree based indices with the melting point (MP).

However, we want to highlight here that one of the reasons could be the non-availability of sufficient experimental data for studying this property of PCBs (see column 5 of Table B.1). Further, as one is well aware that the value of Henry's constant depends on the nature of the gas, solvent, temperature and pressure. While the vertex degree based indices does not consider any of these parameters and depend only on the structure/bond within the molecules. Hence, we naturally arrive at the fact that the VDB indices do not correlate with the PCBs' log Henry constant values.

VDB index	c	m	r	SE	F	SF
ZM_1	2.4283	0.2075	0.67	3.7848	11.6931	0.0041
ZM_2	4.7498	0.1523	0.67	3.7712	11.8789	0.0039
R	-3.4455	3.0119	0.67	3.7795	11.7644	0.004
AZI	-2.5159	0.1655	0.67	3.7671	11.9343	0.0038
H	-7.177	3.6722	0.67	3.7733	11.849	0.0039
ISI	0.5269	0.9828	0.67	3.7752	11.8230	0.0039
$\boldsymbol{SDD}$	**4.225**	**0.3791**	**0.67**	**3.8010**	**11.4736**	**0.0044**

TABLE 5.4: The statistical parameters of linear regression fit with correlation coefficient for VDB index and relative enthalpy of formation (dHf) are given.

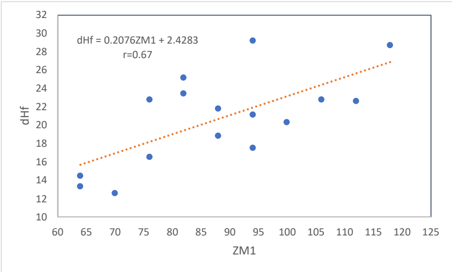
VDB index	c	m	r	SE	F	SF
ZM_1	1.0985	0.0304	0.11	2.225	0.26085	0.6157
ZM_2	1.3181	0.023	0.12	2.2243	0.2834	0.6009
R	0.1711	0.451	0.12	2.2252	0.26843	0.6106
AZI	-0.2930	0.0299	0.13	2.2210	0.3380	0.5681
H	-0.4994	0.566	0.12	2.2245	0.2791	0.6036
ISI	0.7155	0.1501	0.12	2.2248	0.2755	0.6060
$\boldsymbol{SDD}$	**1.4667**	**0.0525**	**0.11**	**2.2268**	**0.2418**	**0.6288**

TABLE 5.5: The statistical parameters of linear regression fit with correlation coefficient for VDB index and log Henry constant ($logH$) are given.

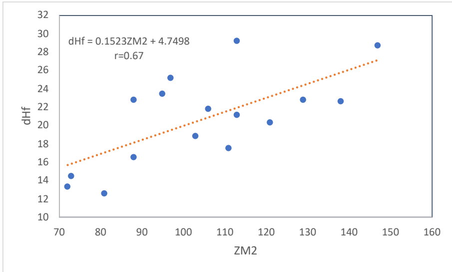
Finally, as an overall observation, we have seen that the SDD index has an excellent correlation mainly with log water solubility and octanol-water partition coefficient.

5.3.3 Predictability Analysis

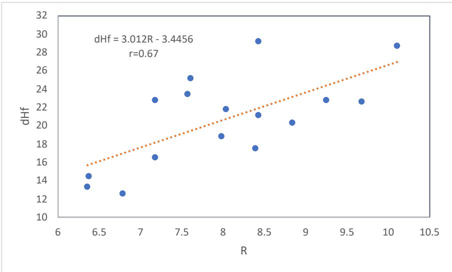
We further analyze these two properties and extend a predictive analysis for the PCB congeners. Firstly, we compute the expected value for the $logSw$ and $logP$ using the parameters obtained from Table 5.1 and Table 5.2, respectively. Figure 5.6(a) and 5.6(b), compares the experimental and the expected value of the $logSw$ and $logP$, respectively, from the available data on PCBs. From the figure, it is pretty evident that the error(less than 0.30) between the experimental and expected values is significantly less. Hence, using the parameters c and m derived from this analysis for the SDD index, we compute the missing data for $logSw$ and $logP$. Columns 7 and 8 of Table B.1 give the expected/predicted value (shown in bold font) for which there was no data available, while the remaining entries (shown in normal font) represent the experimental data.



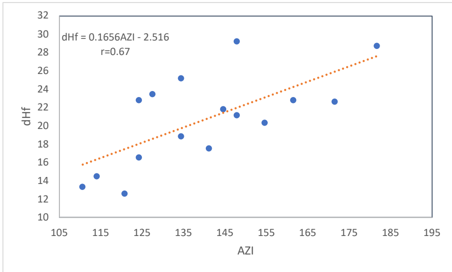
(a) ZM_1 index vs dHf



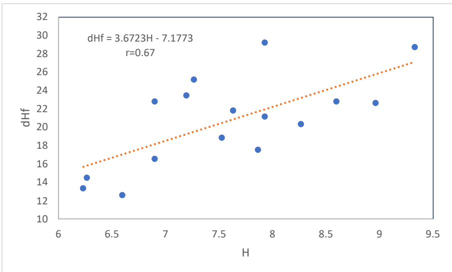
(b) ZM_2 index vs dHf



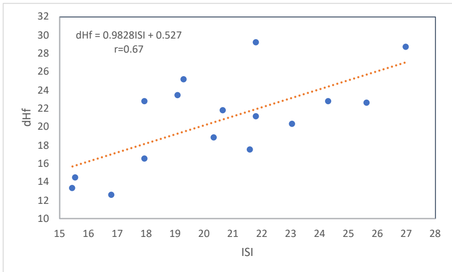
(c) R index vs dHf



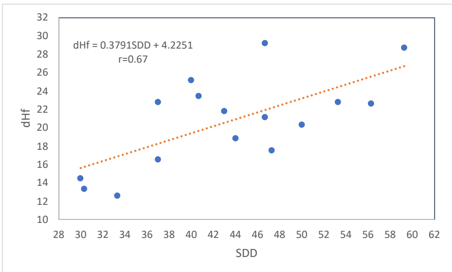
(d) AZI index vs dHf



(e) H index vs dHf



(f) ISI index vs dHf

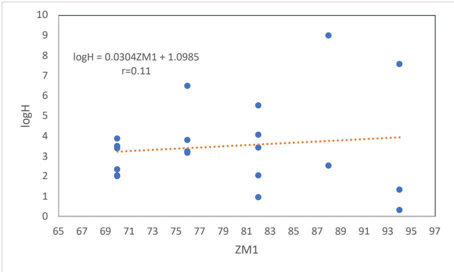


(g) SDD index vs dHf

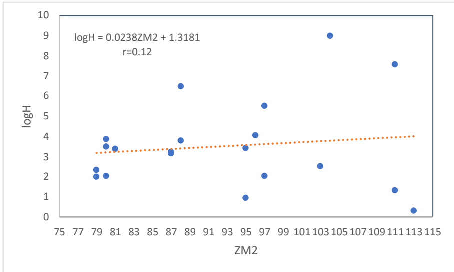
FIGURE 5.4: Correlation of VDB indices with relative enthalpy of formation (dHf).

5.4 Summary

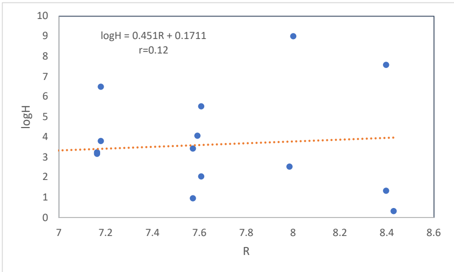
In this chapter, we have studied the correlation and predictive ability of Symmetric division degree index computationally. We have shown that the SDD index is a viable and applicable molecular descriptor by analyzing some physicochemical properties of PCB congeners. Firstly we see that the



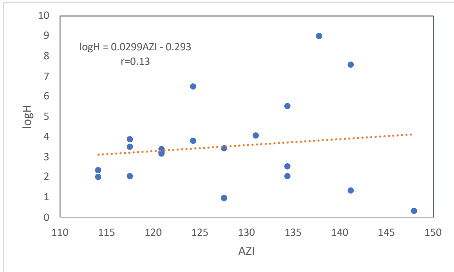
(a) ZM_1 index vs $logH$



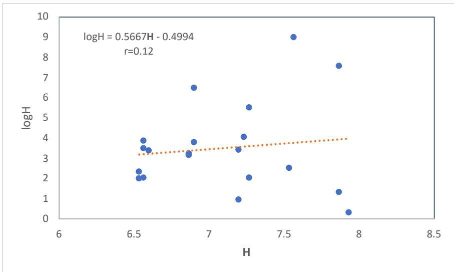
(b) ZM_2 index vs $logH$



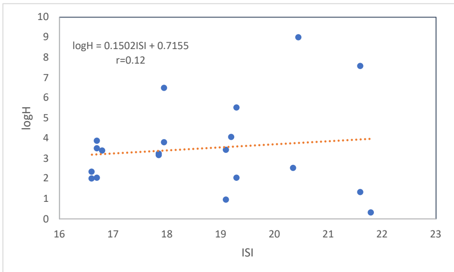
(c) R index vs $logH$



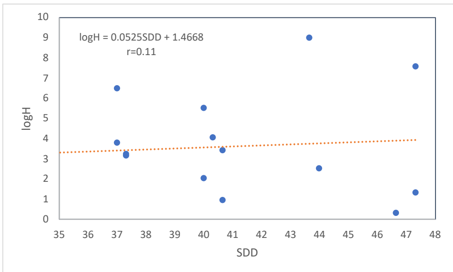
(d) AZI index vs $logH$



(e) H index vs $logH$



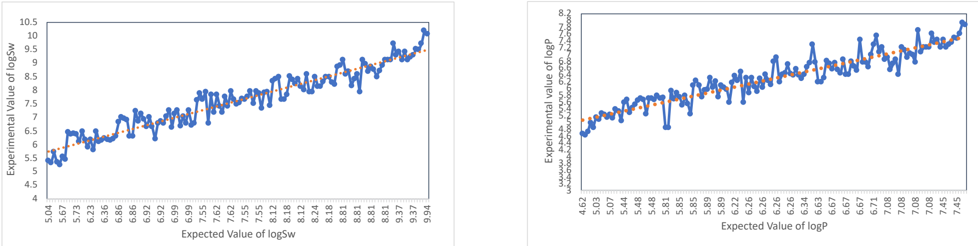
(f) ISI index vs $logH$



(g) SDD index vs $logH$

FIGURE 5.5: Correlation of vertex degree based indices with log Henry constant ($logH$).

vertex degree-based indices have limitations and cannot predict many physicochemical properties, such as Henry's constant. This is observed because inherently, the VDB indices depend only on the structure of the molecules and not on the type of molecules, temperature, and pressure. We also observe from the figures that the data is aligned vertically due to the higher degree of degeneracy (or the discrimination ability is low) shown by the VDB indices on the PCB molecules. However,



(a) $logSw$ (b) $logP$

FIGURE 5.6: Experimental versus Expected Value of $logSw$ and $logP$.

the correlation ability of the VDB indices is quite good on many other considered properties.

In this direction, we obtain that the Symmetric division deg index and the first Zagreb index correlate well from the correlation tables. Further, from Table 5.1 and Table 5.2, we find that the Symmetric division deg index has a better predictive power for log-water solubility and octanol-water partition coefficients, respectively. Moreover, we have shown that the SDD index has better correlation ability with a correlation value of 0.96 for log-water solubility and 0.93 for octanol-water partition coefficient. We have also predicted the values for these two properties of PCB congeners whose experimental values were not available.

****