

**Analisis Hubungan Kuantitatif Struktur dan Kelarutan
Senyawa Aktif Pestisida Organofosfat: Pendekatan Model Linear
dan Metode Kluster**
*(Analysis of Quantitative Structure and Solubility Relationship for
Organophosphate Active Compounds Linear Model and Cluster Model
Approach)*

Is Fatimah¹⁾ dan Jaka Nugraha²⁾

¹⁾ Staf Pengajar Program Studi Ilmu Kimia FMIPA Universitas Islam Indonesia

²⁾ Staf Pengajar Program Studi Statistika FMIPA Universitas Islam Indonesia

ABSTRACT

Study on quantitative solubility and structure relationship of organophosphate pesticide has been conducted. Aim of this research is to analyze relationship between characteristic physical properties (water solubility) of pesticide with molecular structure by a model that can be used to predict physical properties of new compound proposed to be pesticide compound. Analysis was done for 52 organophosphate compounds and its water solubility data. The model predicted by 39 fragmental descriptor based on specific structure of organophosphate compound. Analysis used in modeling are cluster analysis, one way analysis of variance (ANOVA) and regression analysis. Result show water solubility of the pesticide compound positive effect by functional group of P=O, N-primer and C=O in the structure. Functional group of S (sulphur), fen (phenil), C-aromatic, C1-C2, C3-C4, C5-C6, Ortho Aromatic, P=S, C-Cl and log(Mr) decrease water solubility. The model resulted by regression analysis using stepwise and best subset method is : $\ln(1+watsol) = 11.0 - 0.352 n C (\text{aromatic}) - 5.41 n (P=S) - 1.55 n (C-Cl)$, $F\text{-value} = 32.43$ (sign. 0,000), $R^2 = 0.67$ dan $S=2.643$. The model can be used to predict the organophosphate water solubility.

Keywords : QSPR, ANOVA, regression analysis, organophosphate pesticide

PENDAHULUAN

Secara kimia, bahan pestisida tergolong bahan organik berbahaya dan beracun yang perlu dikendalikan keberadaannya dalam lingkungan. Berbagai golongan senyawa pestisida diidentifikasi berdasar data fisik dan toksikologinya, termasuk di dalamnya senyawa pestisida golongan organofosfat. Informasi keamanan bahan pestisida dinyatakan sebagai informasi toksikologi melalui indikator-indikator toksikologi seperti LD₅₀ dan LC₅₀ dengan pengujian pada tikus atau hewan lain berkaitan dengan indikator pencemaran lingkungan. Selain parameter toksikologi, suatu pestisida memiliki parameter fisika yang spesifik berkaitan dengan data keamanan bahan kimia. Parameter spesifik tersebut berkaitan erat dengan sifat keberadaannya di lingkungan meliputi kelarutan dalam air (*water solubility/watsol*), koefisien distribusi oktanol-air (K_{ow}), tekanan uap, dan stabilitas. Kelarutan dalam air dapat digunakan sebagai salah satu parameter patokan untuk memprediksikan akumulasi bahan pestisida pada perairan. Secara umum, jika kelarutan dalam air cukup besar dan bahan memiliki stabilitas tinggi, tinggalnya dalam lingkungan memerlukan

waktu yang lama dan dapat berakibat pencemaran pestisida (Fatimah, 2005). Selain itu, data kelarutan bahan pestisida berkaitan dengan interaksi senyawa aktif pestisida dengan organisme di lingkungan melalui media transport air. Berdasar hal ini, kelarutan bahan pestisida memberikan kontribusi pada sifat toksikologinya terhadap organisme di lingkungan.

Hingga saat ini, eksplorasi dan usaha mensintesis dan mendesain struktur senyawa aktif sebagai pestisida yang aman sangat berkembang. Metode perancangan obat diperlukan untuk mendesain molekul yang memiliki aktivitas biologik, namun memiliki efek merugikan yang lebih kecil dapat dilakukan dengan menggunakan pendekatan pemodelan aktivitas dan serta pemodelan sifat fisika senyawa. Rancangan yang digunakan umumnya diterapkan sebagai upaya untuk mengurangi faktor coba-coba sekaligus menghemat waktu dan biaya. Rancangan atau pemodelan tersebut lebih populer disebut dengan istilah QSAR atau (*Quantitative Structure and Activity Relationship*) atau HKSA (Hubungan Kuantitatif Struktur dan Aktivitas) dan QSPR (*Quantitative Structure*

and Physical Properties Relationship). Asumsi mendasar yang digunakan dalam analisis ini adalah adanya hubungan kuantitatif antara sifat mikroskopis (struktur molekul) dengan sifat makroskopis suatu molekul (Lee dkk., 1996). Pada analisis ini, deskriptor atau sifat mikroskopis dapat diwakili oleh beberapa parameter. Beberapa metode diusulkan antara lain Metode Hansch, Metode Free-Wilson dan Metode QSAR 3 D atau CoMFA (*Comparative Molecular Field Analysis*). Dari ketiga metode tersebut, deskriptor yang sering digunakan dalam analisis adalah deskriptor fragmental dan deskriptor berdasarkan perhitungan kimia kuantum.

Penentuan model hubungan antara variabel prediktor dan respon merupakan permasalahan utama dalam analisis QSPR. Langkah awal sebelum menentukan modelnya adalah pemilihan prediktor yang akan dianalisis. Selanjutnya adalah menetapkan bentuk hubungan setiap prediktor terhadap respon. Hubungan antara prediktor dan respon dapat berupa hubungan linear maupun non linear. Kebanyakan analisis QSPR menggunakan pendekatan linear. Pemilihan model yang terbaik dapat yang berkaitan dengan memprediksi deskriptor yang berperan dalam analisis regresi. Sementara itu jika ingin mengetahui apakah terdapat pengaruh faktor-faktor terhadap respon dapat digunakan analisis variansi. Gabungan antara analisis variansi dan analisis regresi biasa disebut sebagai analisis kovariansi. Beberapa asumsi mendasari kedua analisis tersebut, yaitu normalitas, multikolinearitas, homoskedastisitas, independensi antar respon. Pengembangan dari analisis model linear tersebut biasa dinamakan *Generalized Linear Model* (Model Linear Tergeneralkan).

Hingga saat ini, dalam bidang kimia dan farmasi, analisis kuantitatif hubungan struktur dan aktivitas atau sifat fisika sering dilakukan dengan menentukan model yang diperoleh melalui metode regresi multi komponen. Sangat jarang dilakukan analisis eksplorasi untuk mengungkap hubungan antara struktur dengan aktivitas. Salah satu metode eksplorasi adalah analisis kluster yang biasanya digunakan untuk mengelompokkan unit/objek penelitian berdasarkan kesamaan sifat. Dalam hal ini analisis kluster digunakan untuk melihat apakah senyawa yang memiliki kemiripan struktur juga memiliki kemiripan aktifitas. Demikian juga sebaliknya senyawa yang

memiliki kemiripan aktifitas juga memiliki kemiripan struktur.

Analisis QSAR dan QSPR

Dalam dunia industri kimia dan farmasi, mengidentifikasi sifat senyawa baru yang diusulkan sebagai bahan aktif atau obat merupakan hal yang penting dilakukan sebelum melakukan proses dalam skala industri. Untuk menghindari aktivitas *trial- and -error*, saat ini dikembangkan pemodelan hubungan kuantitatif antara struktur senyawa dan aktivitasnya atau umum disebut sebagai QSAR dan QSPR (Erickson dkk., 2002). Dengan pemodelan tersebut, sifat senyawa yang akan disintesis dapat diprediksi dan diramalkan aktivitas dan sifat fisiknya berkaitan dengan beberapa hal termasuk aktivitas toksikologinya. Meskipun hingga saat ini telah banyak diusulkan beberapa metode pemodelan dalam QSAR, beberapa telaah kritis terhadap beberapa model yang telah digunakan. Untuk itu, analisis kemometrik yang teliti sangat menunjang pengembangan teknik pemodelan dalam QSAR. Dalam hal ini beberapa tahap utama pemodelan yaitu :

- a. Bagaimana mendesain eksperimen
- b. Bagaimana mengidentifikasi dan mengukur sifat biologik dan sifat fisika suatu senyawa
- c. Bagaimana mengkaitkan satu sifat dengan sifat lainnya yang valid atau tepat sebagai prediktor.

Salah satu yang dapat dipilih berdasarkan informasi tersebut adalah pendekatan Desain molekul secara statistika (*statistical molecular design*) atau yang lebih populer disingkat sebagai SMD.

Ide dasar SMD dalam QSAR adalah mendeskripsikan hubungan antara sifat fisika atau aktivitas senyawa yang tersedia dengan prediktor yang mungkin digunakan. Deskriptor yang digunakan paling tidak terdiri dari 10 hingga 20 prediktor. Prediktor dapat dipilih berdasarkan sistem model atau perhitungan energi orbital molekul berdasarkan kimia kuantum serta perhitungan berdasarkan banyaknya fragment atau atom. Langkah awal yang digunakan dalam pemilihan deskriptor adalah menentukan keikutsertaan deskriptor yang terhitung seperti lipofilisitas, ukuran molekul (*volume bulk*), sifat elektronik, reaktivitas kimia, konformasi aktif, dan sebagainya.

Berdasarkan hal tersebut, adalah sangat menarik untuk menerapkan metode analisis kluster untuk mengidentifikasi faktor-faktor (gugus) pada senyawa pestisida yang mempengaruhi aktifitas. Disamping itu juga dilakukan pemodelan menggunakan analisis QSPR terhadap data kelarutan pestisida. Hasil penelitian akan memberikan usulan model hubungan kuantitatif struktur dengan sifat fisika kelarutan dalam air senyawa pestisida sehingga dapat memberikan masukan pada sintesis dan desain bahan aktif pestisida baru. Variabel yang diambil dalam penelitian adalah pemilihan model hubungan kuantitatif meliputi model linear berdasarkan beberapa deskriptor fragmental yang diusulkan berupa unsur penyusun senyawa dan gugus fungsional dalam senyawa. Dengan penekanan tinjauan pada metode pemodelan, senyawa pestisida yang dipilih adalah senyawa golongan organofosfat. Pemilihan golongan senyawa pestisida didasarkan pada pertimbangan kesamaan struktur dan sifat hidrofobisitas senyawa sehingga diharapkan range data yang diambil dalam model tidak terlalu meruah.

Analisis Kluster

Analisis Cluster : tujuan utamanya adalah mengelompokkan objek berdasarkan kesamaan karakteristik diantara objek-objek tersebut. Cluster yang baik bersifat

- a. mempunyai kesamaan yang tinggi antar anggota dalam satu cluster
- b. mempunyai perbedaan yang tinggi antar cluster

Analisis cluster adalah mengelompokkan individu/responden berdasarkan kesamaan sifat. Langkah-langkah analisis cluster

- a. Mengukur kesamaan objek (similarity). Terdapat tiga dasar/metode yaitu korelasi, jarak (distance jika data dalam metrik) dan asosiasi (jika data non metrik)
- b. Standarisasi ukuran, khususnya jika antar variabel mempunyai selisiah ukuran yang cukup besar
- c. Membuat cluster. Ada dua cara yaitu *Hierarchical method (single linkage, complete linkage, Average linkage, Ward's method, centroid method)* dan *Non Hierarchical method (k-mean kluster)*
- d. Interpretasi cluster.
- e. Melakukan validasi dan *profiling cluster* (penjelasan karakteristik tiap cluster). Profiling dapat dilakukan dengan analisis

crostab dan grafik, bila perlu dilanjutkan analisis diskriminan

Asumsi dalam analisis ini adalah tidak ada multikolinear/korelasi antar objek.

Model Linear

Andaikan bahwa sebuah model linear menjelaskan hubungan antara variabel respon kontinu y dan sekumpulan variabel regresor (prediktor) x_1, \dots, x_k . Dalam percobaan dengan $n > k$ observasi, model dugaan diberikan oleh

$$E(y_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} \quad (i=1, \dots, n) \dots\dots\dots(1)$$

atau model observasi diberikan oleh

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \varepsilon_i \quad (i=1, \dots, n) \dots\dots\dots(2)$$

Sebagaimana dikemukakan didepan, diperlukan asumsi $\varepsilon_i \sim iid N(0, \sigma^2)$. Dengan asumsi ini akan diperoleh persamaan estimator dengan maksimum likelihood maupun OLS (ordinary least squares) (Montgomery dan Pect, 1992). Jika digunakan notasi matrik, X (matrik $n.(k+1)$) maka matrik estimator dengan OLS adalah

$$\hat{\beta} = (X^t X)^{-1} X^t y \dots\dots\dots(3)$$

disini vektor y memuat nilai respon. Matrik varians-covarian adalah

$$Var(\hat{\beta}) = \sigma^2 (X^t X)^{-1} \dots\dots\dots(4)$$

Selanjutnya parameter σ^2 diestimasi dengan

$$s^2 = \frac{\sum_{i=1}^n (y_i - \hat{y})^2}{\{n - (k + 1)\}} \dots\dots\dots(5)$$

Sehingga persamaan (6) menjadi bebas dari parameter dan matrik $(X^t X)^{-1}$ dapat dihitung dengan mudah jika disusun rancangan optimal. dari persamaan (5) dan (6) dapat dilakukan inferensi atas koefisien β_i dengan statistik t ataupun estimasi interval untuk mean respon.

Seleksi Prediktor dan Pemilihan Model

Terdapat beberapa metode untuk menseleksi prediktor yang layak masuk dalam model sehingga diperoleh model terbaik, yaitu: (Draper, 1981)

- 1. All possible regression menggunakan kriteria R^2, S^2 dan Mallows' C_p
- 2. Best subset regression menggunakan kriteria R^2, R^2_{adj} dan C_p
- 3. Backward elimination
- 4. Stepwise regression

5. PRESS (Predicted Residual Sum of Squares)
6. ridge regression
7. Principal component regression
8. Latent root regression

Penelitian ini akan memberikan kontribusi pada pengembangan ilmu pengetahuan dan teknologi dan memberikan suatu model yang dapat digunakan untuk memprediksi sifat toksikologi senyawa pestisida baru yang diusulkan dalam sintesis bahan aktif. Pada cakupan yang lebih luas, penelitian memberikan deskripsi mengenai penggunaan pemodelan molekul yang dapat dikembangkan pada aplikasi-aplikasi lainnya dalam sintesis obat atau bahan aktif.

Masalah yang akan dipecahkan dalam penelitian ini :

1. Apakah deskriptor yang berpengaruh terhadap hubungan kuantitatif antara struktur dan sifat fisika kelarutan dalam air senyawa aktif pestisida menggunakan deskriptor fragmental
2. Bagaimana model yang tepat untuk menyatakan secara kuantitatif hubungan antara struktur kimia dan kelarutan senyawa aktif pestisida dalam air?

METODE

Penelitian dilakukan dengan menggunakan data berasal dari sumber database pestisida PAN pesticide Database, AL Pesticide propertise dan Arsuda pesticide data base serta EXTOXNET. Kelompok senyawa yang diuji adalah senyawa pestisida golongan organofosfat. Pestisida organofosfat terpilih disajikan pada Tabel 1. Langkah-langkah penelitian yang akan dilakukan :

1. Menentukan senyawa pestisida sampel dan menentukan kelompok senyawa
2. Mengeksplorasi sifat toksikologi setiap senyawa
3. Menetapkan deskriptor fragmental yang akan diuji
4. Menetapkan deskriptor yang berpengaruh
5. Memilih model terbaik
6. Menguji metode yang diperoleh.

Metode analisis analisis yang digunakan meliputi analisis kluster, analisis variansi satu arah dan analisis regresi.

Analisis kluster

Dimaksudkan untuk mengelompokkan senyawa-senyawa yang memiliki gugus hampir sama. Sehingga pengelompokan didasarkan atas kemiripan gugus. Dalam setiap kluster diidentifikasi faktor (gugus) yang menyebabkan adanya variasi kelarutan. Disamping itu juga dilihat kemiripan gugus pada senyawa-senyawa dengan kelarutan tinggi dan pada senyawa-senyawa dengan kelarutan rendah. Langkah ini digunakan untuk mengidentifikasi gugus apa yang menyebabkan kelarutan tinggi dan gugus-gugus yang menyebabkan kelarutan rendah.

Analisis regresi

dimaksudkan untuk mencari fungsi yang menghubungkan variabel independen (gugus dan berat molekul) dengan variabel dependen (kelarutan). Digunakan analisis regresi linear untuk masing-masing variabel independen (dengan menganggap variabel independen yang lainya tetap) dan Analisis regresi ganda. Analisis regresi ganda dimaksudkan untuk mencari persamaan regresi terbaik dengan menggunakan metode stepwise dan best subset. Asumsinya adalah galat berdistribusi normal dan variabel independenya terkendali (fix), variabel independenya ditransformasi menjadi $Y = \ln(\text{watsol} + 1)$ dengan watsol adalah nilai kelarutan dalam air . Oleh karena variabel berat molekul mempunyai nilai yang cukup besar, maka ditransformasi menjadi $\log(\text{BM})$.

Model linear untuk masing-masing variabel independen (G)

$$Y_{ij} = \beta_0 + \beta_1 G_j + \varepsilon_{ij} \quad i = 1, 2, \dots, 45 \text{ dan } j = 1, 2, 3, \dots, 36 \quad \dots \dots \dots (6)$$

Model linear untuk regresi ganda

$$Y_{ij} = \beta_0 + \beta_1 G_1 + \beta_2 G_2 + \beta_3 G_3 + \dots + \beta_{36} G_{36} + \varepsilon_{ij} \quad \dots \dots \dots (7)$$

Tabel 1. Data pestisida yang diuji dan kelarutannya dalam air

| No. | NAMA | Kelarutan dalam air (ppm) | No. | NAMA | Kelarutan dalam air (ppm) |
|-----|----------------|---------------------------|-----|------------------------|---------------------------|
| 1 | Cyanofenpos | 0,6 | 27 | Fensulfotion | 700 |
| 2 | Cyanofos | 46 | 28 | Etoprop | 750 |
| 3 | Diazinon | 40 | 29 | Fention | 55 |
| 4 | Diklofention | 0,245 | 30 | Fosfietan | 50000000 |
| 5 | Dimetoat | 25000 | 31 | Forat | 22 |
| 6 | dioksabenzofos | 58 | 32 | metil parathion | 500 |
| 7 | Disulfoton | 25 | 33 | Temefos | 1 |
| 8 | Edifenpos | 5 | 34 | Metamidofos | 1000000 |
| 9 | Iprofenpos | 1000 | 35 | Monokrotofos | 1000000 |
| 10 | Karbofention | 0,34 | 36 | Pirimiphos | 9 |
| 11 | Isoxantion | 1,9 | 37 | pirimiphos etil | 93 |
| 12 | Malation | 145 | 38 | Fosfamidon | 1000000 |
| 13 | Pentoat | 11 | 39 | Fonofos | 16,9 |
| 14 | Phosalon | 10 | 40 | glufosinat-ammn. | 1370000 |
| 15 | Phosmet | 25 | 41 | Glifosat isopropilamin | 900000 |
| 16 | Propafos | 125 | 42 | DMPA | 5 |
| 17 | Protiofos | 1,7 | 43 | Glifosat | 12000 |
| 18 | Metidation | 240 | 44 | demeton-s-metil | 330 |
| 19 | Formotion | 2600 | 45 | Bromofos | 40 |
| 20 | Fenitrotion | 14 | 46 | Etion | 240 |
| 21 | Kloropyfos | 2 | 47 | Manazon | 0 |
| 22 | Coumafos | 1,5 | 48 | Etepon | 0 |
| 23 | Demetol | 3,3 | 49 | Klortion | 0 |
| 24 | Mathidation | 1000 | 50 | Dikapton | 0 |
| 25 | Azinpos | 33 | 51 | Phosphirat | 0 |
| 26 | Diklorofos | 10000 | 52 | Dioksation | 0 |

Analisis Variansi satu arah

dimaksudkan untuk menguji pengaruh banyaknya gugus untuk jenis gugus tertentu terhadap kelarutan. Analisis ini mengasumsikan bahwa faktor lainnya dianggap tidak berpengaruh dan data (kelarutan) berdistribusi normal, sehingga dilakukan ditransformasi untuk data kelarutan sebagaimana dalam analisis regresi. Tidak digunakannya analisis variansi multi arah dan analisis interaksi sebab dari data yang tersedia tidak memungkinkan dilakukannya analisis tersebut.

Model linearnya adalah

$$Y_{ij} = \mu + \tau_j + \epsilon_{ij} \quad i= 1,2,...45$$

$$\text{dan } j= 1,2,3,...,36 \quad \dots\dots\dots(8)$$

Dari deskriptor terpilih, dilakukan uji signifikansi terdapat beberapa gugus yang berpengaruh secara signifikan terhadap

kelarutan. Analisis data digunakan paket program Minitab 11 dan SPSS 11.

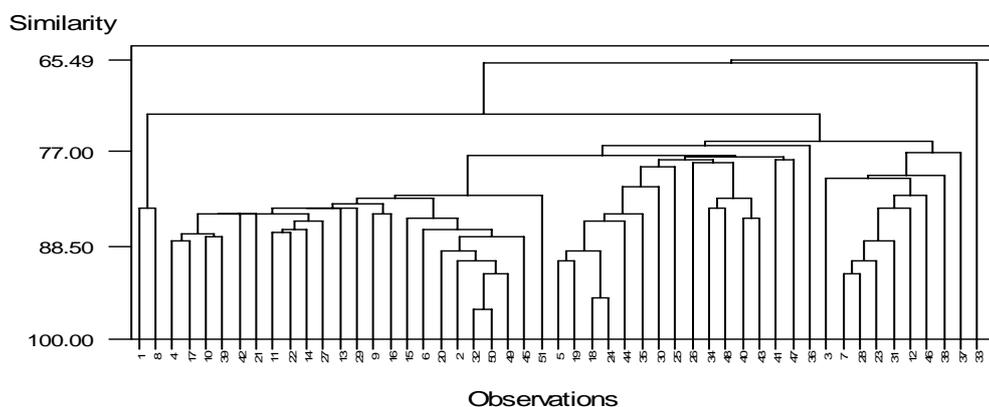
HASIL DAN PEMBAHASAN

Pestisida yang digunakan dalam penelitian ini adalah pestisida golongan organofosfat dengan ciri spesifik strukturnya adalah gugus $\equiv P=O$, $-P-O$, dan gugus $\equiv P=S$. Beberapa senyawa memiliki gugus aromatik dan gugus alifatik dengan panjang rantai bervariasi serta substituen halida pada posisi yang berbeda. Berdasarkan kesamaan struktur yang dimiliki oleh kelompok pestisida organofosfat ditentukan beberapa deskriptor fragmental seperti disajikan pada Tabel 2.

Tujuan dari analisis ini adalah untuk mengetahui apakah terdapat pengaruh deskriptor terhadap kelarutan senyawa (52 jenis senyawa).

Tabel 2. Deskriptor Fragmental yang digunakan

| No. | Deskriptor | Arti deskriptor | No. | Deskriptor | Arti deskriptor |
|-----|-------------------------------------|-------------------------------------------------------------|-----|------------------------|----------------------------------|
| 1 | N C | Banyaknya atom C | 19 | n (-COH=) | Banyaknya gugus -COH= |
| 2 | nH | Banyaknya atom H | 20 | n (CH ₃ O-) | Banyaknya gugus metoksi |
| 3 | nO | Banyaknya atom O | 21 | n (C=O) | Banyaknya gugus karbonil |
| 4 | nCl | Banyaknya atom Cl | 22 | n (\equiv C-Cl) | Banyaknya gugus \equiv C-Cl |
| 5 | nBr | Banyaknya atom Br | 23 | n (orto arom) | Banyaknya gugus pada posisi orto |
| 6 | nP | Banyaknya atom P | 25 | n (para) | Banyaknya gugus pada posisi para |
| 7 | nS | Banyaknya atom S | 26 | n (meta) | Banyaknya gugus pada posisi meta |
| 8 | N N | Banyaknya atom N | 27 | n (P=S) | Banyaknya gugus P=S |
| 9 | n-fen | Banyaknya gugus fenil | 28 | n (-S-) | Banyaknya gugus -S- |
| 10 | N C (alf) | Banyaknya atom C alifatik | 29 | n (-O-R) | Banyaknya gugus -O-R |
| 11 | N C (prim alf) | Banyaknya atom C alifatik primer | 30 | n (N-primer) | Banyaknya N-primer |
| 12 | N C (sek alf) | Banyaknya atom C alifatik sekunder | 31 | n (N sekunder) | Banyaknya N-sekunder |
| 13 | N C (ter alf) | Banyaknya atom C alifatik tersier | 32 | n (N tersier) | Banyaknya N tersier |
| 14 | N C (aromatik) | Banyaknya atom C aromatik | 33 | n (-O-P-) | Banyaknya gugus -O-P- |
| 15 | N (C ₂ =C ₃) | Banyaknya ikatan rangkap dua C ₂ -C ₃ | 34 | n (-NO ₂) | Banyaknya gugus -NO |
| 16 | N (C ₃ =C ₄) | Banyaknya ikatan rangkap dua C ₃ -C ₄ | 35 | Log(BM) | Logaritma Berat molekul senyawa |
| 17 | N (C ₄ =C ₅) | Banyaknya ikatan rangkap dua C ₄ -C ₅ | 36 | n (P=O) | Banyaknya gugus P=O |
| 18 | N (-CH ₂ -O) | Banyaknya gugus -CH ₂ -O | 37 | n (OH) | Banyaknya gugus OH |



Gambar 1. Grafik dendrogram

Analisis kluster

Dari analisis kluster diperoleh grafik dendogram seperti pada Gambar 1. Berdasarkan kluster yang terjadi, dilakukan identifikasi kluster yang memuat senyawa dengan kelarutan tinggi (lebih dari 1 juta) dan kluster dengan kelarutan rendah (kurang dari 5). Hal ini dilakukan untuk mengidentifikasi apakah dalam kluster yang sama/ berdekatan juga mempunyai nilai kelarutan yang sama. Berdasar dendogram di atas, diperoleh tiga kluster untuk kelarutan tinggi (kluster tinggi) yaitu :

Kluster T1 terdiri atas senyawa nomor 30, 35, 44 dan 25

Kluster T2 terdiri atas senyawa nomor 34 dan 38

Kluster T3 terdiri atas senyawa nomor 40 dan 43

Dalam kluster T1 terdapat dua senyawa dengan nilai kelarutan rendah yaitu nomor 44 dan 25. Sementara itu dalam kluster T2, senyawa nomor 38 mempunyai kelarutan relatif rendah. Pada kluster T3, senyawa nomor 43 juga relatif rendah dibanding nomor 40.

Senyawa dengan kelarutan rendah (kluster rendah), dapat dikelompokkan menjadi

Kluster R1 terdiri dari senyawa nomor : 1, 8, 4, 17, 10, 39, 42, 21, 11, 22, 14, 27, 13, 29, 9, 16, 15, 6, 20, 2, 32, 50, 49, 45, 52.

Kluster R2 terdiri dari senyawa nomor : 34, 48, 40, 43.

Kluster R3 terdiri dari senyawa nomor : 7, 28, 23, 31, 12, 46

Kluster R4 terdiri dari senyawa nomor : 33 dan 53

Pada kluster R1 terdapat 3 senyawa yang mempunyai kelarutan relatif tinggi dibanding anggota senyawa yang lain yaitu senyawa nomor 27, 9 dan 32. Pada kluster R3 terdapat dua senyawa dengan kelarutan relatif tinggi yaitu nomor 28 dan 12.

Dengan adanya nilai kelarutan yang sangat berbeda untuk senyawa yang terletak pada kluster yang sama (baik kluster rendah maupun tinggi) maka dengan mencermati gugus-gugus yang terdapat pada senyawa tersebut diduga bahwa Gugus P=O memberikan efek positif (memperbesar) kelarutan

Gugus-gugus C-Cl, Ortho-aromatis, P=S, C3=C4, C5=C6, C-aromatik, Cl dan fen memberikan efek negatif (memperkecil) kelarutan.

Untuk membuktikan dugaan tersebut perlu dilihat perubahan banyaknya gugus untuk

masing-masing jenis gugus tersebut terhadap kelarutan. Data disajikan pada Tabel 3.

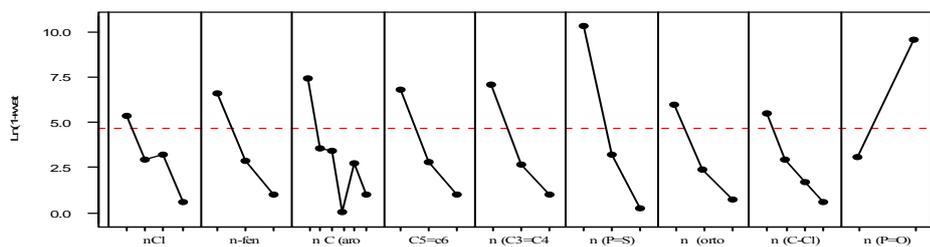
Dari Tabel 3 dan Gambar 2 terlihat bahwa gugus P=O memberikan pengaruh positif terhadap kelarutan pada sebagian besar senyawa. Kontribusi positif dimungkinkan oleh karena interaksi ionisasi senyawa oleh air terjadi pada gugus fosfat senyawa. Muatan formal yang dihasilkan oleh gugus P=O tersebut pada suatu senyawa meningkatkan kepolaran berdasar elektron π yang dimilikinya dan berakibat meningkatkan daya terlarut dalam berdasar prinsip *like-dissolve like*. Sebagian kecil senyawa yang memuat gugus P=O diidentifikasi memiliki kelarutan kecil, dimungkinkan oleh adanya gugus lain yang memberikan respon negatif pada senyawa itu. Sementara itu gugus P=S, Ortho, C1=C2, C3=C4, C5=C6, Cl dan gugus Fen (fenil) memberikan pengaruh negatif terhadap kelarutan. Keberadaan gugus-gugus fungsional tersebut terhadap kelarutan dipengaruhi oleh kestabilan struktur gugus fungsional. Adanya ikatan rangkap dengan posisi C1=C2 disertai dengan adanya ikatan rangkap C3=C4 dan C5=C6 menunjukkan adanya ikatan konjugasi pada struktur senyawa. Ikatan konjugasi terbentuk memiliki kecenderungan untuk mempertahankan kestabilan rantai hidrokarbon yang menurunkan interaksi ionisasi senyawa dalam air (oleh gugus pospat) oleh distribusi elektron yang merata di dalam molekul. $\log(BM)$ memberikan efek negatif (memperkecil) kelarutan memberikan makna bahwa semakin besar berat molekul ukuran senyawa semakin besar dan sulit larut di dalam air. Selain itu, berdasar struktur senyawa-senyawa yang digunakan dalam pemodelan ini, berat molekul terkait dengan jumlah atom (dominan atom C) yang berefek non polar sehingga menurunkan daya solvasi senyawa.

Pada beberapa senyawa yang memuat gugus positif dan gugus negatif secara bersamaan, sulit digeneralisir gugus mana yang lebih dominan. Hal ini menyangkut banyaknya gugus dan pengaruh gugus lainnya. Disamping itu sangat memungkinkan adanya pengaruh faktor lain. Pada senyawa-senyawa yang tidak memuat gugus positif dan gugus negatif, nilai kelarutannya masih bervariasi walaupun variasinya kecil. Hal ini berarti ada faktor lain yang mempengaruhi kelarutan, kemungkinan besar adalah ukuran molekul. Hasil uji anava satu arah dan analisis regresi sederhana untuk masing-masing gugus disajikan pada Tabel 4.

Tabel 3 Data Identifikasi Pengaruh Gugus Fungsional terhadap Kelarutan

| Gugus | Banyak Gugus | Banyak Senyawa | Rata-rata Kelarutan | Nilai min s/d Nilai maks |
|------------|--------------|----------------|---------------------|--------------------------|
| P=O | 0 | 39 | 802 | 0 s/d 25000 |
| | 1 | 13 | 4253375 | 0 s/d 50000000 |
| P=S | 0 | 12 | 4607823 | 0 s/d 50000000 |
| | 1 | 37 | 845 | 0 s/d 25000 |
| | 2 | 3 | 1 | 1 |
| Fen | 0 | 27 | 2048981 | 0 s/d 50000000 |
| | 1 | 22 | 120,2 | 0 s/d 1000 |
| | 2 | 3 | 2,2 | 0,6 s/d 5 |
| Ortho | 0 | 34 | 1627166 | 0 s/d 50000000 |
| | 1 | 17 | 89,6 | 0 s/d 700 |
| | 2 | 1 | 1 | 1 |
| C5=C6 | 0 | 25 | 2212900 | 0 s/d 50000000 |
| | 1 | 24 | 110,3 | 0 s/d 1000 |
| | 2 | 3 | 2,20 | 0,6 s/d 5 |
| C3=C4 | 0 | 25 | 2212900 | 0 s/d 50000000 |
| | 1 | 24 | 110,3 | 0 s/d 1000 |
| | 2 | 3 | 2,20 | 0,6 s/d 5 |
| C1=C2 | 0 | 25 | 2212900 | 0 s/d 50000000 |
| | 1 | 24 | 110,3 | 0 s/d 1000 |
| | 2 | 3 | 2,20 | 0,6 s/d 5 |
| C-Aromatik | 0 | 23 | 2405322 | 0 s/d 50000000 |
| | 1 | 1 | 33 | 33 |
| | 3 | 2 | 51 | 9 s/d 93 |
| | 5 | 1 | 0 | 0 |
| | 6 | 22 | 118,8 | 0 s/d 1000 |
| | 12 | 3 | 2,20 | 0,6 s/d 5 |
| Cl | 0 | 39 | 1392695 | 0 s/d 50000000 |
| | 1 | 6 | 166669 | 0 s/d 1000000 |
| | 2 | 5 | 2009 | 0 s/d 10000 |
| | 3 | 2 | 1 | 0 s/d 2 |

Main Effects Plot - Means for Ln(1+wat



Gambar 2. Plot rata-rata log kelarutan untuk masing –masing gugus.

Tabel.4 Nilai signifikansi hasil uji Anava dan Regresi terhadap variabel

| Gugus | ANOVA | Regresi | Pengaruh |
|----------------|-------|---------|----------|
| n C | 0,296 | 0,000 | - |
| nS | 0,003 | 0,022 | - |
| n-fen | 0,003 | 0,001 | - |
| N C (aromatik) | 0,003 | 0,000 | - |
| n (C1=c2) | 0,000 | 0,000 | - |
| n (C3=C4) | 0,000 | 0,000 | - |
| C5=c6 | 0,000 | 0,000 | - |
| n (-CH2-O) | 0,075 | 0,004 | - |
| n (C=O) | 0,005 | 0,017 | + |
| n (C-Cl) | 0,123 | 0,017 | - |
| n (orto arom) | 0,013 | 0,003 | - |
| n (P=S) | 0,000 | 0,000 | - |
| n (N-primer) | 0,010 | 0,064 | + |
| n (P=O) | 0,000 | 0,000 | + |
| log BM | | 0,000 | - |

Dengan ANAVA satu arah disimpulkan bahwa terdapat 11 gugus yang pengaruhnya signifikan (dengan $\alpha=0,05$). Faktor log BM tidak dianalisis dengan ANAVA karena bersifat kontinu. Sementara itu dengan analisis regresi diperoleh 14 gugus termasuk log BM yang pengaruhnya signifikan (dengan $\alpha=0,05$). Pada analisis regresi, pengamatan yang masuk dalam outlier tidak digunakan untuk mengestimasi koefisien regresi. Sedangkan pada anava, semua data digunakan dalam perhitungan. Hal ini mengakibatkan banyaknya gugus yang signifikan pada analisis regresi lebih banyak. Sehingga dapat disimpulkan bahwa gugus yang berpengaruh terhadap kelarutan senyawa adalah gugus S, fen, C-aromatik, C1-C2, C3-C4, C5-C6, C=O, Ortho Aromatik, P=S, N-primer, P=O dan log(BM).

Kelemahan kedua analisis ini adalah menganggap gugus/faktor lain tetap. Dari data yang ada asumsi ini jelas tidak dipenuhi. Mengingat keterbatasan data, maka pemenuhan asumsi tersebut tidak dapat dilakukan. Demikian juga analisis interaksi antar gugus tidak dapat dilakukan, sebab diperlukan tambahan data untuk senyawa-senyawa yang lain.

Analisis Regresi Ganda

Pemilihan regresi terbaik digunakan metode stepwise dan metode best subset. Dari dua metode tersebut diperoleh persamaan :

$$\ln(1+watsol) = 11.0 - 0.352 n C (aromatik) - 5.41 n (P=S) - 1.55 n (C-Cl)$$

$$\text{Nilai T : } (14,78) \quad (-3,13) \\ (-7,16) \quad (-3,21)$$

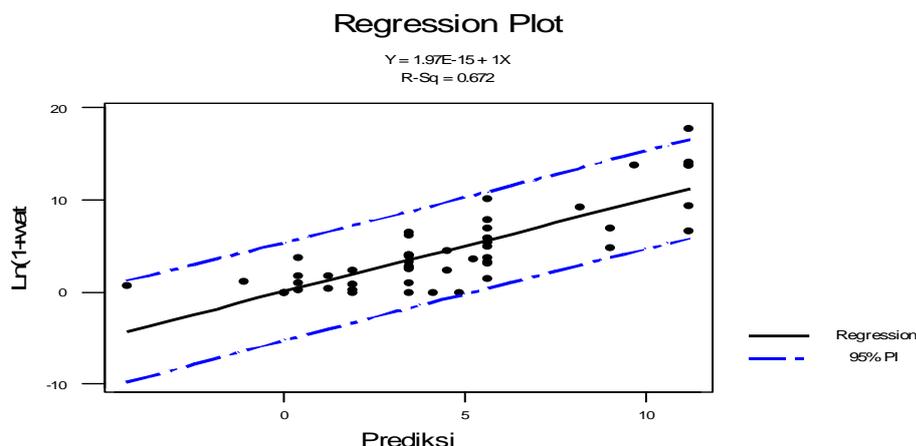
$$\text{Nilai sign. : } (0,000) \quad (0,003) \\ (0,000) \quad (0,002)$$

$$F = 32,43 \text{ (sign. } 0,000) \text{ , } R^2 = 0,67 \text{ dan } S=2,643$$

Watsol = kelarutan senyawa dalam air

Dalam persamaan ini ternyata tidak semua gugus yang secara parsial berpengaruh tetapi tidak masuk dalam persamaan regresi. Hal ini sangat mungkin terjadi sebab persamaan regresi disusun untuk keperluan prediksi, yang menjadi pertimbangan dalam pemilihan persamaan diantara semua persamaan yang ada (model linear) adalah sebuah persamaan yang mempunyai galat terkecil. Jadi pertimbangan bukan semata-mata pada adanya hubungan variabel independen (gugus) dengan variabel dependen (kelarutan). Oleh karena itu untuk mengetahui ada-tidaknya hubungan antara variabel independen dengan variabel independen sebaiknya digunakan Analisis variansi atau bisa juga digunakan analisis korelasi.

Ketepatan prediksi dapat diketahui dari nilai koefisien determinasi (R^2), dalam kasus ini diperoleh sebesar 0,67. Kesesuaian antara nilai prediksi dan nilai observasi dapat juga dinyatakan dalam plot grafik yang disajikan pada Gambar 3.



Gambar 3. Plot nilai prediksi dan nilai observasi dengan regresi linear sederhana

Berdasar plot yang diperoleh, nampak bahwa terdapat beberapa amatan berada diluar daerah interval konfidensinya. Titik titik tersebut biasa disebut titik outlier. Dari analisis output (lampiran 2) diperoleh persamaan regresi $y = 0.000 + 1.00 x$ ($t=10, 12$) nilai intersep mendekati nol dan nilai slope signifikan, sehingga dapat disimpulkan bahwa persamaan regresi yang diperoleh dari metode stepwise dan metode best subset diatas masih layak digunakan untuk keperluan prediksi.

KESIMPULAN

Dengan menggunakan Analisis Kluster, Analisis variansi satu arah dan analisis regresi dapat disimpulkan bahwa gugus yang mempengaruhi kelarutan senyawa adalah
a. gugus P=O, N-primer dan C=O memberikan efek positif (memperbesar) terhadap kelarutan
b. gugus S, fen, C-aromatik, C1-C2, C3-C4, C5-C6, Ortho Aromatik, P=S, gugus C-Cl dan log(BM) memberikan efek negatif (memperkecil) kelarutan.

Model yang diperoleh dari analisis regresi dengan menggunakan metode stepwise dan metode best subset adalah
 $\text{Ln}(1+\text{watsol}) = 11.0 - 0.352 n \text{ C (aromatik)} - 5.41 n \text{ (P=S)} - 1.55 n \text{ (C-Cl)}$
dengan nilai $F = 32,43$ (sign. 0,000), $R^2 = 0,67$ dan $S=2,643$

DAFTAR PUSTAKA

Anonim, EXTOWNET, <http://ptcl.chem.ox.co.uk/EXTOWNET>

- Draper, N.R., Smith, H., 1966. *Applied Regression Analysis*, 2nd Edition, John Wiley and Sons, New York.
- McCullagh, P., Nelder, J., 1983. *Generalized Linear Models*, Chapman and Hall, London.
- Fatimah, 2005. *Panduan Pengetahuan Bahan Kimia*, Program Studi D3 Kimia Analis UII.
- Erickson, L., Anderson, P.M., Johansson, E., Lundstedt, T., 2002. Statistical Molecular Design-Acore Concept in Multivariate QSAR and Combinatorial Technologies, Part I-General Principles and Application to Lead Optimization, *Homepage of Chemometrics*, November 2002.
- Erickson, L., Anderson, P.M., Johansson, E., Lundstedt, T., 2002. Statistical Molecular Design-Acore Concept in Multivariate QSAR and Combinatorial Technologies, Part II-QSAR Application, *Homepage of Chemometrics*, Desember 2002.
- Erickson, L., Anderson, P.M., Johansson, E., Lundstedt, T., 2003. Statistical Molecular Design-Acore Concept in Multivariate QSAR and Combinatorial Technologies, Part III-QSAR directed Virtual learning, *Homepage of Chemometrics*, February 2003.
- Free, S.M., Wilson, J.W., 1964. A Mathematical Contribution to Structure -Activity Studies, *J.Med.Chem.*, **7**, 395-399.

- Kubinyi, H., 1993. *QSAR; Hansch Analysis and Related Approaches*, VCH Verlagsgesellschaft, Weinheim.
- Lee, K.W., Kwon, S.Y., Hwang, S., Lee, J.U., Kim, H., 1996. Quantitative Structure-Activity Relationship (QSAR) Study on C-7 Substituted Quinolone, *Bull. Korean Chem. Soc.*, **17**, 147-152.
- Sudarmanto, B.S., 2002. *Analisis Hubungan Kuantitatif Struktur-Aktivitas Seri Senyawa Analog Kurkumin Sebagai Antioksidan Menggunakan Deskriptor Berdasarkan Perhitungan Kimia Kuantum*, Tesis, Universitas Gadjah Mada, Jogjakarta.
- Vighi, M dan Funari, E, 1995. *Pesticide Risk in Ground Water*, Lewis Publisher, New York.
- Vogue, P.A., Kerle, A., Jenkin, J., 2002. OSU Extension Pesticide Properties Database, *OSU bulletin*.

**Lampiran 1.
Stepwise Regression**

F-to-Enter: 4.00 F-to-Remove: 4.00
Response is Ln(1+wat) on 37 predictors, with N = 52

| Step | 1 | 2 | 3 |
|-----------|-------|--------|--------|
| Constant | 9.680 | 10.494 | 10.978 |
| n (P=S) | -6.03 | -6.11 | -5.41 |
| T-Value | -6.81 | -7.80 | -7.16 |
| n (C-Cl) | | -1.95 | -1.55 |
| T-Value | | -3.86 | -3.21 |
| n C (aro) | | | -0.35 |
| T-Value | | | -3.13 |
| S | 3.24 | 2.87 | 2.64 |
| R-Sq | 48.14 | 60.24 | 66.96 |

Regression Analysis

The regression equation is
 $\text{Ln}(1+\text{watsol}) = 11.0 - 0.352 \text{ n C (aromatik)} - 5.41 \text{ n (P=S)} - 1.55 \text{ n (C-Cl)}$

| Predictor | Coef | StDev | T | P | VIF |
|-----------|---------|--------|-------|-------|-----|
| Constant | 10.9782 | 0.7426 | 14.78 | 0.000 | |
| n C (aro | -0.3516 | 0.1125 | -3.13 | 0.003 | 1.2 |
| n (P=S) | -5.4071 | 0.7551 | -7.16 | 0.000 | 1.1 |
| n (C-Cl) | -1.5494 | 0.4825 | -3.21 | 0.002 | 1.1 |

S = 2.643 R-Sq = 67.0% R-Sq(adj) = 64.9%

Analysis of Variance

| Source | DF | SS | MS | F | P |
|------------|----|---------|--------|-------|-------|
| Regression | 3 | 679.72 | 226.57 | 32.43 | 0.000 |
| Error | 48 | 335.32 | 6.99 | | |
| Total | 51 | 1015.05 | | | |

| Source | DF | Seq SS |
|----------|----|--------|
| n C (aro | 1 | 278.25 |
| n (P=S) | 1 | 329.45 |
| n (C-Cl) | 1 | 72.03 |

Unusual Observations

| Obs | n C (aro | Ln(1+wat | Fit | StDev Fit | Residual | St Resid |
|-----|----------|----------|--------|-----------|----------|----------|
| 21 | 6.0 | 1.099 | -1.187 | 1.283 | 2.285 | 0.99 X |
| 30 | 0.0 | 17.728 | 10.978 | 0.743 | 6.749 | 2.66R |
| 33 | 12.0 | 0.693 | -4.056 | 1.197 | 4.749 | 2.02R |
| 51 | 5.0 | 0.000 | 4.572 | 1.390 | -4.572 | -2.03RX |

R denotes an observation with a large standardized residual
 X denotes an observation whose X value gives it large influence.
 Durbin-Watson statistic = 1.86

Lampiran 2. (Analisis regresi antara nilai prediksi dan observasi)**Regression**

The regression equation is

$$y = 0.000 + 1.00 x$$

| Predictor | Coef | StDev | T | P |
|-----------|---------|---------|-------|-------|
| Constant | 0.0000 | 0.5861 | 0.00 | 1.000 |
| x | 1.00000 | 0.09886 | 10.12 | 0.000 |

S = 2.581 R-Sq = 67.2% R-Sq(adj) = 66.5%

Analysis of Variance

| Source | DF | SS | MS | F | P |
|------------|----|---------|--------|--------|-------|
| Regression | 1 | 681.85 | 681.85 | 102.32 | 0.000 |
| Error | 50 | 333.19 | 6.66 | | |
| Total | 51 | 1015.05 | | | |