

**KEY LINK BETWEEN REGRESSION
COEFFICIENTS AND STRUCTURAL
CHARACTERIZATION OF
HYDROCARBONS IN GASOLINE (MOTOR
SPIRIT) FOR CONTINUOUS BLENDING
PROCESS APPLICATIONS**

Santanu Talukdar

**Manager-Analytical and Systems Engineering
Field Engineering Services Division
Yokogawa India Ltd.
Tel(O): +918051586000
E-Mail: Santanu.Talukdar@in.yokogawa.com**

Key words: FT-NIR, PCA, PCR, PIONA, Regression Coefficients, Cross Validation, Prediction Accuracy, Recipe, Continuous Blending, Batch Blending

LIST OF ABBREVIATIONS:

1. FCCG: The gasoline cut from fluidized catalytic cracker, known as fluidized cat cracker gasoline.
2. CCG: The gasoline cut from catalytic cracker, known as cat cracker gasoline.
3. Raffinate: The bottom portion from splitting unit of Aromatic Recovery Unit (ARU).
4. Reformate: Light reformate cut, low in aromatics, from Catalytic Reforming Unit (CRU).
5. Naptha: Naptha cut, light aromatics known as light aromatic naptha (LAN).
6. RON: Physical Property data, known as Research Octane Number.
7. MON: Physical Property data, known as Motor Octane Number.
8. RVP: Physical Property data, known as Reid Vapor Pressure.
9. REC70/REC100: Physical Property data, known as Recovery @ 70degC, Recovery @ 100degC.
10. Benzene: Concentration of Benzene.
11. PIONA: Hydrocarbon structures lumped in terms of cut points. Paraffin (P), Isoparaffin (I), Olefin (O), Napthene (N), Aromatics (A).
12. FT-NIR: Fourier Transform Near Infrared Analysis.
13. PCA/PCs: Principal Component Analysis/ Principal Components.
14. PCR: Principal Component Regression.
15. RMSEP: Root Mean Square Error of Prediction.
16. RMSE: Root Mean Square Error.
17. Recipe: Blend Recipe.
18. BPC: Blend Property Control.
19. BRC: Blend Ratio Control.
20. P+I: Paraffin + Isoparaffin structures.
21. QC Lab: Quality Control Laboratory inside the refinery complex where primary ASTM tests and QA/QC criteria in terms of repeatability/reproducibility is established.
22. ASTM: American Society of Testing and Materials.

INTRODUCTION:

In this application, a Fourier Transform Near InfraRed Analyzer (FT-NIR) is used to predict properties in Gasoline (Motor Spirit or MS). The predicted properties are given as feedback to Advanced Process Control Software as **Blend Property Control (BPC)**. The FT-NIR, together with BPC and **BRC (Blend Ratio Control)**, optimizes blending in the gasoline pool so that target values in the final properties are achieved with minimum reblend and product giveaways.

Analysis is made on the basis of key structural characteristics in PIONA classification and its impact on property values.

In this study the key link between regression coefficients and structural characteristics of hydrocarbons is highlighted. The regression coefficients play a major role in prediction accuracy of the regression model developed. The step changes of the coefficients in terms of process recipe changes and process operating variable changes are highlighted.

The impact of continuous blending is to keep such coefficient changes to a minimum in order to ensure a highly accurate inferential property prediction from NIR. Such coefficient changes are caused by scheduling from bringing changes in components in the various blend recipes. BPC decides the feasibilities of the various blend recipes in meeting the desired fuel property specifications. BPC also optimizes the blend recipes with the optimization function being either on maximizing profit or minimizing product inventory. BPC also decides the optimum ratios of the various components in the blend recipes. BRC sets the ratio control after getting the information from BPC. The control is done on individual control streams through the flow controller per setpoints determined by BPC.

This control logic is primarily on the continuous blending process rather than on the batch blending process.

Next, the impact on prediction accuracy by FTNIR by the influence of such regression coefficients is highlighted.

The present report excludes process unit studies and its hydrocarbon structural characterization in terms of spectral absorptions, as it is beyond present scope and content.

FT-NIR and Model Development:

Thirty-seven process samples are taken for this particular application. Each sample is tagged with sample name and tested in QC Lab for physical properties, namely, RON, MON, RVP, REC at 70degC, REC at 100degC, and Benzene. Each sample is injected three times to get better repeatability for the spectral database. Hence, the total sample size is 118. However, because there are some missing data in certain properties, the effective sample size therefore varies from 105 to 111.

The process sample is selected after a detailed investigation on actual process runs. Accordingly, the blend recipes selected to prepare the realistic final blends are:

1. FCCG + CCG (recipe 1)
2. FCCG + CCG + raffinate (ARU) (recipe 2)
3. FCCG + CCG + naptha (light aromatic naptha) (recipe 3)
4. FCCG + CCG + reformat (light reformat, low aromatic cut, CRU) (recipe 4)

The final model is prepared from a spectrum database selected from the samples in the above blend recipes. In the final model, therefore, the necessary samples are from the above four recipes. In the blend, the rundown streams are FCCG + CCG and the control components are reformat or raffinate or naptha. This blend should meet the desired final fuel property specifications. The control point in terms of giveaway is determined by the average of RON and MON which is known as AKI. Once AKI giveaway is controlled, other physical properties are also determined so that the final product meets its quality specifications. The primary components of the blends as FCCG or CCG are directly determined from the process. Depending upon the operations, scheduling has the flexibility to divert either reformat (light aromatic cut) from the CRU, raffinate from ARU, or the light aromatic naptha (LAN), as per the defined recipes from above.

This blend is not allowed if either of the primary components FCCG or CCG is not available for the recipe. This blend also does not allow other typical components, e.g., heavy aromatic naptha (HAN), unstabilized naptha or high aromatic cut of reformat from CRU.

A Data Matrix is developed with samples in rows taken from the above blend recipes, samples being of low benzene grade and spectral wave numbers in columns. The cells are the various NIR absorbances. The wavenumber selected ranges from 9194 to 5955 cm⁻¹ to take care of total PIONA absorbances within this range. Wavenumbers lower than 5955 and higher than 9194 are rejected after studying the spectral characteristics because of noise. The wavenumbers are known as X-variables in the Data Matrix.

The total Data Matrix being multivariate in nature, the raw data is decomposed through Principal Component Analysis (PCA). Linear Regression is performed on this multivariable data matrix by Principal Component Regression (PCR) to find the best estimate. The responses (physical properties) are known as Y-variables in the Data Matrix. In our applications the Y-variables are RON, MON, RVP, REC70, REC100, and Benzene.

After the data decomposition into PCs (Principal Components), the samples are represented into score plots in any PC1, PC2 co-ordinates. Similarly the X-variables are represented as Loading Plot in any PC1, PC2 co-ordinates. The original Data Matrix in terms of samples, X-variables and Y-variables is now transformed as Data Matrix = (Score) * (Loading) + Error. Finally, the response variable Y is a linear combination of the X-variables. The coefficients are the regression coefficients.

The X-variables are all linearly independent. The main process variables (PIONA) are selected in terms of wavelength absorbances. Accordingly, the main PCs are selected for the model in such a way that the Root Mean Square Error of Prediction (RMSEP) is a minimum. All the PCs are orthogonal to each other.

Regression Coefficients Analysis for Hydrocarbon Structural Characterization:

The regression coefficients for each response variable Y are analyzed in detail. These regression coefficients are generated on the basis of the developed model described above. A summary is tabulated below in Table 1 as indices. The indices are scaled and marked as high (+) or high (-) depending upon their values. The regression coefficients are plotted in Figures 1 to 5 in Appendix 1.

Table 1. Indices

Property	Aromatics	Paraffin+ Isoparaffin	Aromatics+ Olefins	Napthenes
RON	high +	high +/high -	high +	
MON	high -	high +/high -	high -	
REC at100degC		high -	high -	
RVP		high +/high -		
Benzene	high +		high +	

The influences of the above regression coefficients on hydrocarbon structures can be summarized as following:

1. The structural composition of RON -- as given in Figure 1, Appendix 1 – (a) implies that the model has high positive regression coefficients in the aromatics and olefins absorbance range as well as high positive and negative regression coefficients in the P+I absorbance ranges; and (b) implies that for samples having strong NIR absorbances in aromatics, olefins and P+I ranges will have a positive correlation with RON because of the high influence of the regression coefficients for such absorbances in cases where such samples also have high positive correlation with the X-variables in the PCs. Such high absorbances will depend on the type of primary crude, its processing at the CDU, and the operational variables at the cracking units. For example, paraffin-based crude types will be responsible for low RON because of their high negative regression coefficients and strong NIR absorbances in that region. However RON can be increased by the high positive coefficients with strong paraffin-based absorbance by manipulating the operational and the catalytic conditions of the crackers.

2. Changes in the crude type, or in the operating mode of the cracker, or in the product feed will have a likely effect on RON primarily due to the changes in the absorbance coefficients by NIR. Also note the differences in regression coefficients between RON and MON. This implies that RON will always be higher than MON for similar absorbances. For REC at 100 °C, aromatics + olefins have strong negative regression coefficients. More aromatics and olefins absorbance in the sample will lower the recovery, and vice versa for less. RVP is characterized by strong P+I regression coefficients. More P+I absorbances, RVP will be high, and vice versa for less. Benzene is strongly characterized by high positive regression coefficients in the aromatics and olefins absorbance ranges.
3. The blend recipe directly impacts property values. For example, in a blend recipe, increasing component ratios of either FCCG or CCG will increase RON due to increase of aromatics + olefins absorbance coefficients. But high aromatics have a negative correlation and hence will decrease REC100. RVP will be influenced by lighter components, as P+I. More P+I components will increase RVP. Lighter components, such as naphtha, reformat and raffinate, will typically decrease RON and increase RVP. MON will have a lower value than RON because of its structural characteristics. Refer to the Figures in Appendix 1.
4. Blended samples having similar hydrocarbons structural (PIONA) absorbances by NIR will give a high inferential prediction output with good accuracy. Large variances in samples, as from changes in crude type or varying operating modes of the reactors, will change the NIR absorbance coefficients significantly. This will increase the prediction inaccuracy of the model. Therefore, such extreme samples must be incorporated during the model development stage and a new set of regression coefficients generated. This is equivalent to selecting samples such that the variance is maximized.
5. An unknown spectrum type for the new sample, such as by adding new component to the blend recipe or totally new blend recipe, will make the model unsteady. Process outliers are a measure of how close the spectrums of the sample match, and this in turn means how close the NIR absorbance coefficients match. In the case of a mismatch, an alarm triggers.
6. Process operation variables have a direct impact on hydrocarbon structure. Typical are catalyst characteristics (yield and properties), coking, reactor temperature, heat exchanger plugging and fouling, feed contaminants, and others. Operation variables such as process leaks will also trigger outliers, because of the large variance in the sample. The robust model needs to incorporate all such borderline extreme samples so that the above process samples are represented in the model.

Changes in regression coefficients due to changes in blend recipes (variances in samples) in different models in similar hydrocarbon structural absorbances (PIONA) – Study for RON

Please refer to Appendix 2, Figures 1 to Figure 5.

1. Figure 1 for RON is taken as the base case for the model of FCCG + CCG only. As discussed above, RON is influenced by strong positive regression coefficients in aromatics and olefins absorbance, and both positive and negative regression coefficients in P+I absorbance.
2. In Figure 2, the model of FCCG + CCG + reformat is studied. The changes in the regression coefficients in similar absorbance ranges are noticeable. This lowers RON from the base case in Figure 1. This is because of high negative regression coefficients in P+I and aromatics + olefins absorbance. The allowance of model 1 to predict conditions of model 2 is erroneous because of the new sample reformat in the recipe. This means that the recipe condition is changed for prediction. In a similar way, the allowance of model 2 to predict conditions of model 1 is erroneous because of the changed recipe, i.e., a recipe without reformat.
3. In Figure 3, naptha is added to base case 1. The new model is prepared with FCCG + CCG + naptha as components. The model depresses RON because of high negative regression coefficients in P+I absorbances and changes in regression coefficients in aromatics+olefins absorbance. The allowance of model 3 to predict conditions of model 2 or model 1 is erroneous.
4. In Figure 4, raffinate is added to base case 1. The new model is prepared with FCCG + CCG + raffinate. The model depresses RON, as predicted, because of a high negative regression coefficients in P+I absorbance. Since this component raffinate is from the ARU (Aromatic Recovery Unit), it has positive regression coefficients of aromatic and olefin absorbance. The allowance of model 4 to predict conditions of model 3 or model 2 or model 1 is erroneous.
5. In Figure 5, a new model is prepared with the above four recipes as the calibration samples set. No major spectrum shift is noticed. The final blend sample will belong to a similar calibration sample population in case the sample is similar to any of the defined blend recipes. The new model 5 can then predict the conditions of models 1 to 4.
6. If there are major spectrum shifts by merging all the samples in recipes, then a single model would not be sufficient. Multiple models are required. This is best tackled by batch blending.

Cross Validation of the Model

The developed model is validated by employing a Cross Validation Technique, a powerful and effective statistical tool. A total sample size of 108 samples is segregated into 6 segments, each segment containing 18 samples.

1. The model is calibrated by keeping the first segment out of the domain.
2. The model is validated by taking the first segment as an unknown sample and predicted.

3. The steps are repeated until all segments are tested. This process is known as internal validation.

After doing Cross Validation Analysis on the model, the model is checked and external validation performed with external samples. In case the external samples are similar to the training samples, then the uncertainty in prediction is within “Prediction $\pm 2*RMSEP$ ”.

Measurement Errors

The measurement errors could be of the following four types:

1. Primary Measurement Error
2. Sampling Error
3. NIR Absorption Error
4. Wavenumber Noise

Typically considering 1 and 3 as reproducible errors, the other gross errors produced in measurement are due to types 2 and 4.

In this applications, we find the sampling error to be very high in RVP. This is because of vapor pressure loss during the sampling time, the measurement time, and finally the NIR spectrum absorption time. Because of such a high error, the slope of the RVP Regression is poor.

The Calibration Slope can be adjusted by biasing the slope from the NIR Analyzer after downloading the model to the analyzer.

Future Model Improvements

The Model should be upgraded regularly in the future to make it more robust, more accurate and to take into consideration all process information and process operation variables. For example, with 37 sample sizes, only 4 samples belong to the category of recipe 1. This total size is too narrow to accommodate all process variances.

Conclusions

1. How close the NIR absorbance coefficients of a new sample match the calibration sample set determines the accuracy of the model.
2. NIR absorbance coefficients will match if the sample components in the blend recipe are in close agreement.
3. Scheduling generates multiple blend recipes from a set of linear programs.
4. A single model is prepared incorporating the various blend recipes by merging the samples in various blend recipes.
5. BPC decides upon the feasibility of each of the blend recipes in order to meet property specifications and optimization functions.
6. The optimization function is on maximizing profit or minimizing product inventory.

7. BPC also decides on the optimum ratios of the various components in any blend recipe.
8. BRC controls the streams to achieve the set points as determined by BPC.
9. Typically per recipe, 30 to 50 samples are required to achieve statistically significant correlation coefficients. Later on, the sample size can be increased to incorporate more variances in sample, and thus make the model more robust.
10. The prediction accuracy is tested.
11. Sample variances increase because of process and operation variables.
12. In Batch Blending, unlike Continuous Blending as discussed above, multiple models are prepared, each model for each batch. Depending upon the batch switch, there is a necessity of a model switch.

APPENDIX 1- FIGURES

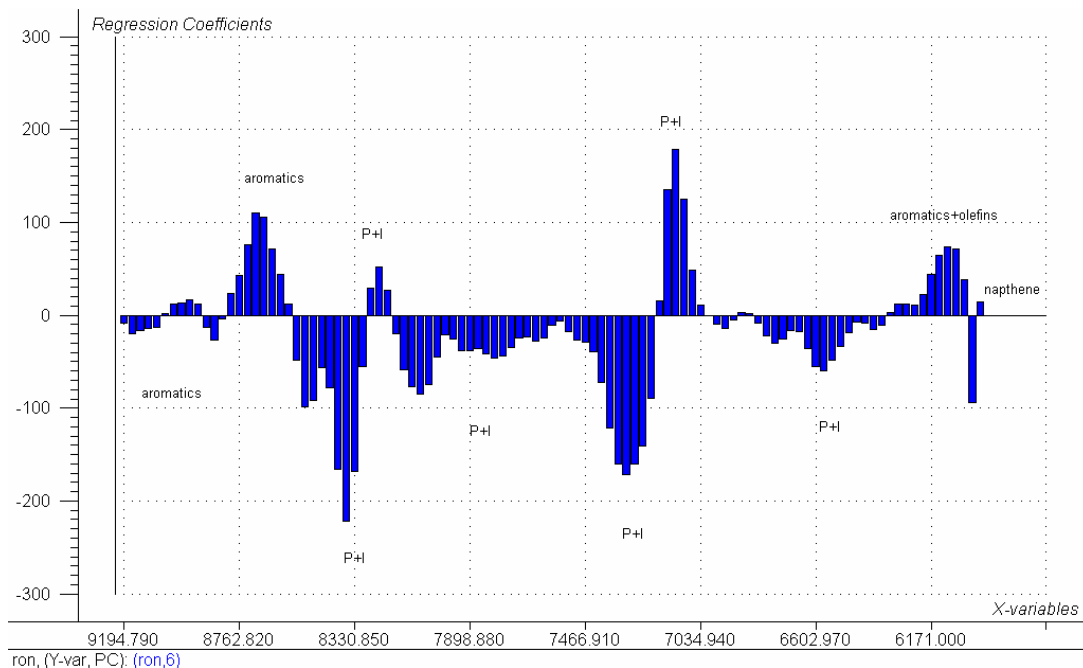


Fig 1: RON from a single model of merged samples from four feasible recipes.

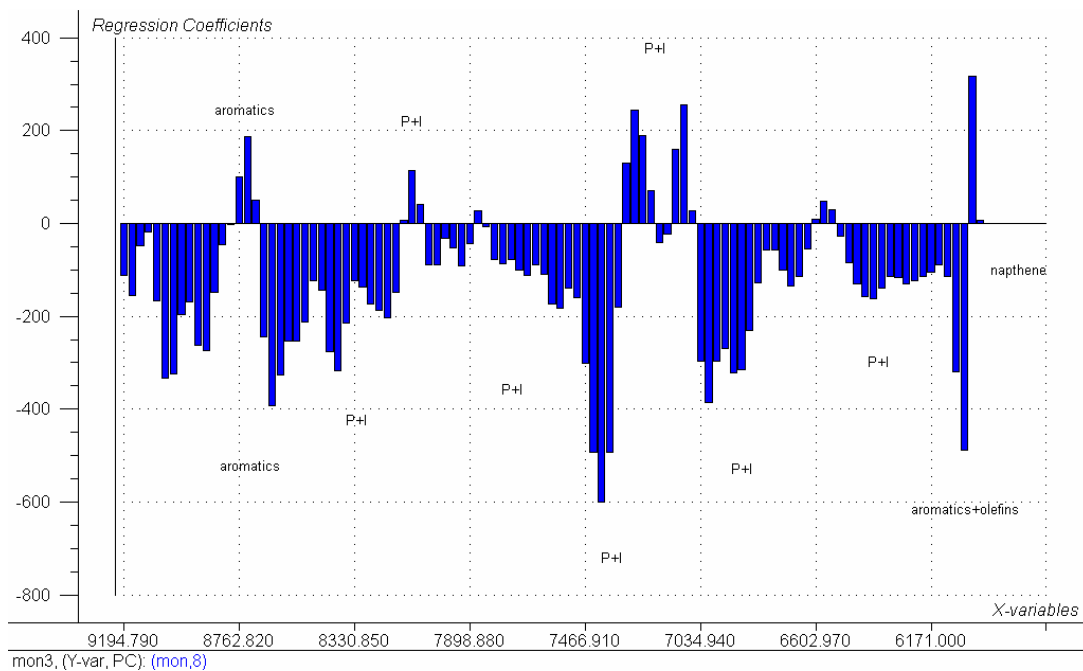


Fig 2: MON from a single model of merged samples from four feasible recipes.

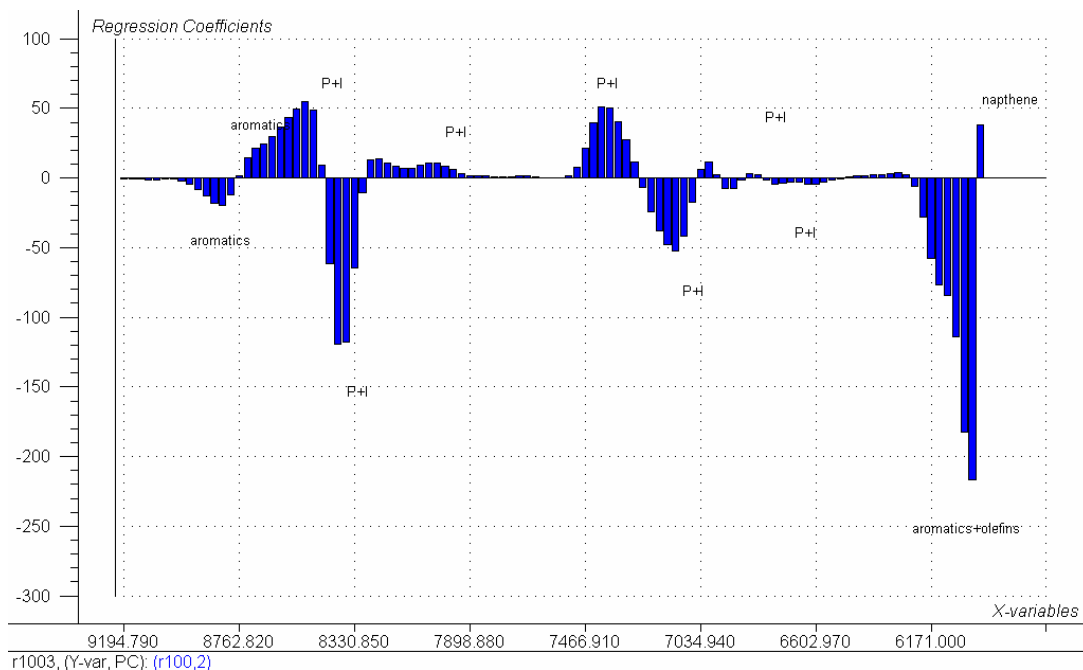


Fig 3: [Rec@100](#) degC from a single model of merged samples from four feasible recipes.

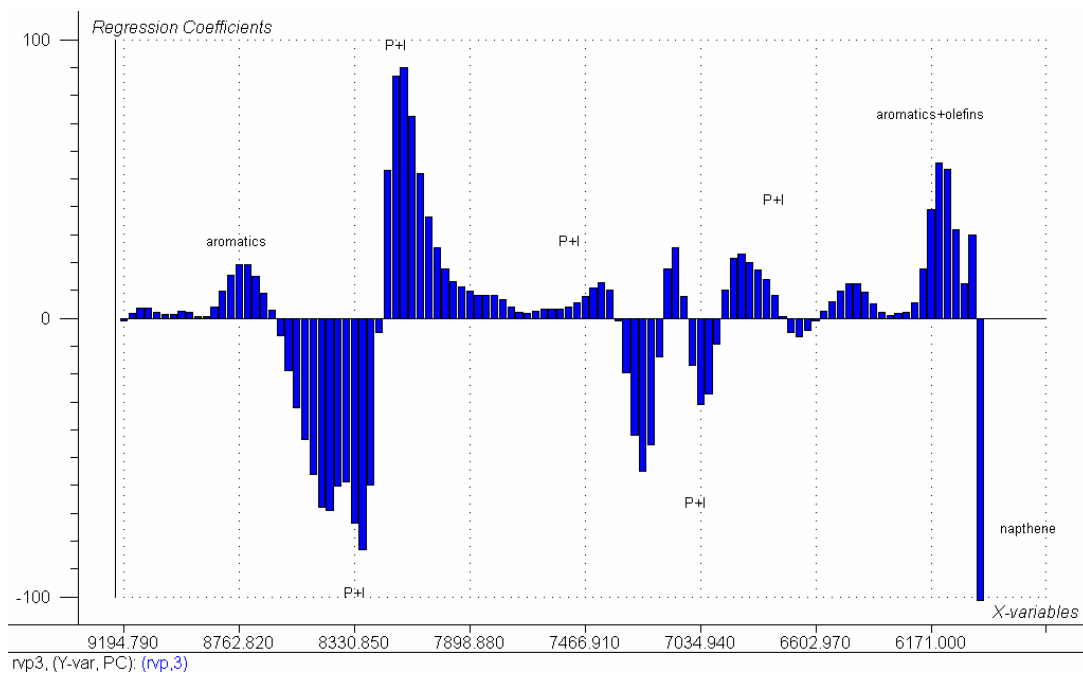


Fig 4: RVP from a single model of merged samples from four feasible recipes.

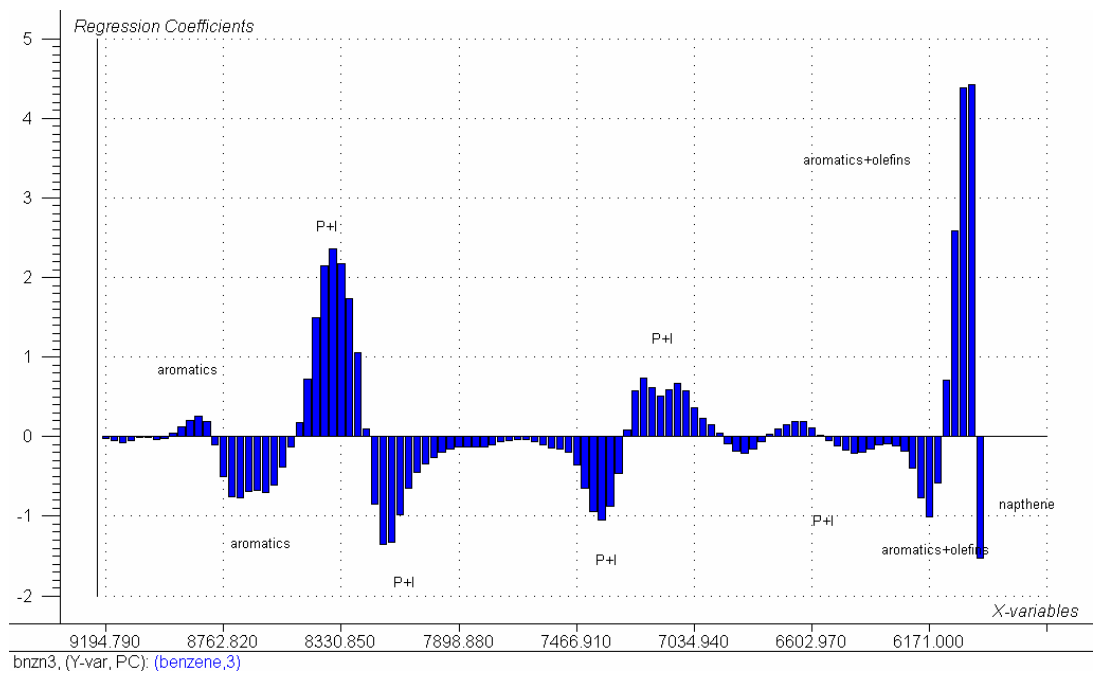


Fig 5: Benzene from a single model of merged samples from four feasible recipes.

APPENDIX 2:

Study for RON in various models due to changes in the regression coefficients in similar hydrocarbon structure absorbances (PIONA), by varying the blend recipes.

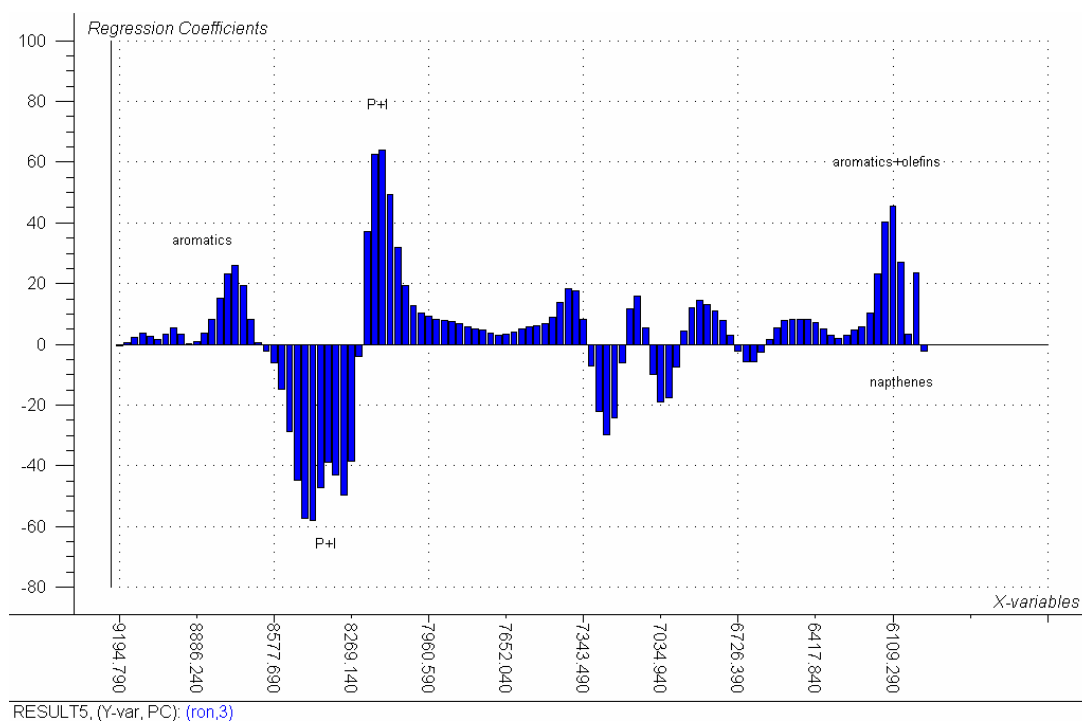


Fig 1: Result of RON from model of FCCG + CCG only. Shows a strong influence of regression coefficients in the aromatics, olefins and P+I absorbance ranges.

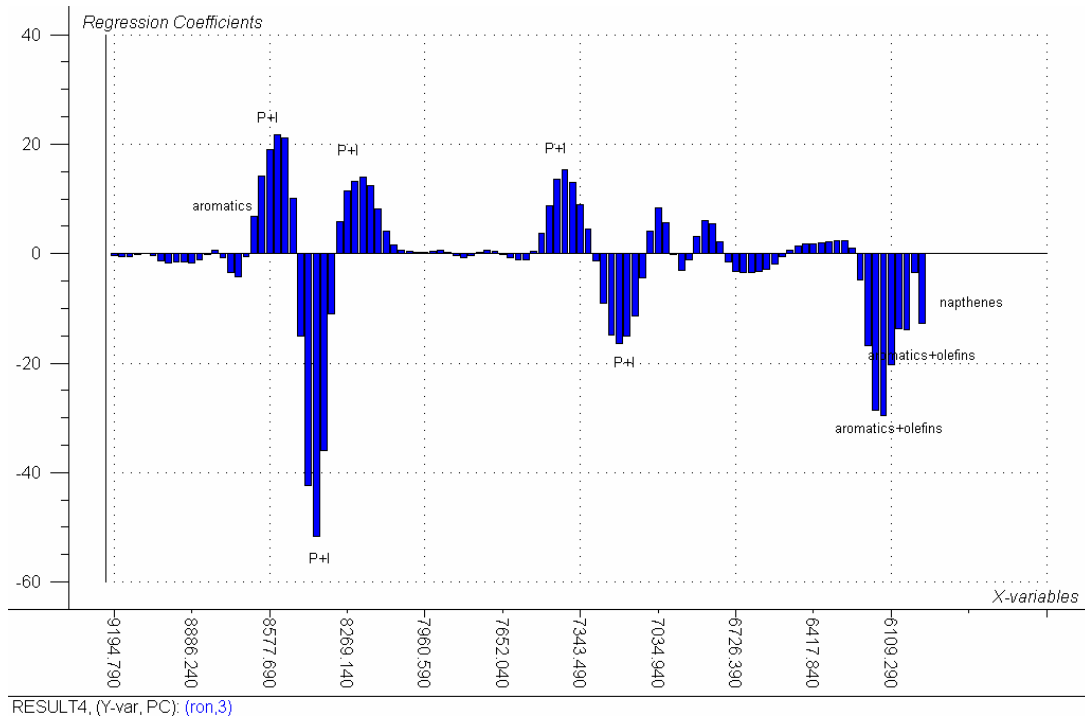


Fig 2: Result of RON from model of FCCG + CCG + reformate, low aromatic cut. Note changes in regression coefficients in P+I, aromatics and olefins absorbances. Will decrease RON from base case in Fig 1. Reformate is control component for RON.

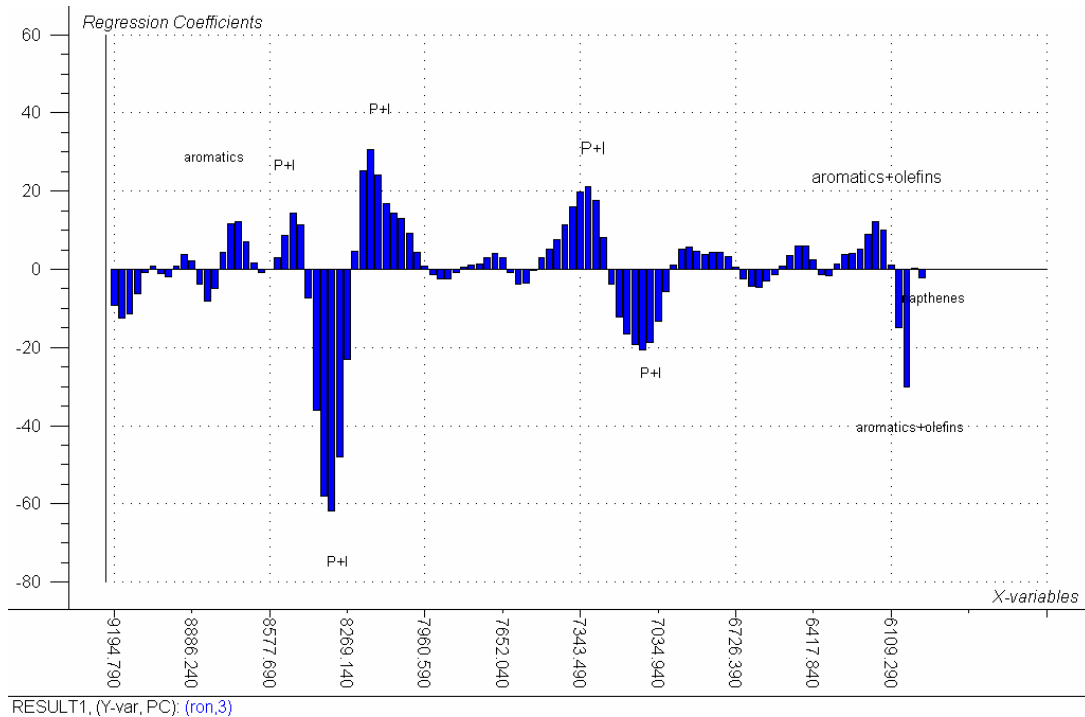


Fig 3: Result of RON from model of FCCG + CCG + naphtha. Note the changes in regression coefficients in P+I, aromatics and olefins absorbances. Will decrease RON from base case in Fig 1. Naptha is the control component for RON

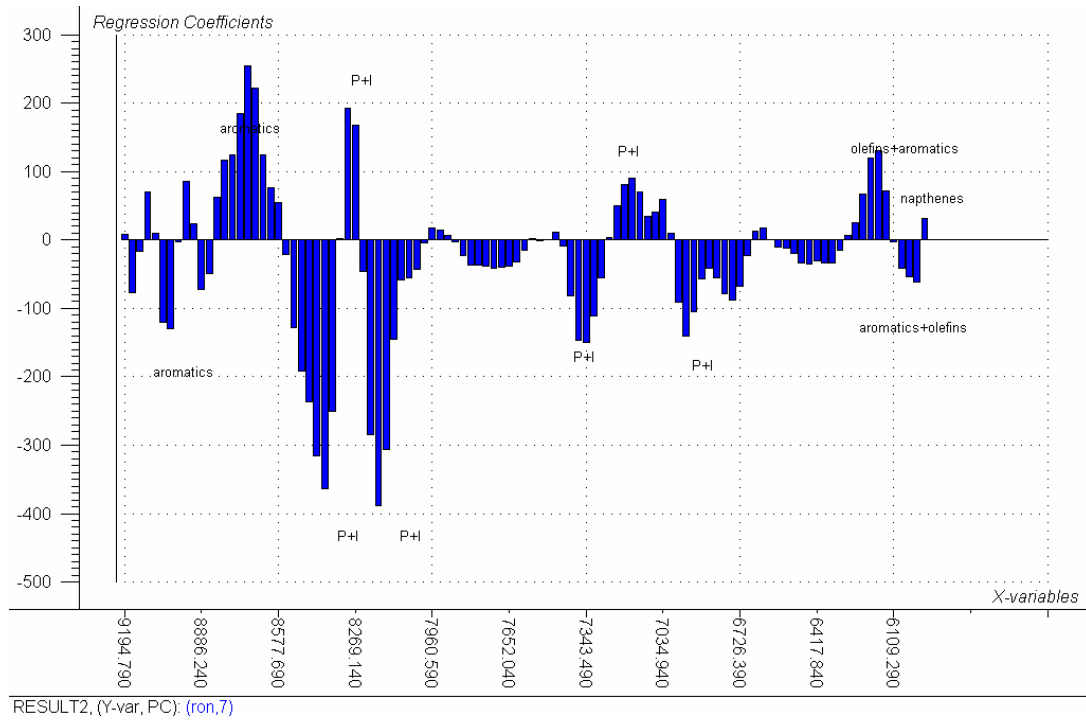


Fig 4: Result of RON from model of FCCG +CCG+raffinate. The strong regression coefficients in P+I absorbance will decrease RON drastically. Note the positive influence of the regression coefficients in aromatics absorbance, which is because of the fact that the raffinate is from ARU (aromatic recovery unit). Raffinate is also the control component for RON.

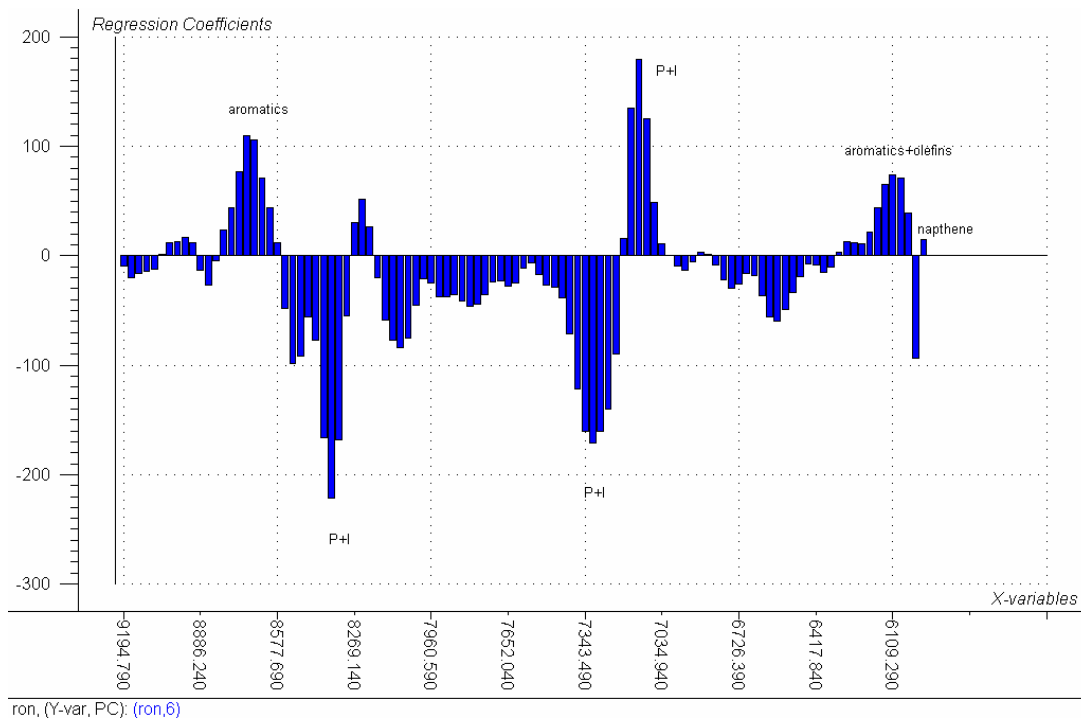


Fig 5: Result of RON from a single model by merging the four recipes. The changes in the regression coefficients in the PIONA absorbance ranges from the above models are noticeable. This model can predict for a sample belonging to any of the defined recipes as the sample population.