QSPR Modeling of Anticancer Drugs Using Uphill Topological Indices

Anwar Saleh[1], Sara Bazhear[2,*], Najat Muthana[1]

[1]*Department of Mathematics and Statistics, Faculty of Science, University of Jeddah, Jeddah 21589, Saudi Arabia*

[2]*Department of Mathematical Science, College of Engineering, University of Business and Technology, Jeddah, Saudi Arabia*

*Corresponding author: s.bazhear@ubt.edu.sa

Abstract. Over the past thirty years, cancer has affected more than ten million people worldwide annually. While diverse treatments exist (chemotherapy, surgery, radiation, immunotherapy, stem cell transplants), anti-cancer drugs remain essential. This research introduces five novel uphill topological indices derived from molecular graphs to establish Quantitative Structure-Property Relationships (QSPR) for ten essential anticancer drugs: carmustine, convolutamine F, raloxifene, tambjamine K, pierocellin B, caulibugulone E, convolutamide A, daunorubicin, deguelin, and podophyllotoxin. Power regression analysis correlated these indices with six experimental properties: boiling point (BP), melting point (MP), enthalpy (E), molar refraction (MR), molar volume (MV), and surface tension (ST). Results demonstrate statistically significant correlations ($p < 0.05$), with the uphill sigma index (UPSIG) emerging as the optimal predictor for BP ($R^2 = 0.929$), MP ($R^2 = 0.907$), and E ($R^2 = 0.828$), while the uphill Albertson index (UPAL) excelled for MV ($R^2 = 0.918$). Predictive validity was confirmed via multilinear regression ($R^2 > 0.934$ for BP/E/MV/MR), establishing these indices as powerful tools for rational anticancer drug design.

1. Introduction

Cancer is a life-threatening disease that can develop in individuals of any age or gender and may originate in any part of the body. The human body is composed of countless cells that undergo natural division and growth. However, when certain cells divide abnormally, evade programmed cell death, and invade surrounding tissues, they can form malignant tumors, leading to cancer [1] Chemical graph theory represents an interdisciplinary approach that combines principles from

graph theory and chemistry to analyze molecular structures and establish quantitative structure-property relationships [2]. In this mathematical framework, molecules are represented as graphs where vertices correspond to atoms and edges represent chemical bonds.

A fundamental contribution of chemical graph theory is the development of topological indices - numerical invariants derived from molecular graphs that encode structural information. These indices serve as quantitative descriptors that correlate with various physicochemical properties, biological activities, and reactivity patterns of chemical compounds [3]. Formally, a topological index based on the degree of the vertices $TI(G)$ for a molecular graph G is defined as:

$$TI(G) = \sum_{uv \in E(G)} f(d_u, d_v).$$

where $E(G)$ represents the edge set, d_u, d_v denotes the degrees of the vertices u and v, and f is a suitably chosen function that captures specific structural features [4]. The predictive power of topological indices has been demonstrated in numerous QSAR (Quantitative Structure-Activity Relationship) and QSPR (Quantitative Structure-Property Relationship) studies, making them invaluable tools in computational chemistry and drug design [5–10].

Topological indices provide a quantitative framework for predicting the physicochemical properties of chemical compounds. Formally defined, a topological index is a mathematical function that maps a molecular graph to a non-negative real number. The foundational Wiener index, introduced by Wiener [11, 12], originally termed the path number, established a critical structure-property relationship for alkane boiling points and remains a cornerstone of chemical graph theory.

Subsequent research has generated diverse topological indices grounded in varied mathematical frameworks, with broad applications across chemical graph theory and related disciplines [13–21]. Among these, graph irregularity quantification has emerged as a significant focus in chemical and network modeling. The Albertson irregularity index (Albertson index) has received particular scholarly attention due to its extensive application and theoretical examination [14]. It is defined as:

$$\mathrm{AL}(G) = \sum_{xy \in E(G)} |d_x - d_y|,$$

where d_x and d_y denote vertex degrees. Key theoretical developments for this index are documented in [22–24].

To address computational limitations inherent in the absolute value operation of the Albertson index, the σ-irregularity index (also termed the sum-connectivity index) was developed as an alternative formulation:

$$\sigma(G) = \sum_{xy \in E(G)} (d_x - d_y)^2 \tag{1.1}$$

Gutman et al. [25, 26] established the foundational results for this index. Further extending this paradigm, Alwardi et al. [27] introduced the first and second entire Zagreb indices:

$$M_1^{\mathcal{E}}(G) = \sum_{x \in V \cup E} d_x^2,$$

$$M_2^{\mathcal{E}}(G) = \sum_{\substack{x\, adjacent\, to\, y \\ or\, x\, incident\, to\, y}} d_x d_y.$$

where $L(G)$ denotes the line graph of G. These indices have stimulated significant research activity [28–31], reflecting their utility in the development of structural descriptors.

In [32], same authors have introduced the concept of uphill degree of the vertex in a graph as follows: Let $G = (V, E)$ be a simple undirected graph. For a vertex $v \in V(G)$:

- An **uphill path** from u to v is a path $P = u = v_1, v_2, \dots, v_k = v$, where $\deg(v_i) \le \deg(v_{i+1})$ for all $1 \le i \le k - 1$.
- Vertex v is **uphill adjacent** to u if there exists an uphill path from v to u.
- The **uphill neighborhood** of v is:

$$N_{\mathrm{up}}(v) = \{u \in V(G) \mid v \text{ is uphill adjacent to } u\}.$$

- The **uphill degree** of v is:

$$d_{\mathrm{up}}(v) = |N_{\mathrm{up}}(v)|.$$

Also in [32, 33] same authors introduced and studied the following uphill indices

(1) **First Uphill Zagreb Index**:

$$UPM_1(G) = \sum_{v \in V(G)} \left(d_{\mathrm{up}}(v)\right)^2.$$

(2) **Second Uphill Zagreb Index**:

$$UPM_2(G) = \sum_{uv \in E(G)} d_{\mathrm{up}}(u) \cdot d_{\mathrm{up}}(v).$$

(3) **Modified First Uphill Zagreb Index**:

$$MUPM_1(G) = \sum_{uv \in E(G)} \left(d_{\mathrm{up}}(u) + d_{\mathrm{up}}(v)\right).$$

(4) **Forgotten Uphill Zagreb Index**:

$$UPF(G) = \sum_{v \in V(G)} \left(d_{\mathrm{up}}(v)\right)^3.$$

In this paper, we define five new versions of uphill topological indices and considered nine uphill topological indices to study the QSPR analysis of anti-cancer drugs.

Definition 1.1. *Let $G = (V, E)$ be a graph. Then*

(1) *The modified forgotten uphill index denoted by MUPF(G) is defined as,*

$$MUPF(G) = \sum_{uv \in E(G)} (d_{up}(u)^2 + d_{up}(v)^2).$$

(2) *The uphill hyper index denoted by UPH(G) and defined as*

$$UPH(G) = \sum_{uv \in E(G)} (d_{up}(u) + d_{up}(v))^2.$$

(3) *The uphill Albertson index (UPAL) is defined as*

$$UPAL(G) = \sum_{uv \in E(G)} |d_{up}(u) - d_{up}(v)|.$$

(4) *The uphill sigma index denoted by UPSIG(G) and defined as,*

$$UPSIG(G) = \sum_{uv \in E(G)} (d_{up}(u) - d_{up}(v))^2.$$

(5) *The uphill Geometrical index denoted by UPGA(G) and defined as,*

$$UPGA(G) = \sum_{uv \in E(G)} \frac{2\sqrt{(d_{up}(u) * d_{up}(v))}}{(d_{up}(u) + d_{up}(v))}.$$

The aim of this paper is to check the chemical significance of the newly defined indices and also the regression analysis is made to check the efficiency of these indices to model physico-chemical properties. We use the above defined 9 uphill topological indices to model the physical properties boiling point (BP), melting point (MP), enthalpy (E), molar refraction (MR), molar volume (MV) and surface tension (ST) for the 10 famous anti-cancer drugs.

2. Data and Methods

The experimental physicochemical properties of the ten anticancer drugs, their chemical and molecular structures in Figure 1 and Table 1 , were rigorously sourced from established chemical databases to ensure accuracy and reproducibility. The values for the boiling point (BP), the melting point (MP), the enthalpy (E), the molar refraction (MR), the molar volume (MV) and the surface tension (ST) were curated from PubChem (National Center for Biotechnology Information, accessible via http://pubchem.ncbi.nlm.nih.gov/), and ChemSpider (Royal Society of Chemistry, accessible via http://www.chemspider.com/Search.aspx). These databases provide peer-reviewed, experimentally validated data essential for robust quantitative structure-property relationship (QSPR) modeling. Molecular graphs of each drug were constructed using standard graph-theoretic representations, where atoms correspond to vertices and chemical bonds to edges. Nine uphill topological indices (defined in Section 1) were computed from these graphs to serve as predictive variables, see Table 2. Statistical analyses, including power regression and multi-linear regression, were performed using SPSS21 software.

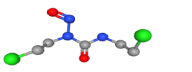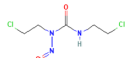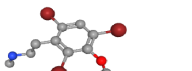FIGURE 1. Chemical and molecular structures of anti cancer drugs
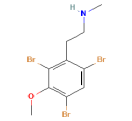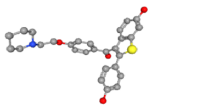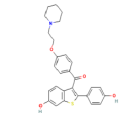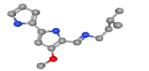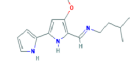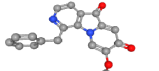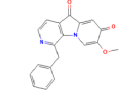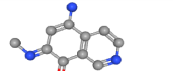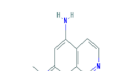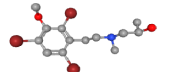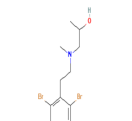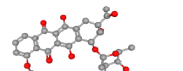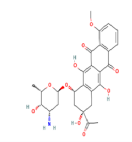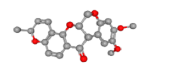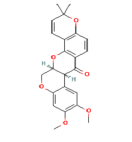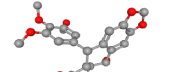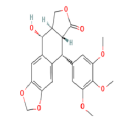
Drug name	Molecular graph	Chemical structure
Carmustine		
Convolutamine F		
Raloxifene		
Tambjamine K		
Pterocellin B		
Caulibugulone E		
Convolutamide A		
Daunorubicin		
Deguelin		
Podophyllotoxin		

TABLE 1. Experimental physicochemical property data for anticancer drugs.

Drugs	BP	MP	E	MR	MV	ST
Carmustine	309.6	120.99	63.8	46.6	146.4	50.4
Convolutamine F	387.7	128.67	63.7	73.8	220.1	41.8
Raloxifene	728.2	289.58	110.1	136.6	367.3	57.3
Tambjamine K	391.7		64.1	76.6	235.1	37
Pterocellin B	521.6	199.88	79.5	87.4	228.3	68
Caulibugulone E	373	129.46	62	52.2	139.1	53.2
Convolutamide A	629.9		97.9	130.1	396	51.9
Daunorubicin	770	208.5	117.6	130	339.4	87.4
Deguelin	560.1	213.39	84.3	105.1	314.2	43.2
Podophyllotoxin	597.9	235.86	93.6	104.3	302.4	52.8

TABLE 2. Uphill topological indices of anti-cancer drugs

Drugs	UPM_1	UPM_2	UPF	MUPF	$MUPM_1$	UPH	UPAL	UPSIG	UPGA
Carmustine	123	90	473	197	59	377	9	17	10.5722
Convolutamine F	452	792	2702	1618	302	3202	22	34	29.8371
Raloxifene	1399	1197	10819	2999	419	5393	93	605	25.8798
Tambjamine K	295	225	1533	583	125	1033	35	133	15.3132
Pterocellin B	704	631	4558	1506	256	2768	50	244	22.9196
Caulibugulone E	472	450	3084	938	160	1838	14	38	14.8505
Convolutamide A	3156	2898	45412	6036	482	11832	44	240	26.9297
Daunorubicin	3517	3564	41883	7718	694	14846	76	590	34.6802
Deguelin	1737	1725	16295	3936	450	7386	72	486	27.8591
Podophyllotoxin	1261	1172	10379	2752	382	5096	68	408	32.0002

3. QSPR Analysis

Quantitative Structure-Property Relationship (QSPR) analysis was employed to evaluate the efficacy of nine uphill topological indices in predicting six critical physicochemical properties of anticancer drugs. Using power regression models of the form $Y = a \cdot X^b$, each index was tested against boiling point (BP), melting point (MP), enthalpy (E), molar refraction (MR), molar volume (MV), and surface tension (ST).

For enhanced predictive accuracy, a multi-linear regression model integrating multiple indices was developed. Model validity was verified through ANOVA testing and residual analysis, ensuring the indices' utility in rational drug design.

3.1. **Power Model Regression Analysis.** This subsection presents a comprehensive power model regression analysis examining the predictive capacity of nine uphill indices for six critical properties of anticancer drugs. For each index, we provide:

(1) Complete statistical parameters quantifying regression performance
(2) The derived power regression equation for each anticancer drug property
(3) A multi-property visualization depicting relationships between the index and all six pharmacological characteristics

The analysis employs power regression models of the form:

$\ln(P) = \ln(a) + b \cdot \ln(X)$ (Power function: $P = a \cdot (X)^b$ where P represents the drug property, X denotes the uphill index, and a, b are regression coefficients. The Power regression equations for physical properties using the 9 uphill indices are shown Tables 3, 5, 7, 9, 11, 13, 15, 17 and 19. The relations between the physical properties of anticancer drugs with the uphill indices are depicted in Figures 2 to 10.

TABLE 3. Power regression equations for physical properties using UPM_1 predictor

Property	Power Regression Equation
Boiling Point (BP)	$BP = \exp(69.2805) \times UPM_1^{0.2969}$
Melting Point (MP)	$MP = \exp(34.9507) \times UPM_1^{0.2464}$
Energy (EN)	$EN = \exp(20.3452) \times UPM_1^{0.2079}$
Molar Refractivity (MR)	$MR = \exp(14.4955) \times UPM_1^{0.2634}$
Molar Volume (MV)	$MV = \exp(29.6538) \times UPM_1^{0.3136}$
Surface Tension (ST)	$ST = \exp(5.0578) \times UPM_1^{0.3151}$

TABLE 4. Power Regression Model Summary and ANOVA Results

Property	R	R^2	Adj. R^2	Std. Error of Estimate	F	p-value
BP	0.933	0.871	0.849	0.127	40.407	0.0007
MP	0.782	0.611	0.546	0.219	9.434	0.0219
EN	0.854	0.729	0.683	0.141	16.113	0.0070
MR	0.678	0.460	0.370	0.318	5.108	0.0645
MV	0.865	0.748	0.706	0.203	17.805	0.0056
ST	0.675	0.456	0.366	0.383	5.033	0.0660

TABLE 5. Power regression equations for physical properties using UPM_2 predictor

Property	Power Regression Equation
Boiling Point (BP)	$BP = \exp(88.8592) \times UPM_2^{0.2604}$
Melting Point (MP)	$MP = \exp(46.9975) \times UPM_2^{0.2027}$
Energy (EN)	$EN = \exp(25.4922) \times UPM_2^{0.1747}$
Molar Refractivity (MR)	$MR = \exp(11.6857) \times UPM_2^{0.2963}$
Molar Volume (MV)	$MV = \exp(35.3040) \times UPM_2^{0.2883}$
Surface Tension (ST)	$ST = \exp(4.3269) \times UPM_2^{0.3392}$

TABLE 6. Power Regression Model Summary and ANOVA Results for UPM_2

Property	R	R²	Adj. R²	Std. Error of Estimate	F	p-value
BP	0.864	0.747	0.705	0.178	17.709	0.006
MP	0.679	0.461	0.372	0.257	5.139	0.064
EN	0.757	0.574	0.503	0.177	8.073	0.030
MR	0.805	0.649	0.590	0.256	11.074	0.016
MV	0.840	0.705	0.656	0.219	14.327	0.009
ST	0.768	0.589	0.521	0.333	8.614	0.026

TABLE 7. Power regression equations for physical properties using UPF predictor

Property	Power Regression Equation
Boiling Point (BP)	$BP = \exp(74.6147) \times UPF^{0.22096}$
Melting Point (MP)	$MP = \exp(38.0254) \times UPF^{0.18074}$
Energy (EN)	$EN = \exp(21.5311) \times UPF^{0.15417}$
Molar Refractivity (MR)	$MR = \exp(15.0146) \times UPF^{0.19959}$
Molar Volume (MV)	$MV = \exp(32.4206) \times UPF^{0.23216}$
Surface Tension (ST)	$ST = \exp(4.9363) \times UPF^{0.24642}$

TABLE 8. Power Regression Model Summary and ANOVA Results for UPF

Property	R	R^2	Adj. R^2	Std. Error of Estimate	F	p-value
BP	0.919	0.845	0.819	0.140	32.664	0.001
MP	0.759	0.576	0.506	0.228	8.159	0.029
EN	0.838	0.702	0.652	0.148	14.136	0.009
MR	0.680	0.462	0.373	0.317	5.160	0.064
MV	0.847	0.718	0.671	0.214	15.273	0.008
ST	0.699	0.489	0.403	0.371	5.732	0.054

TABLE 9. Power regression equations for physical properties using MUPF predictor

Property	Power Regression Equation
Boiling Point (BP)	$BP = \exp(69.9442) \times MUPF^{0.2642}$
Melting Point (MP)	$MP = \exp(37.0176) \times MUPF^{0.2127}$
Energy (EN)	$EN = \exp(21.5142) \times MUPF^{0.1785}$
Molar Refractivity (MR)	$MR = \exp(9.9357) \times MUPF^{0.2860}$
Molar Volume (MV)	$MV = \exp(27.0676) \times MUPF^{0.2926}$
Surface Tension (ST)	$ST = \exp(3.8403) \times MUPF^{0.3186}$

TABLE 10. Power Regression Model Summary and ANOVA Results for MUPF

Property	R	R^2	Adj. R^2	Std. Error of Estimate	F	p-value
BP	0.890	0.792	0.757	0.162	22.792	0.003
MP	0.723	0.523	0.443	0.242	6.572	0.043
EN	0.785	0.616	0.552	0.168	9.638	0.021
MR	0.789	0.622	0.559	0.266	9.868	0.020
MV	0.865	0.747	0.705	0.203	17.756	0.006
ST	0.731	0.535	0.457	0.354	6.902	0.039

TABLE 11. Power regression equations for physical properties using MUPM1 predictor

Property	Power Regression Equation
Boiling Point (BP)	$BP = \exp(59.9162) \times MUPM1^{0.3794}$
Melting Point (MP)	$MP = \exp(31.8395) \times MUPM1^{0.3100}$
Energy (EN)	$EN = \exp(19.6221) \times MUPM1^{0.2540}$
Molar Refractivity (MR)	$MR = \exp(7.3315) \times MUPM1^{0.4348}$
Molar Volume (MV)	$MV = \exp(21.1815) \times MUPM1^{0.4333}$
Surface Tension (ST)	$ST = \exp(3.3752) \times MUPM1^{0.4472}$

TABLE 12. Power Regression Model Summary and ANOVA Results for MUPM1

Property	R	R^2	Adj. R^2	Std. Error of Estimate	F	p-value
BP	0.881	0.776	0.739	0.168	20.820	0.004
MP	0.727	0.528	0.450	0.241	6.721	0.041
EN	0.771	0.594	0.526	0.173	8.778	0.025
MR	0.827	0.684	0.631	0.243	12.986	0.011
MV	0.883	0.779	0.743	0.190	21.202	0.004
ST	0.708	0.501	0.418	0.367	6.035	0.049

TABLE 13. Power regression equations for physical properties using UPH predictor

Property	Power Regression Equation
Boiling Point (BP)	$BP = \exp(59.8327) \times UPH^{0.2627}$
Melting Point (MP)	$MP = \exp(33.5039) \times UPH^{0.2083}$
Energy (EN)	$EN = \exp(19.4497) \times UPH^{0.1769}$
Molar Refractivity (MR)	$MR = \exp(7.9450) \times UPH^{0.2910}$
Molar Volume (MV)	$MV = \exp(22.7757) \times UPH^{0.2909}$
Surface Tension (ST)	$ST = \exp(2.8919) \times UPH^{0.3285}$

TABLE 14. Power Regression Model Summary and ANOVA Results for UPH

Property	R	R²	Adj. R²	Std. Error of Estimate	F	p-value
BP	0.878	0.771	0.733	0.169	20.239	0.004
MP	0.703	0.494	0.410	0.250	5.860	0.052
EN	0.772	0.597	0.530	0.172	8.879	0.025
MR	0.797	0.635	0.574	0.261	10.434	0.018
MV	0.853	0.728	0.683	0.210	16.068	0.007
ST	0.749	0.561	0.487	0.344	7.654	0.033

TABLE 15. Power regression equations for physical properties using UPAL predictor

Property	Power Regression Equation
Boiling Point (BP)	$BP = \exp(138.7414) \times UPAL^{0.3552}$
Melting Point (MP)	$MP = \exp(51.1746) \times UPAL^{0.3482}$
Energy (EN)	$EN = \exp(32.7067) \times UPAL^{0.2519}$
Molar Refractivity (MR)	$MR = \exp(30.9658) \times UPAL^{0.2760}$
Molar Volume (MV)	$MV = \exp(55.3896) \times UPAL^{0.4050}$
Surface Tension (ST)	$ST = \exp(20.4109) \times UPAL^{0.1968}$

TABLE 16. Power Regression Model Summary and ANOVA Results for UPAL

Property	R	R²	Adj. R²	Std. Error of Estimate	F	p-value
BP	0.958	0.917	0.904	0.102	66.632	0.000
MP	0.948	0.899	0.882	0.112	53.179	0.000
EN	0.887	0.787	0.752	0.125	22.210	0.003
MR	0.610	0.372	0.267	0.343	3.549	0.109
MV	0.958	0.918	0.905	0.115	67.343	0.000
ST	0.362	0.131	-0.014	0.484	0.904	0.378

TABLE 17. Power regression equations for physical properties using UPSIG predictor

Property	Power Regression Equation
Boiling Point (BP)	$BP = \exp(171.6777) \times UPSIG^{0.2141}$
Melting Point (MP)	$MP = \exp(63.1561) \times UPSIG^{0.2096}$
Energy (EN)	$EN = \exp(37.4870) \times UPSIG^{0.1547}$
Molar Refractivity (MR)	$MR = \exp(42.2938) \times UPSIG^{0.1375}$
Molar Volume (MV)	$MV = \exp(75.0393) \times UPSIG^{0.2322}$
Surface Tension (ST)	$ST = \exp(23.6222) \times UPSIG^{0.1131}$

TABLE 18. Power Regression Model Summary and ANOVA Results for UPSIG

Property	R	R²	Adj. R²	Std. Error of Estimate	F	p-value
BP	0.964	0.929	0.917	0.094	78.539	0.000
MP	0.953	0.907	0.892	0.107	58.738	0.000
EN	0.910	0.828	0.799	0.112	28.870	0.002
MR	0.507	0.257	0.133	0.373	2.075	0.200
MV	0.917	0.841	0.814	0.161	31.652	0.001
ST	0.347	0.120	-0.026	0.487	0.822	0.400

TABLE 19. Power regression equations for physical properties using UPGA predictor

Property	Power Regression Equation
Boiling Point (BP)	$BP = \exp(72.6367) \times UPGA^{0.6176}$
Melting Point (MP)	$MP = \exp(37.1124) \times UPGA^{0.5059}$
Energy (EN)	$EN = \exp(22.8418) \times UPGA^{0.4062}$
Molar Refractivity (MR)	$MR = \exp(4.9458) \times UPGA^{0.9030}$
Molar Volume (MV)	$MV = \exp(21.3549) \times UPGA^{0.7726}$
Surface Tension (ST)	$ST = \exp(3.4416) \times UPGA^{0.7939}$

TABLE 20. Power Regression Model Summary and ANOVA Results for UPGA

Property	R	R²	Adj. R²	Std. Error of Estimate	F	p-value
BP	0.777	0.603	0.537	0.223	9.123	0.023
MP	0.642	0.413	0.315	0.269	4.216	0.086
EN	0.667	0.445	0.353	0.202	4.819	0.071
MR	0.930	0.865	0.843	0.159	38.477	0.000
MV	0.852	0.727	0.681	0.211	15.958	0.007
ST	0.681	0.463	0.374	0.380	5.183	0.063

We summarize the correlations between the nine uphill indices and the six anticancer drug properties presented in Tables 4, 6, 8, 10, 12, 14, 16, 18 and 20. Bold values indicate the highest correlation for each property. Table 21 summarize the correlations between the nine uphill indices and the six anticancer drug properties.

TABLE 21. Summary of results from power regression model for anticancer drugs.

Property	Predictors	R	Best predictor
Boiling Point (BP)	UPM1	0.933	
	UPM2	0.864	
	UPF	0.919	
	MUPF	0.890	UPSIG
	MUPM1	0.881	
	UPH	0.878	
	UPAL	0.958	
	UPSIG	**0.964**	
	UPGA	0.777	
Melting Point (MP)	UPM1	0.782	
	UPM2	0.679	
	UPF	0.759	
	MUPF	0.723	UPSIG
	MUPM1	0.727	
	UPH	0.703	
	UPAL	0.948	
	UPSIG	**0.953**	
	UPGA	0.642	
	UPM1	0.854	
	UPM2	0.757	

Property	Predictors	R	Best predictor
Energy (E)	UPF	0.838	UPSIG
	MUPF	0.785	
	MUPM1	0.771	
	UPH	0.772	
	UPAL	0.887	
	UPSIG	**0.910**	
	UPGA	0.667	
Molar Refractivity (MR)	UPM1	0.678	UPGA
	UPM2	0.805	
	UPF	0.680	
	MUPF	0.789	
	MUPM1	0.827	
	UPH	0.797	
	UPAL	0.610	
	UPSIG	0.507	
	UPGA	**0.930**	
Molar Volume (MV)	UPM1	0.865	UPAL
	UPM2	0.840	
	UPF	0.847	
	MUPF	0.865	
	MUPM1	0.883	
	UPH	0.853	
	UPAL	**0.958**	
	UPSIG	0.917	
	UPGA	0.852	
Surface Tension (ST)	UPM1	0.675	UPH
	UPM2	0.768	
	UPF	0.699	
	MUPF	0.731	
	MUPM1	0.708	
	UPH	**0.749**	
	UPAL	0.362	
	UPSIG	0.347	
	UPGA	0.681	

The correlations between the physicochemical properties of the 10 anticancer drugs and the uphill indices are graphically depicted in Figures 2 to 10.
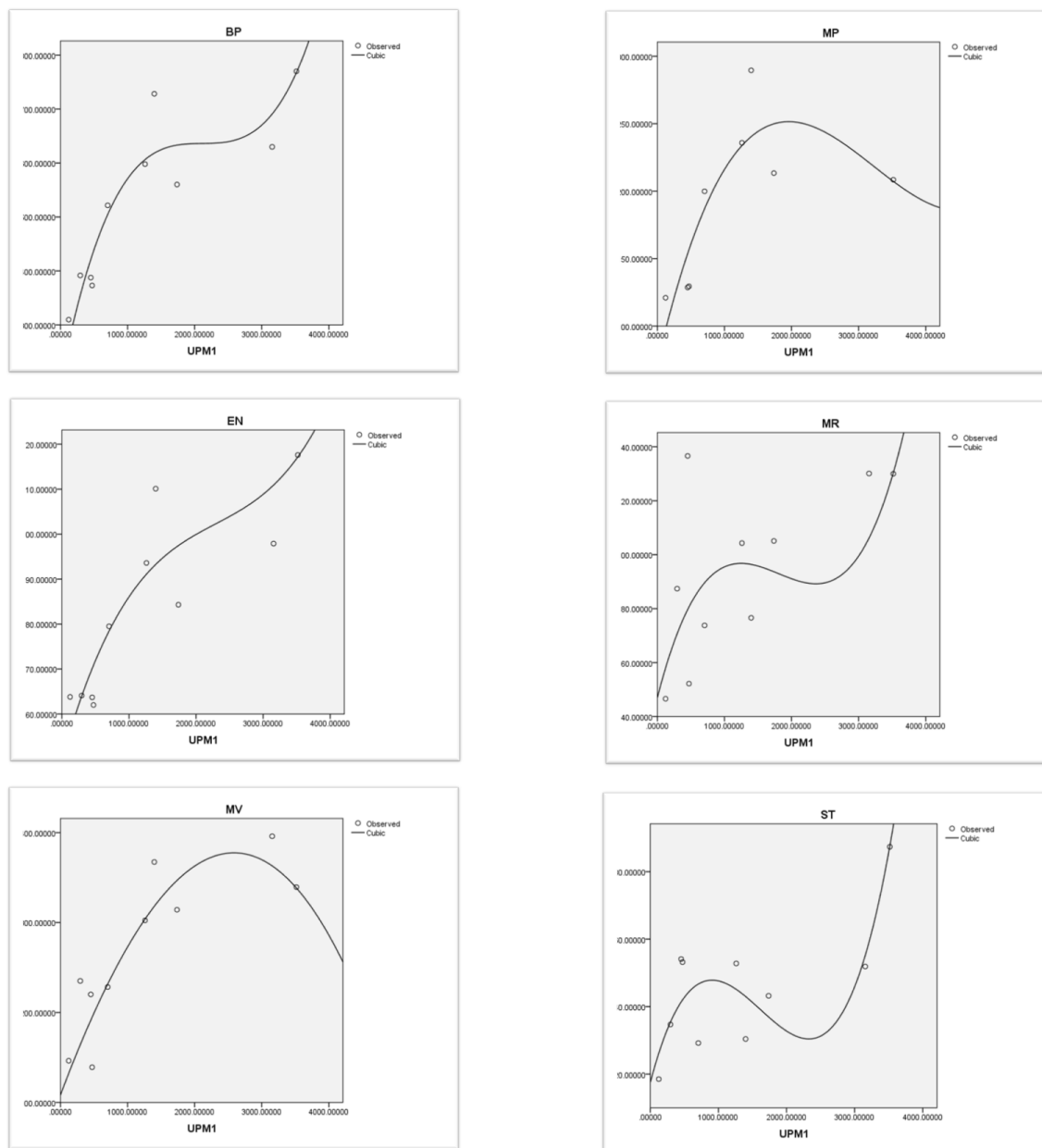

FIGURE 2. Correlations of the physical properties with UPM_1 index for anticancer drugs.
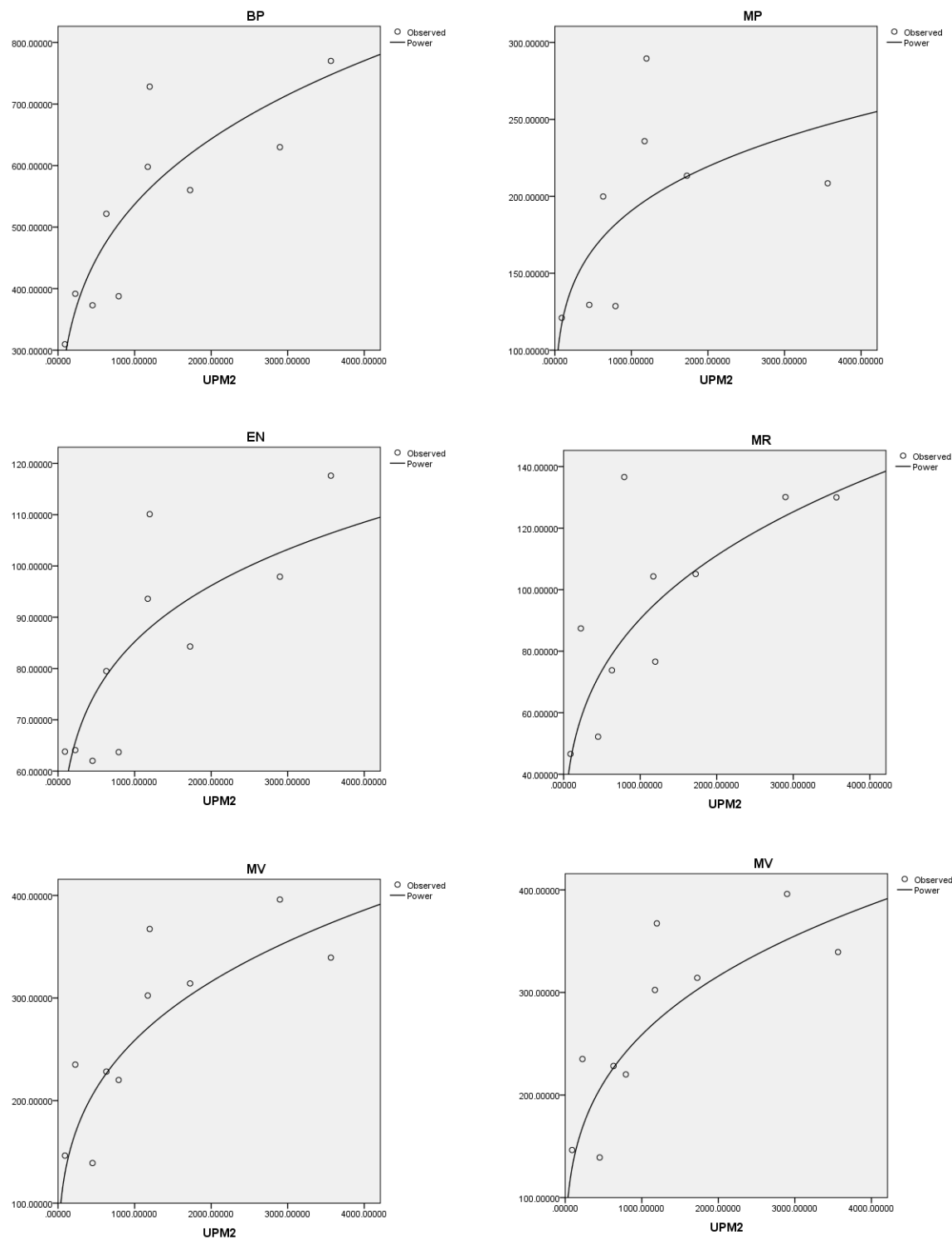
FIGURE 3. Correlations of the physical properties with UPM_2 index for anticancer drugs.

FIGURE 4. Correlations of the physical properties with *UPF* index for anticancer drugs.

FIGURE 5. Correlations of the physical properties with *MUPF* index for anticancer drugs.

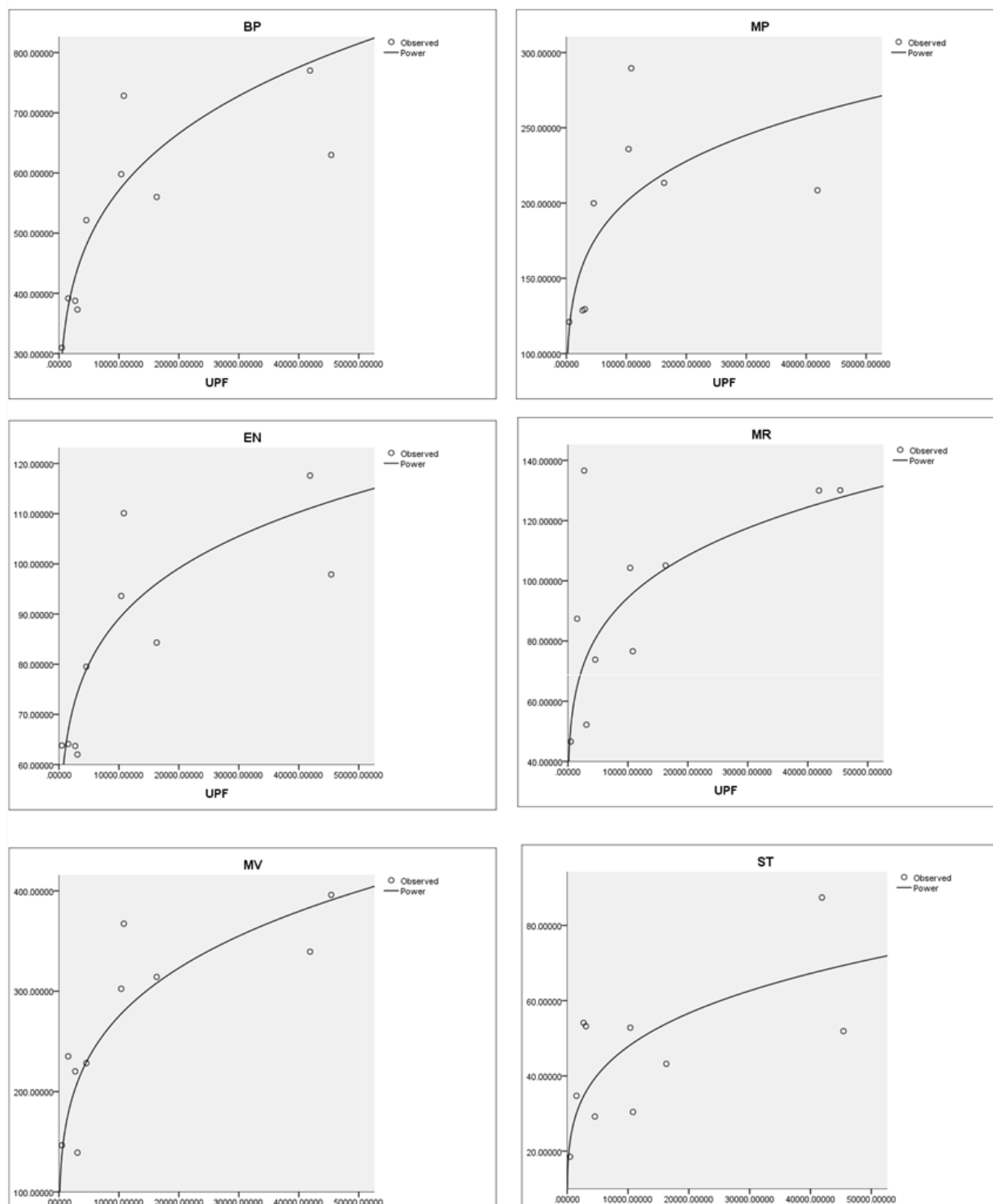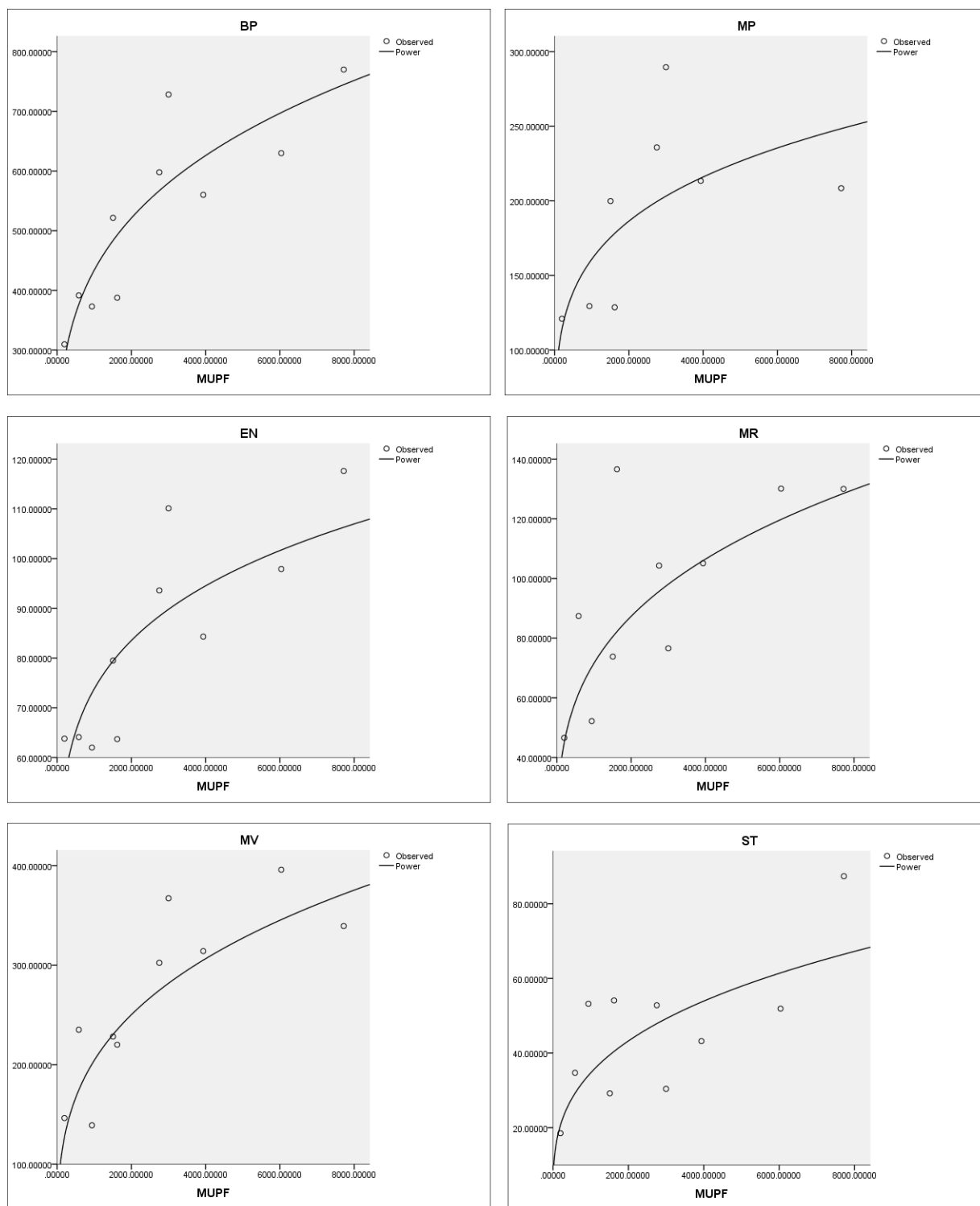FIGURE 6. Correlations of the physical properties with $MUPM_1$ index for anticancer drugs.

FIGURE 7. Correlations of the physical properties with *UPH* index for anticancer drugs.

FIGURE 8. Correlations of the physical properties with *UPAL* index for anticancer drugs.

FIGURE 9. Correlations of the physical properties with *UPSIG* index for anticancer drugs.

FIGURE 10. Correlations of the physical properties with *UPGA* index for anticancer drugs.
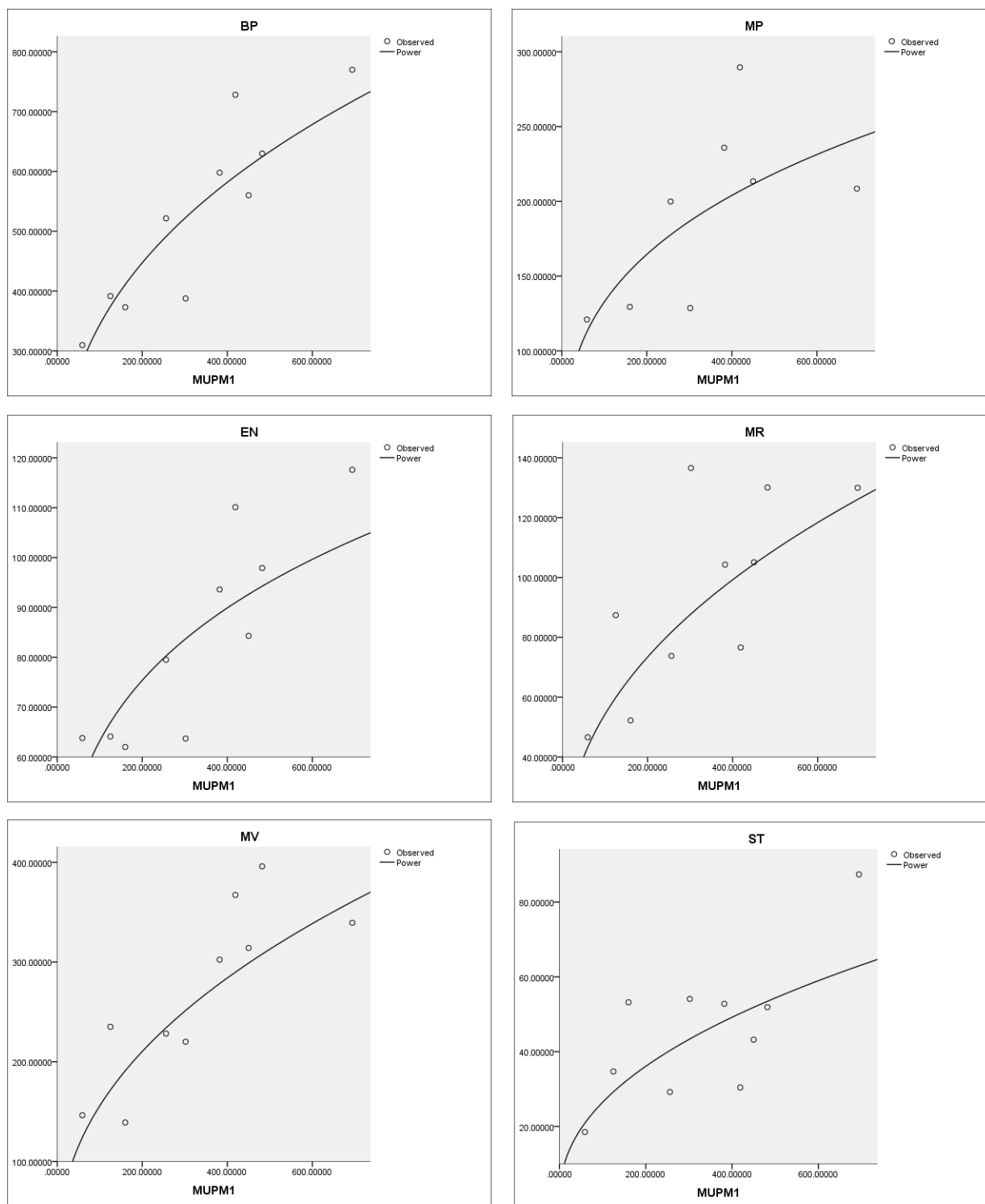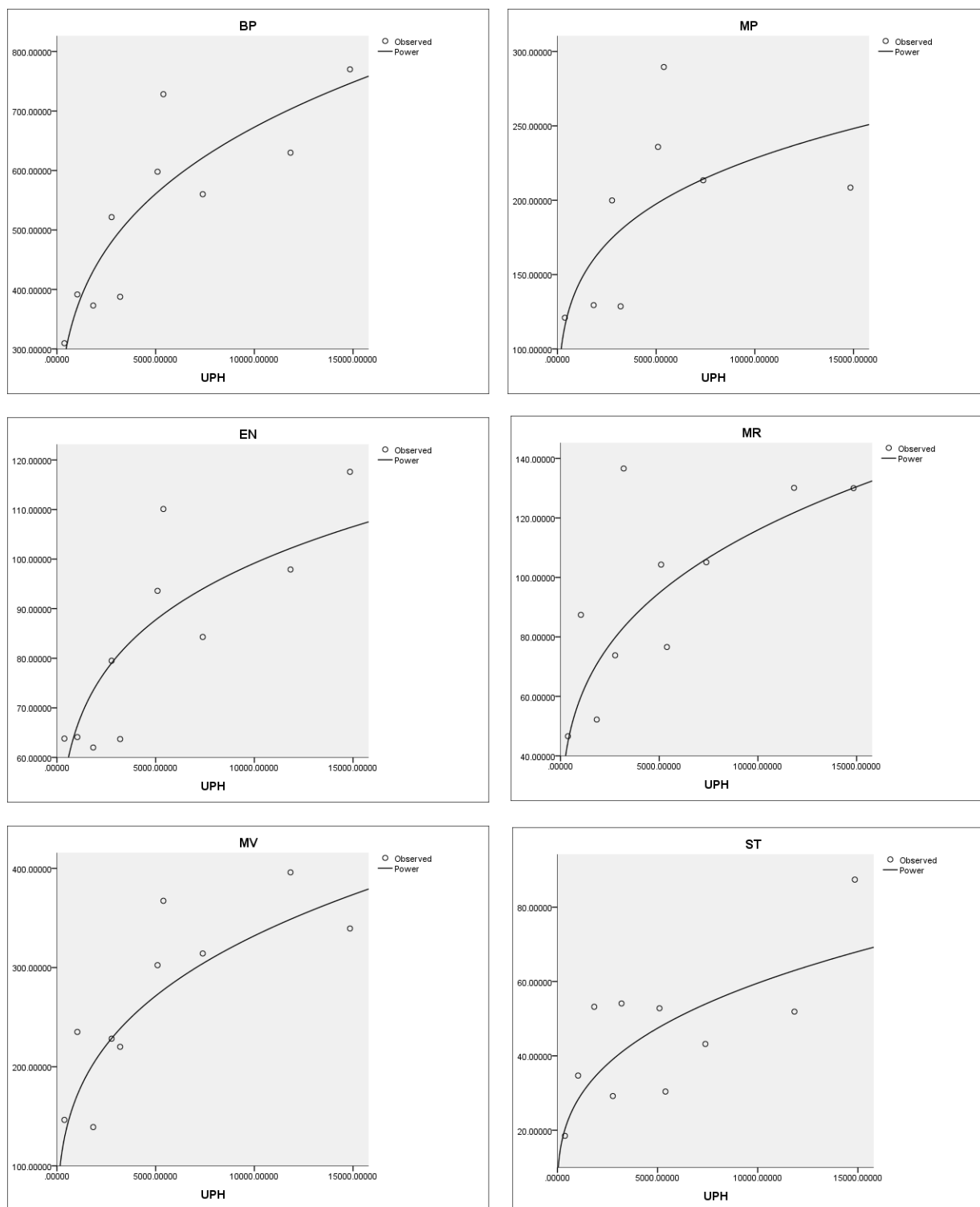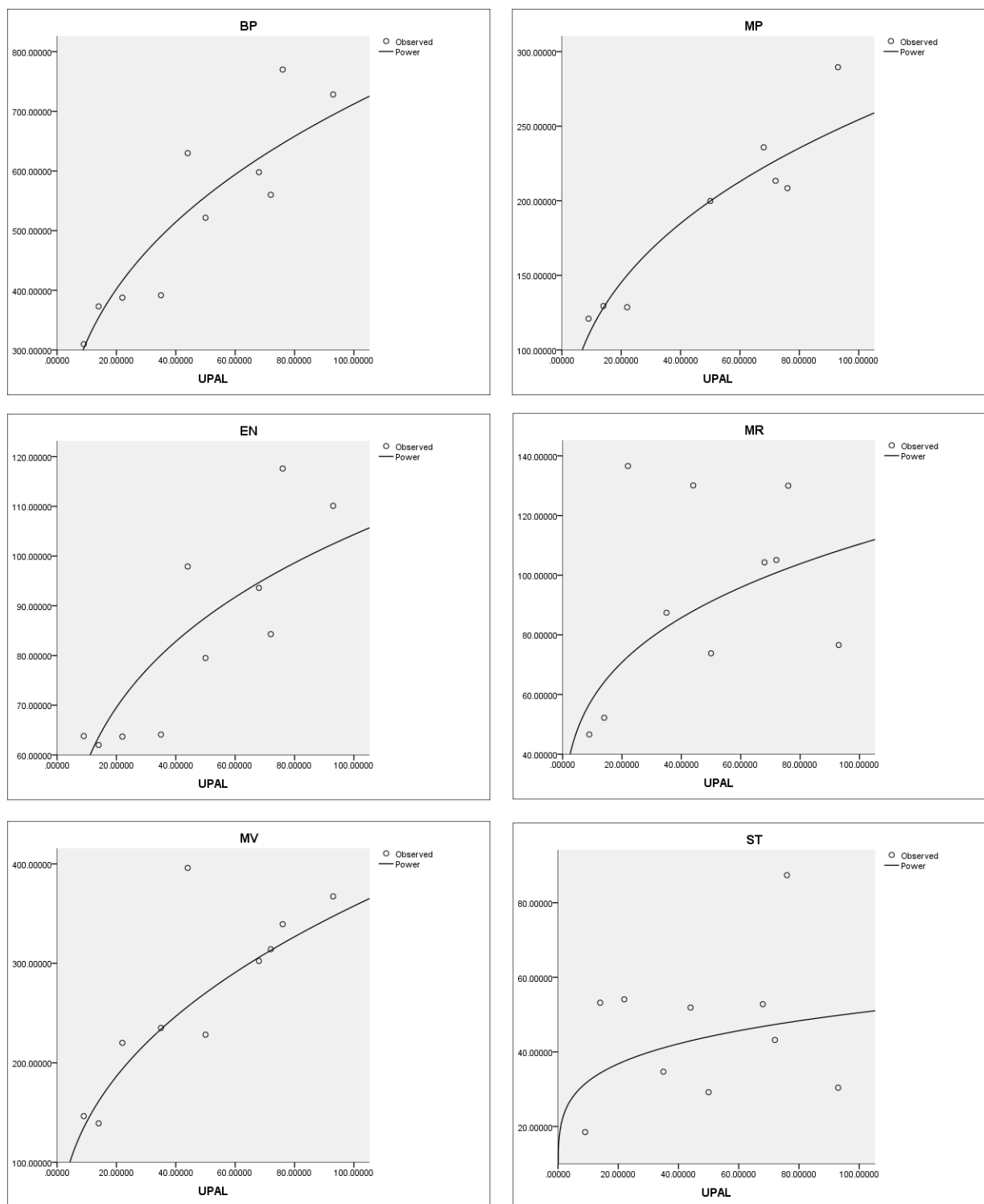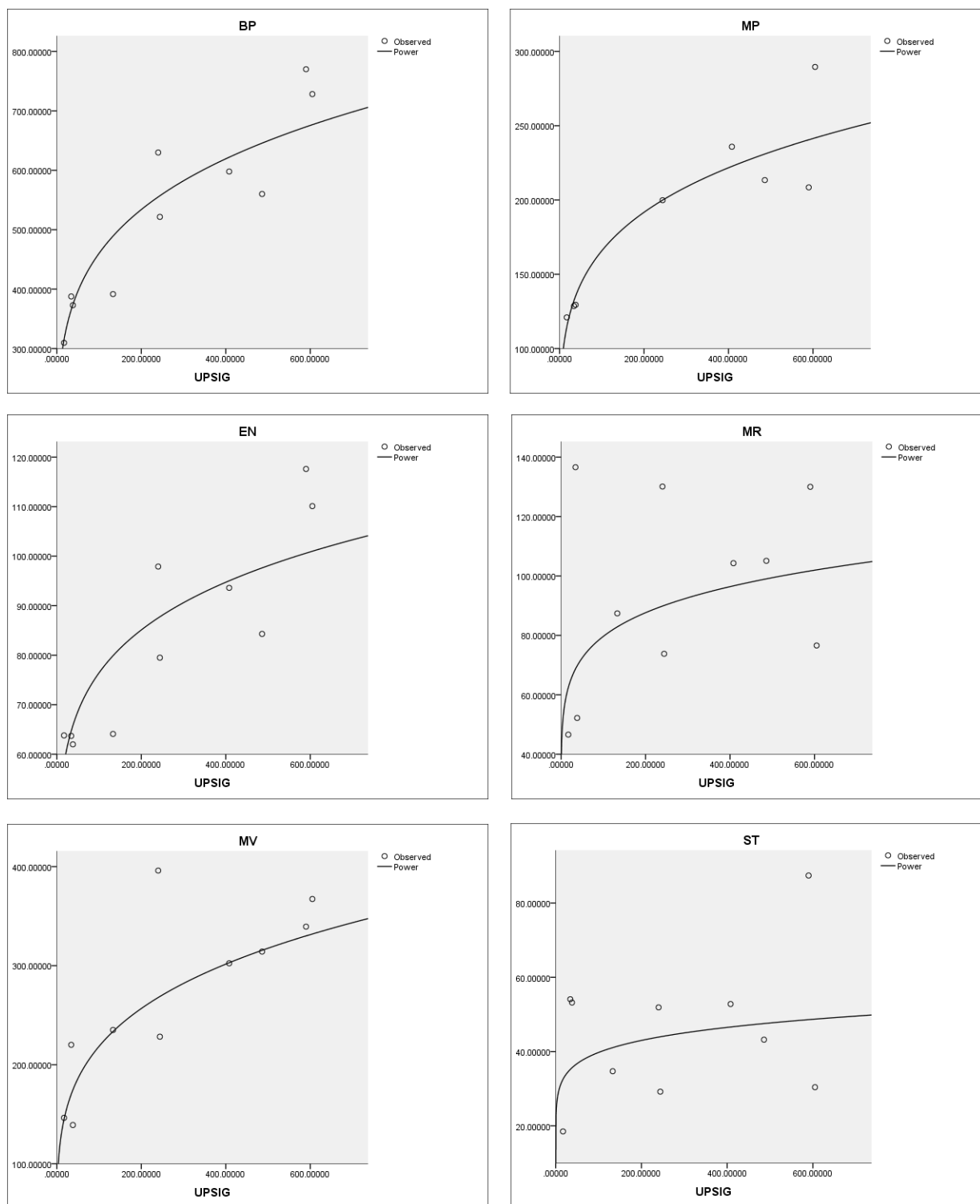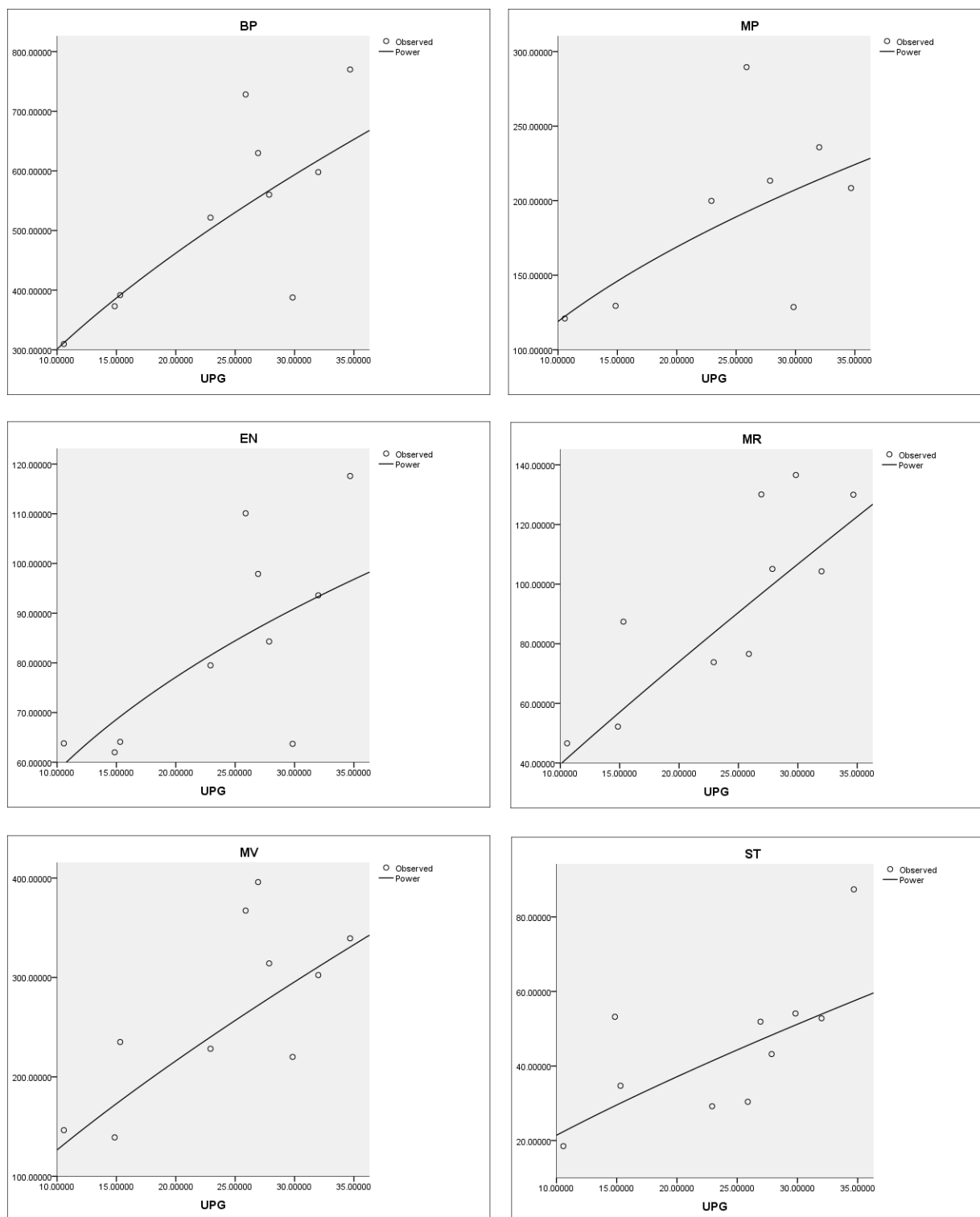
3.2. **Multi-linear Regression Model.** This subsection details a quantitative structure-property relationship (QSPR) analysis employing multi-linear regression to enhance predictive accuracy. The model incorporates five molecular descriptors: UPM_1, UPM_2, UPF, $MUPF$, and $MUPM_1$. For all anticancer drug properties examined, the regression analysis yielded strong statistically significant correlations ($p < 0.05$), as evidenced by Tables 22 and 23. Comparative between the multi-linear prediction and experimental values of the properties for anti cancer drugs is shown in Figure 11.

TABLE 22. Multi-linear Regression Equations

Property	Regression Equation
ST	$ST = 15.2598 + 0.1259 \cdot UPM1 + 0.3073 \cdot UPM2 - 0.0079 \cdot UPF - 0.1470 \cdot MUPF - 0.0075 \cdot MUPM1$
MV	$MV = 153.1277 - 0.5290 \cdot UPM1 - 1.1674 \cdot UPM2 + 0.0391 \cdot UPF + 0.4981 \cdot MUPF + 1.0579 \cdot MUPM1$
MR	$MR = 56.7307 - 0.3058 \cdot UPM1 - 0.1285 \cdot UPM2 + 0.0141 \cdot UPF + 0.0912 \cdot MUPF + 0.4528 \cdot MUPM1$
EN	$EN = 55.3563 - 0.0045 \cdot UPM1 - 0.1419 \cdot UPM2 + 0.0017 \cdot UPF + 0.0590 \cdot MUPF + 0.0737 \cdot MUPM1$
BP	$BP = 250.3262 + 0.1827 \cdot UPM1 - 0.7611 \cdot UPM2 + 0.0064 \cdot UPF + 0.2048 \cdot MUPF + 1.0190 \cdot MUPM1$

TABLE 23. Multilinear Regression Model Summary and ANOVA Results

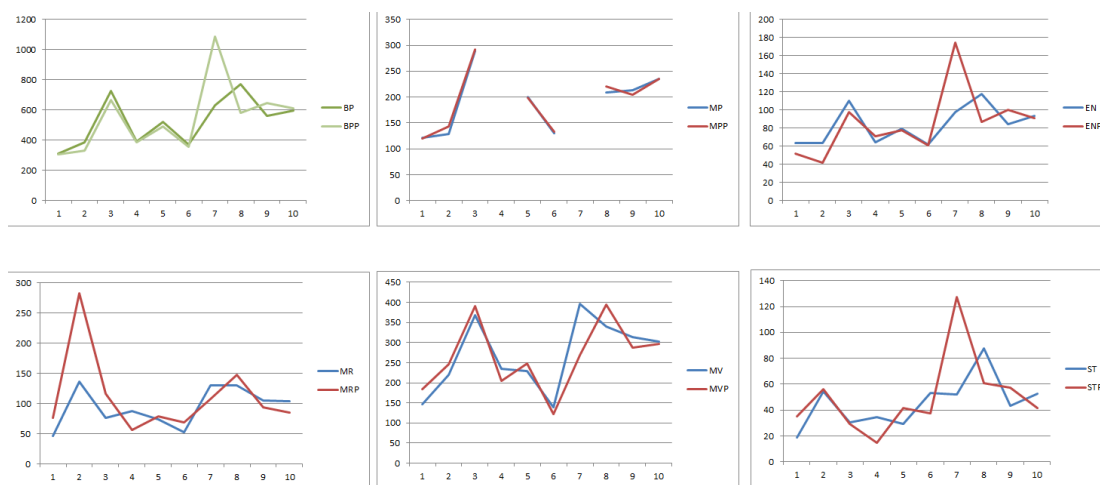| Dependent	R	R^2	Adj. R^2	Std. Error	F	p-value
ST	0.925	0.8558	0.6755	10.9542	4.7465	0.078
MV	0.990	0.9805	0.9560	18.6216	40.1547	0.002
MR	0.946	0.8941	0.7618	15.7665	6.7556	0.044
EN	0.968	0.9371	0.8585	7.7444	11.9214	0.016
BP	0.985	0.9710	0.9348	40.3535	26.8241	0.004

FIGURE 11. Comparative between the multi-linear prediction and experimental values of the properties for anti cancer drugs.

3.3. **Residual Analysis.** Residual analysis serves as a critical diagnostic tool for assessing the performance and reliability of regression models in predicting physicochemical properties. The residuals, defined as the differences between experimental and predicted values (Residual = Experimental – Predicted), provide insights into model accuracy, bias, and potential limitations.

3.3.1. *Comparative Analysis of Experimental and Predicted Values.* Table 24 presents a comprehensive comparison between experimental and predicted values across five physicochemical properties. The multi-linear regression models demonstrate varying degrees of predictive accuracy across different properties and compounds.

TABLE 24. Comparison of experimental and predicted physicochemical properties for anticancer drugs using multi-linear regression.

Drug	Surface Tension (ST)		Molar Volume (MV)		Molar Refractivity (MR)	
	Exp	Pred	Exp	Pred	Exp	Pred
Carmustine	50.4	25.3	146.4	162.0	46.6	58.9
Convolutamine F	41.8	54.1	220.1	220.5	73.8	139.1
Raloxifene	57.3	29.8	367.3	375.8	136.6	90.9
Tambjamine K	37.0	22.8	235.1	217.0	76.6	69.0
Pterocellin B	68.0	38.5	228.3	243.3	87.4	77.9
Caulibugulone E	53.2	49.5	139.1	135.2	52.2	56.0
Convolutamide A	51.9	53.5	396.0	392.5	130.1	128.3
Daunorubicin	87.4	82.6	339.4	348.2	130.0	131.9
Deguelin	43.2	53.3	314.2	294.2	105.1	96.4
Podophyllotoxin	52.8	44.8	302.4	298.6	104.3	90.8

Drug	Energy (E)		Boiling Point (BP)	
	Exp	Pred	Exp	Pred
Carmustine	63.8	58.8	309.6	307.8
Convolutamine F	63.7	63.3	387.7	386.5
Raloxifene	110.1	105.4	728.2	705.3
Tambjamine K	64.1	68.3	391.7	389.6
Pterocellin B	79.5	78.1	521.6	497.2
Caulibugulone E	62.0	61.8	373.0	368.9
Convolutamide A	97.9	98.8	629.9	639.2
Daunorubicin	117.6	111.5	770.0	736.2
Deguelin	84.3	95.9	560.1	623.7
Podophyllotoxin	93.6	91.5	597.9	608.0

For surface tension (ST), the model shows particularly strong predictive capability for Caulibugulone E (Residual = 3.68) and Convolutamide A (Residual = -1.59), while exhibiting larger discrepancies for Pterocellin B (Residual = 29.51) and Raloxifene (Residual = 27.53). The

molar volume (MV) predictions display excellent overall accuracy, with residuals generally below 20 units across all compounds, indicating the model's robustness in predicting this property.

3.3.2. *Residual Distribution and Model Performance.* The detailed residual values presented in Table 25 reveal several important patterns in model performance. The energy (E) predictions demonstrate the highest accuracy among all properties, with residuals ranging from -11.55 to 6.09 and a mean residual of 0.33, indicating minimal systematic bias. In contrast, molar refractivity (MR) predictions show the widest variation, with residuals spanning from -65.34 to 45.72, suggesting potential limitations in the model's ability to capture the structural features influencing this property.

TABLE 25. Residual values (Experimental - Predicted) for physicochemical properties of anticancer drugs.

Drug	Res_ST	Res_MV	Res_MR	Res_E	Res_BP
Carmustine	25.136	-15.631	-12.303	4.993	1.807
Convolutamine F	-12.291	-0.399	-65.342	0.450	1.187
Raloxifene	27.534	-8.464	45.718	4.680	22.915
Tambjamine K	14.206	18.122	7.608	-4.217	2.140
Pterocellin B	29.510	-14.961	9.504	1.381	24.443
Caulibugulone E	3.680	3.924	-3.846	0.246	4.054
Convolutamide A	-1.594	3.473	1.825	-0.876	-9.327
Daunorubicin	4.759	-8.765	-1.933	6.091	33.795
Deguelin	-10.143	19.999	8.724	-11.553	-63.609
Podophyllotoxin	8.028	3.826	13.489	2.059	-10.095

The residual analysis summary in Table 26 quantifies model performance through mean absolute error (MAE) and root mean square error (RMSE) metrics. The energy (E) model achieves the lowest error values (MAE = 3.65, RMSE = 4.92), confirming its superior predictive capability. Surface tension (ST) and boiling point (BP) models show moderate performance with MAE values of 13.69 and 17.34, respectively, while molar refractivity (MR) exhibits the highest prediction errors (MAE = 17.03, RMSE = 26.34).

TABLE 26. Residual analysis summary for physicochemical property predictions using multi-linear regression.

Property	MAE	RMSE
ST	13.6882	16.7921
MV	9.7564	11.7782
MR	17.0293	26.3425
E	3.6545	4.9230
BP	17.3374	25.5449

3.3.3. *Statistical Characteristics of Residuals.* Table 27 provides a statistical summary of the residual distributions, offering insights into model bias and variability. The near-zero mean residuals for all properties (ranging from 0.11 to 8.88) indicate that the regression models are generally unbiased, with no systematic over- or under-prediction tendencies. However, the substantial standard deviations, particularly for molar refractivity (27.77) and boiling point (26.92), highlight significant variability in prediction accuracy across different molecular structures.

TABLE 27. Statistical summary of residuals for physicochemical property predictions.

Property	Mean	Std Dev	Min	Max
ST	8.8825	15.0213	-12.2914	29.5105
MV	0.1125	12.4147	-15.6308	19.9992
MR	0.3443	27.7651	-65.3425	45.7181
E	0.3254	5.1779	-11.5528	6.0909
BP	0.7311	26.9157	-63.6094	33.7947

The residual ranges reveal important outliers in the dataset. For instance, Deguelin shows a substantial negative residual for boiling point (-63.61), while Raloxifene exhibits a large positive residual for molar refractivity (45.72). These outliers may indicate either limitations in the current set of topological indices or the presence of unique structural features not adequately captured by the regression models.

The residual analysis demonstrates that multi-linear regression models using uphill topological indices provide reasonable predictive capability for anticancer drug properties, though with varying accuracy across different properties. The energy (E) model shows superior performance (MAE = 3.65), indicating these indices effectively capture electronic characteristics. Surface tension (ST) and molar volume (MV) predictions show moderate accuracy, while molar refractivity (MR) and boiling point (BP) exhibit higher errors, suggesting limitations in modeling polarizability and complex intermolecular interactions.

Specific outliers, such as Deguelin's large BP residual (-63.61) and Raloxifene's MR residual (45.72), highlight structural features not adequately captured by current descriptors. The linear regression approach and limited descriptor set may oversimplify complex molecular interactions. Future improvements should incorporate additional electronic descriptors, explore non-linear modeling techniques, and expand the dataset diversity.

Despite these limitations, the models offer practical utility for initial drug screening, with the residual analysis framework providing clear guidance for ongoing model refinement in QSPR studies.

Discussion

The present investigation introduced five novel uphill topological indices modified forgotten uphill index (MUPF), uphill sigma index (UPSIG), uphill hyper index (UPH), uphill geometrical index (UPGA), and uphill Albertson index (UPAL) alongside four established uphill indices, to evaluate their efficacy in predicting six critical physicochemical properties of ten anticancer drugs. Power regression analysis revealed statistically significant correlations ($p < 0.05$) between these indices and key properties, including boiling point (BP), melting point (MP), enthalpy (E), molar refraction (MR), molar volume (MV), and surface tension (ST). The uphill sigma index (UPSIG) emerged as the optimal predictor for BP ($R^2 = 0.929$), MP ($R^2 = 0.907$), and E ($R^2 = 0.828$), while UPAL demonstrated superior performance for MV ($R^2 = 0.918$). These outcomes underscore the capability of uphill indices to encode structural features of molecular graphs that govern physicochemical behaviors, particularly thermodynamic and bulk properties.

The robustness of these indices was further validated through multi-linear regression models incorporating multiple descriptors (UPM$_1$, UPM$_2$, UPF, MUPF, MUPM$_1$). These models achieved enhanced predictive accuracy, with $R^2 > 0.934$ for BP, E, MV, and MR, highlighting synergistic effects among indices. For instance, the model for BP ($R^2 = 0.971$) significantly outperformed single-index power regressions. This aligns with prior studies emphasizing combinatorial descriptors for comprehensive QSPR modeling. Notably, UPSIG's dominance for energy-related properties (BP, MP, E) likely stems from its formulation summing squared differences in uphill degrees which amplifies sensitivity to structural irregularities influencing thermal stability. Conversely, UPAL's strength for MV, defined via absolute differences in uphill degrees, reflects its utility in modeling steric properties.

However, surface tension (ST) exhibited weaker correlations ($R^2 < 0.6$) across all indices, with UPH ($R^2 = 0.561$) yielding the highest but still moderate fit. This suggests that ST is influenced by factors beyond topological structure (e.g., intermolecular forces or polarity), which degree-based indices may not fully capture. Similar limitations were observed for MR, where UPGA ($R^2 = 0.865$) outperformed uphill indices but underscored the need for hybrid descriptors integrating electronic parameters.

Conclusion

This study defined five novel uphill topological indices (MUPF, UPSIG, UPH, UPGA, UPAL) and applied nine uphill indices to establish QSPR models for ten anticancer drugs. Power regression identified UPSIG as the optimal predictor for BP, MP, and E, and UPAL for MV, validating their utility in quantifying molecular graph features linked to physicochemical properties. Multi-linear regression further enhanced predictive power ($R^2 > 0.934$ for BP/E/MV/MR), demonstrating the synergy of combinatorial descriptors. These results affirm uphill indices as robust tools for rational drug design, enabling efficient prediction of key properties without resource-intensive

experimentation. This work advances graph-theoretic approaches in computational oncology, offering a pathway to accelerate anticancer drug discovery and optimization.

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] M.A. Bright, The National Cancer Institute's Cancer Information Service: A Premiere Cancer Information and Education Resource for the Nation, J. Cancer Educ. 22 (2007), S2–S7. https://doi.org/10.1007/bf03174340.

[2] N. Trinajstic, Chemical Graph Theory, CRC Press, (2018). https://doi.org/10.1201/9781315139111.

[3] J.C. Dearden, The Use of Topological Indices in QSAR and QSPR Modeling, in: K. Roy, (eds) Advances in QSAR Modeling. Challenges and Advances in Computational Chemistry and Physics, Vol. 24, Springer, Cham, (2017). https://doi.org/10.1007/978-3-319-56850-8_2

[4] R. Todeschini, V. Consonni, Molecular Descriptors for Chemoinformatics, Wiley, 2009. https://doi.org/10.1002/9783527628766.

[5] M.S. Sardar, M.S. Iqbal, M.M. Hassan, C. Bu, S. Hussain, Improved QSAR Methods for Predicting Drug Properties Utilizing Topological Indices and Machine Learning Models, Eur. Phys. J. E 48 (2025), 25. https://doi.org/10.1140/epje/s10189-025-00491-6.

[6] H.M. Yasin, M. Suresh, J.H.H. Bayati, Topological Indices and QSPR/QSAR Analysis of Some Drugs Being Investigated for the Treatment of Alzheimer's Disease Patients, Baghdad Sci. J. 22 (2025), 242–272. https://doi.org/10.21123/bsj.2024.10866.

[7] M. Hasani, M. Ghods, Topological Indices and QSPR Analysis of Some Chemical Structures Applied for the Treatment of Heart Patients, Int. J. Quantum Chem. 124 (2023), e27234. https://doi.org/10.1002/qua.27234.

[8] N.U.H. Awan, A. Ghaffar, F.M. Tawfiq, G. Mustafa, M. Bilal, et al., QSPR Analysis for Physiochemical Properties of New Potential Antimalarial Compounds Involving Topological Indices, Int. J. Quantum Chem. 124 (2024), e27391. https://doi.org/10.1002/qua.27391.

[9] M. Shanmukha, N. Basavarajappa, K. Shilpa, A. Usha, Degree-Based Topological Indices on Anticancer Drugs with QSPR Analysis, Heliyon 6 (2020), e04235. https://doi.org/10.1016/j.heliyon.2020.e04235.

[10] X. Shi, R. Cai, J. Ramezani Tousi, A.A. Talebi, Quantitative Structure–Property Relationship Analysis in Molecular Graphs of Some Anticancer Drugs with Temperature Indices Approach, Mathematics 12 (2024), 1953. https://doi.org/10.3390/math12131953.

[11] H. Wiener, Structural Determination of Paraffin Boiling Points, J. Am. Chem. Soc. 69 (1947), 17–20. https://doi.org/10.1021/ja01193a005.

[12] D.H. Rouvray, The Rich Legacy of Half a Century of the Wiener Index, in: Topology in Chemistry, Elsevier, 2002: pp. 16–37. https://doi.org/10.1016/b978-1-898563-76-1.50006-8.

[13] I. Gutman, Degree-Based Topological Indices, Croat. Chem. Acta 86 (2013), 351–361. https://doi.org/10.5562/cca2294.

[14] M.O. Albertson, The Irregularity of a Graph, Ars Comb. 46 (1997), 219–225.

[15] A. Alqesmah, A. Saleh, R. Rangarajan, A.Y. Gunes, I.N. Cangul, Distance Eccentric Connectivity Index of Graphs, Kyungpook Math. J. 61 (2021), 61–74. https://doi.org/10.5666/KMJ.2021.61.1.61.

[16] S. Hayat, A. Arif, L. Zada, A. Khan, Y. Zhong, Mathematical Properties of a Novel Graph-Theoretic Irregularity Index with Potential Applicability in QSPR Modeling, Mathematics 10 (2022), 4377. https://doi.org/10.3390/math10224377.

[17] X. Shi, R. Cai, J. Ramezani Tousi, A.A. Talebi, Quantitative Structure–Property Relationship Analysis in Molecular Graphs of Some Anticancer Drugs with Temperature Indices Approach, Mathematics 12 (2024), 1953. https://doi.org/10.3390/math12131953.

[18] S. Hayat, F. Asmat, Sharp Bounds on the Generalized Multiplicative First Zagreb Index of Graphs with Application to QSPR Modeling, Mathematics 11 (2023), 2245. https://doi.org/10.3390/math11102245.

[19] M.K. Jamil, M. Imran, K. Abdul Sattar, Novel Face Index for Benzenoid Hydrocarbons, Mathematics 8 (2020), 312. https://doi.org/10.3390/math8030312.

[20] M.W. Rasheed, A. Mahboob, I. Hanif, Uses of Degree-Based Topological Indices in QSPR Analysis of Alkaloids with Poisonous and Healthful Nature, Front. Phys. 12 (2024), 1381887. https://doi.org/10.3389/fphy.2024.1381887.

[21] F. Bashir Farooq, N. Ul Hassan Awan, S. Parveen, N. Idrees, S. Kanwal, et al., Topological Indices of Novel Drugs Used in Cardiovascular Disease Treatment and Its QSPR Modeling, J. Chem. 2022 (2022), 9749575. https://doi.org/10.1155/2022/9749575.

[22] H. Abdo, N. Cohen, D. Dimitrov, Graphs with Maximal Irregularity, Filomat 28 (2014), 1315–1322. https://doi.org/10.2298/fil1407315a.

[23] I. Gutman, P. Hansen, H. Mélot, Variable Neighborhood Search for Extremal Graphs. 10. Comparison of Irregularity Indices for Chemical Trees, J. Chem. Inf. Model. 45 (2005), 222–230. https://doi.org/10.1021/ci0342775.

[24] P. Hansen, H. Mélot, Variable Neighborhood Search for Extremal Graphs. 9. Bounding the Irregularity of a Graph, DIMACS Ser. Discr. Math. Theor. Comput. Sci. 69 (2005), 253–264

[25] I. Gutman, M. Togan, A. Yurttas, et al., Inverse Problem for Sigma Index, MATCH Commun. Math. Comput. Chem. 79 (2018) 491–508.

[26] A. Jahanbanı, S. Ediz, The Sigma Index of Graph Operations, Sigma J. Eng. Nat. Sci. 37 (2019), 155–162.

[27] A. Alwardi, A. Alqesmah, R. Rangarajan, I.N. Cangul, Entire Zagreb Indices of Graphs, Discret. Math. Algorithms Appl. 10 (2018), 1850037. https://doi.org/10.1142/s1793830918500374.

[28] A. Alqesmah, K.A.A. Alloush, A. Saleh, G. Deepak, Entire Harary Index of Graphs, J. Discret. Math. Sci. Cryptogr. 25 (2021), 2629–2643. https://doi.org/10.1080/09720529.2021.1888435.

[29] A. Saleh, S.H. Alsulami, On the Entire Harmonic Index and Entire Harmonic Polynomial of Graphs, Symmetry 16 (2024), 208. https://doi.org/10.3390/sym16020208.

[30] A. Saleh, A. Aqeel, I.N. Cangul, On the Entire ABC Index of Graphs, Proc. Jangjeon Math. Soc. 23 (2020), 39–51.

[31] A. Saleh, I.N. Cangul, On the Entire Randic Index of Graphs, Adv. Appl. Math. Sci. 20 (2021), 1559–1569.

[32] A. Saleh, S. Bazhear, N. Muthana, On the Uphill Zagreb Indices of Graphs, Int. J. Anal. Appl. 20 (2022), 6. https://doi.org/10.28924/2291-8639-20-2022-6.

[33] A. Saleh, S. Bazhear, N. Muthana, Uphill Zagreb Indices of Some Graph Operations for Certain Graphs, J. Appl. Math. Inform. 40 (2022), 959–977.