H. CHELARIU
I. M. STANCU-MINASIAN

Utilization of Statistical Mathematical Methods in Working out Quality Functions in Mixture Problems*

As a result of the increasing complexity of modern refineries, the economic analysis of crude-oil processing operations was transformed into a set of problems, numerous, varied and difficult to deal with, concerning their optimal and rapid solving.

Due to the complexity of processes in modern refineries, the economic analysis of crude-oil processing operations requires the solving of a set of numerous and varied problems, whose optimal and rapid solving is very difficult. Among the current problems of a refinery are: finding out the most economic assortments of crude-oil, from the viewpoint of processing; determining through computation the rated outputs of various processes and their optimization; costs determination for each product; establishing the optimal mixtures of petrols and gas-oils; the economic substantiation of new technological plants; settling the optimal processing plan of a refinery.

In the present paper, the approached problem is that of oil products mixtures: the carrying out of the mixture's physical model, i.e. the determination of functional relations between mixture characteristics and components' characteristics, by using statistico-mathematical models.

When modelling petrol mixture process [3], [5], [6], [7], the additivity of their main characteristics was used, a fact which permitted to

*) G.I.P. — Borzești Research Team: eng. H. Chelariu, eng. D. Ștefănescu, eng. C. Stancu, eng. C. Uzum, eng. I. Burlacu.
Research Team of the Computation Centre in the Economic Cybernetics Department: V. Ionescu, I. M. Stancu-Minasian, L. Spiru, Cs. Fabian.

express the mixture characteristics as a weighted mean of components' characteristics, i.e.:

$$C = \frac{\sum_{i=1}^{n} c_i x_i}{\sum_{i=1}^{n} x_i}$$

where:

C represents the mixture characteristic,

c_i — i-component characteristic,

x_i — amount of component i, taking part in the mixture.

Practically applied, for each particular case, this relation does not always offer the expected results, since its utilization may lead to obtaining mixtures with characteristics which deviate from the desired value, by amounts exceeding the limits accepted by standards (STAS). For these cases, the problem arises how to determine such a relation between mixture characteristic and components' characteristics, to obtain the characteristic's value, within the accepted limits.

In the concrete case, studied at the G.I.P.-Borzeşti refinery, the following situation exists:

The refinery produces petrol types as follows:

— catalytic-cracking petrol (CC)
— thermal-cracking petrol (CT)
— reformed petrol (heavy reformed RC)
— aromatized petrol
— refined petrol (refined RC)
— heavy xylenes
— normal pentane (nC5)
— isopentane (C5).

For commercialization, petrol mixtures are carried out, of which following products result:

— petrol with octane number 75 R
— petrol with octane number 88 R
— petrol with octane number 90 R.

Since petrol with octane number 88 R is directly obtained from catalytic-cracking petrol, with no admixture of other components, there arises no problems. Therefore, in this paper, we shall leave aside this commercial assortment.

When analyzing the technical characteristics of components and products obtained by mixture, it was stated that octane number and distillation are the most important. The other technical characteristics are implicitly achieved, since their values, for the components, are situated under the limits required by the STAS (Romanian standards) for the commercial product, and no abnormal phenomena take place by mixing.

Starting from these remarks, when establishing the physical model * of the mixture, only the evinced characteristics (octane number, and distillation at 10 cc, 50 cc, 90 cc and final) were taken into consideration.

Since it resulted, from the achieved laboratory experiments, that both the octane number and the distillation have not an additive behaviour, we went further to statistically determining the functional relation between mixture characteristic and components characteristics.

With a view to facilitating the solving of the complete mixture model, we tried to render the respective relations linear, i.e. of the from:

$$y = \frac{\sum_{i=1}^{n} k_i \theta_i x_i}{\sum_{i=1}^{n} x_i} = \sum_{i=1}^{n} k_i \theta_i \left(\frac{x_i}{\sum_{i=1}^{n} x_i} \right) = \sum_{i=1}^{n} k_i \theta_i x_i' = \sum_{i=1}^{n} a_i y_i \qquad (1)$$

where x_i $(i = 1, 2, \ldots, n)$ represents the amount of component i, entering the mixture; θ_i — i-component characteristic; y — mixture characteristic; k_i — sought for coefficients of regression line. Though a nonlinear programming model would have better reflected the existing correlations, it would have been more difficult to approach in the present stage, therefore, in order to facilitate the solving of the mixture problem, we nevertheless resorted to a linear relation.

We shall use, in determining coefficients k_i, the correlation method consisting in numerically expressing the link between two or several magnitudes, on the basis of a sufficient number of observations.

According to its analytic expression, the correlation may be linear or nonlinear. The linking line between variables can be mathematically described by means of a regression function. We used such a function, in which we considered products $\theta_i x_i$ as independent variables, and we attempted to determine a linear regresssion of the from

$$y = \sum_{i=1}^{n} a_i y_i \qquad (2)$$

The determination of the corresponding regression area was carried out in the following stages:

1. determination of an experimenting plan, which could cover the variation intervals of the independent variables (amounts composing the mixture), within the hypothesis of a small and random variation of mixture components (variations determined by the production process);

2. achieving of the scheduled experimenting plan;

3. drawing up of experimental-data processing programs for the I.B.M. 360/40F computer;

*) The physical model refers to quality relations, while the mixtures general model also includes quantity relations, as well as the function to be optimized.

4. experimental-data processing and interpreting, on the basis of the computer program;

5. refining the obtained results.

I. EXPERIMENTING PLAN

We mention that, in the calculations, values x_i were replaced by x_i', which represent percents, thus rendering the relation applicable for any unit of measure, hence also for production and laboratory. Relation (1) must be valid for any percent x_i' from 0 to 100.

The laboratory experiments were conducted in such a way, that points $(x_1, x_2, ..., x_n)$ be distributed as uniformly as possible in domain

$$R^n = \{(x_1^*, ..., x_n^*);\ 0 \leqslant x_i^* \leqslant 100,\ i = 1, ..., n\}$$

of the n — dimensional space.

Ten levels were considered for each variable, a fact which lead to a very high number of experiments (10^8), a number exceedign by far the laboratory's capacity. This is why, it was decided to carry out 250 experiments determined on the basis of some generated random numbers. Each number has eight digits, corresponding to the eight components forming the mixture.

With a view to determining the amounts composing the laboratory mixture, from each component the unit amount, used in preparing mixtures (namely a tenth of the mean daily production — when taking also routine inspections into account), was considered.

The actual amounts composing the mixture were obtained by multiplying the figure corresponding to the component, in the generated random number, by the corresponding component unit.

Starting from this experimenting plan, the laboratory achieved the respective mixtures and analyzed, both the characteristics of the respective components, and those of the obtained mixture.

When estimating results obtained from data processing, one must also take into account the analysis errors which can occur with the experimental values; these errors are admitted according to STAS, as follows:

— for octane number:

1) parallel determinations — the same engine or the same operator $\pm$ 0.5 octane

2) parallel determinations — various engines and various operators $\pm$ 1 octane

— for distillation

1) parallel determinations — the same engine $\pm$ 1 cc
— the same apparatus $\pm$ 0.5 cc

2) parallel determinations — various operators or various laboratories, 4 cc for initial, 2 cc or 1 cc for intermediary and final, and 0.2 cc for residue.

II. EXPERIMENTAL–DATA PROCESSING PROGRAMS

The main program for experimental-data processing, with a view to determining the regression line of each characteristic, is the REMUS (Selective Multiple Regression) program [4]. This program also determines, simultaneously with the regression-line coefficients, a series of statistical indexes, among which the multiple correlation coefficient, the mean square deviation, the variation coefficient, which allow to estimate the quality of adjusting the determined regression surface.

In order to better establish to what extent the determined functional relations correctly reflect the mixing process, differences $y_{calc.} - y_{exp.}$ between mixture experimental and calculated characteristics were computed. These differences, considered for all experiments, permit us to obtain mixtures whose characteristics are within the limits required by STAS.

III. DATA PROCESSING AND INTERPRETATION OF RESULTS

Since, when carrying out planned experiments, one stated that all components' characteristics varied from an experiment to another, because of the production process, there arised the idea of obtaining two model types, namely:

a. a model of the form

$$y = \sum_{i=1}^{8} a_i (c_i x_i)$$

where products $c_i x_i$ were considered as variables, so that coefficients a_i also contain the c_i perturbations of characteristics.

b. a model of the form:

$$y = \sum_{i=1}^{8} a_i x_i + \sum_{i=1}^{8} b_i c_i$$

where the c_i perturbations of characteristics are included in coefficient b_i.

Starting from these considerations, as well as from the necessity of obtaining a model as good as possible, a series of selections of the 250 experiments was done, and 10 models were obtained, namely;

A. Form a) models (with 8 initial variables)

A1. Model including all unselected 250 experiments

A2. Model which retained only experiments for which the values of the respective mixture characteristic were placed within the interval $M \pm \sigma$, where M is the mean of the values, and σ — the mean square deviation.

A3. Model which retained only those experiments which were placed, from the viewpoint of the mixture characteristic, within the interval $M \pm \sigma$ — simultaneously for all five characteristics (CO, 10 cc, 50 cc, 90 cc and final). Totally, 101 experiments were retained in this model.

Table I

	Statistical index Model Characteristic		Multiple correlation	Multiple determination	Mean square deviation	Variation coefficient	Mean	Standard deviation	% of deviations within STAS limits
A1	250 experiments unselected *R*	C.O.	0.918	84.31	1.218	1.42	85.57	3.0763	65.60%
		10CC	0.934	87.20	2.400	3.67	65.48	6.7097	83.60%
		50CC	0.774	59.91	8.992	8.11	110.87	14.2020	62.40%
		90CC	0.843	71.07	2.605	1.55	167.99	4.8432	75.20%
		final	0.466	21.68	7.517		197.218	8.4937	60.80%
A5	116 experiments unselected *R*	C.O.	0.900	81.06	0.7239	0.82	88.2	1.6633	82.75%
		10CC	0.931	86.65	2.743	4.22	65.06	7.5075	74.06%
		50CC	0.902	81.35	4.507	3.94	114.31	10.4365	66.38%
		90CC	0.827	68.43	2.416	1.43	169.107	4.2991	74.13%
		final	0.403	16.25	9.512	4.82	197.323	10.3936	49.13%
B1	250 experiments *E*	C.O.	0.928	86.13	0.9114	1.06	86.3639	2.4475	81.01%
		10CC	0.916	83.85	1.602	2.44	65.7589	3.9881	92.89%
		50CC	0.857	73.38	3.051	2.73	111.9707	5.9143	81.5%
		90CC	0.812	65.98	3.165	1.90	166.7263	5.4269	81.31%
		final	0.822	67.60	2.395	1.21	197.7907	4.2076	84.14%
A3	101 experiments from 250, no deviation *R*	C.O.	0.902	81.32	0.9232	1.07	86.2178	2.1358	77.22%
		10CC	0.895	80.18	1.704	2.62	65.1337	3.8279	94.06%
		50CC	0.500	24.97	8.964	7.94	112.9505	10.3485	59.4%
		90CC	0.762	58.04	2.382	1.42	168.0396	3.6767	84.15%
		final	0.672	45.16	2.514	1.28	197.0693	3.3948	82.17%

A2	250 exper., by eliminating as against mean *R* by independently eliminating	C.O.	0.896	80.22	1.089	1.26	86.3639	2.4475	70.37%
		10CC	0.917	84.14	1.588	2.41	65.7589	3.9881	88.83%
		50CC	0.595	35.43	7.072	6.29	112.3750	8.8006	72.87%
		90CC	0.826	68.19	2.706	1.62	167.128	4.7969	70%
		final	0.671	45.00	3.118	1.58	197.8062	4.2043	67%
A6	77 exper. out of 116 *R*	C.O.	0.798	63.69	0.5290	0.60	87.89	0.8779	96.10%
		10CC	0.932	86.80	2.386	3.70	64.5649	6.568	80.52%
		50CC	0.911	82.92	3.717	3.28	113.21	8.9951	68.83%
		90CC	0.816	66.60	2.479	1.47	168.8182	4.289	77.92%
		final	0.387	14.96	11.12	5.65	196.7143	12.0544	50.65%
B2	101 exper. out of 250, no deviation *E*	C.O.	0.921	84.75	0.834	0.97	86.217	2.135	83.16%
		10CC	0.903	81.54	1.645	2.52	65.133	3.827	94.06%
		50CC	0.843	71.09	3.019	2.69	112,123	5.615	68.31%
		90CC	0.809	65.49	2.210	1.32	167.94	3.7618	91.09%
		final	0.701	49.10	2.415	1.23	197.064	3.3858	83.16%
A4	190 exper. out of 250 R	C.O.	0.917	84.09	1.169	1.36	85.751	2.9317	72.10%
		10CC	0.906	82.05	1.953	3.00	65.168	4.609	90.00%
		50CC	0.686	47.00	7.241	6.46	112.039	9.9464	68.44%
		90CC	0.802	64.34	2.414	1.43	168.231	4.0418	78.42%
		final	0.721	51.97	3.038	1.54	197.55	4.3834	72.63%
B3	190 exper. out of 250 R	C.O.	0.947	89.76	0.9383	1.09	85.75	2.9317	78.42%
		10CC	0.904	81.79	1.967	3.02	65.168	4.609	87.36%
		50CC	0.875	76.56	3.695	3.31	111.57	7.6326	54.21%
		90CC	0.852	72.61	10.67	6.72	158.79	20.3953	43.15%
		final	0.817	66.77	2.531	1.28	197.53	4.3909	79.47%
B4	77 exper. out of 116 E	C.O.	0.7966	63.4639	0.53065	0.6038	87.893	0.8779	94.80%
		10CC	0.9384	88.0612	2.26941	3.5149	64.5649	6.5680	88.31%
		50CC	0.9228	85.1531	3.4563	3.0519	113,253	8.9702	58.44%
		90CC	0.8846	78.25	9.4868	5.9899	158.380	20.3449	42.85%
		final	0.4236	17.9455	10.926	5.5553	196.688	12.062	46.75%

A4. Model which retained only those experiments which were included into the interval $M \pm \sigma$, for at least four out of the five characteristics (totally 190 experiments).

A5. Model which retained only those experiments for which the mixture's octane number was within the interval 86—93. This model was imposed by the fact that the 250 experiments did not cover also product CO (octane number) 75, hence a general model could not be determined for both products (CO 75 and CO 90). This is why we tried to obtain the model for CO 90, retaining, from the 250 experiments, all those which covered octane number 90 (totally 116 experiments).

A6. Model which retained, from the 116 experiments in model A5, only these for which mixture's octane number was placed into the interval $M \pm \sigma$.

B. Form b) models (with 16 initial variables)

B1. Model which retained, for each characteristic, only those experiments for which the respective mixture characteristic is placed within the interval $M \pm \sigma$.

B2. Model in which those experiments were selected, for which all five characteristics had mixture values included into the interval $M \pm \sigma$ (101 experiments).

B3. Model which retained experiments for which at least four from the five characteristics had values within the interval $M \pm \sigma$.

B4. Model corresponding to A6, but obtained for 16 variables.

Statistical indexes (multiple correlation, multiple determination, mean square deviation, variation coefficient, mean, standard deviation) as well as deviation percentage, within the STAS limits, are presented in Table 1.

CONCLUSIONS

1°. Since the characteristics of the 8 components, which cooperate to bring about final products CO 75 and CO 90, have a great variation area (e.g., octane number varies from 63 to 120), a unique model for both final products cannot be determined, a fact which lead us to the attempt of obtaining, from the 250 experiments, a model for CO 90, since these experiments had values close to 90. But it is found, by examining, column „Mean" in Table 1, that, since the experiments were not directed only towards CO 90 from the beginning, the obtained models lead to octane numbers of the mixtures in the range 82—89.8. Experiments must be completed with other ones, for covering the range 88—92, as to obtain a resulting octane number of at least 90. It is also necessary to plan and to carry out another set of experiments for CO 75.

2°. From the laboratory checkings of both model types, *a* and *b*, as well as from the study of the parameters in Table 1, it results that *b*-type models are much better, as concerns accuracy, but more difficult to apply, because both the amounts x_i, and the characteristics of components C_i, must not be outside the variation interval, resulting from the processing of the 250 experiments. This requirement cannot practically be achieved,

since the components obtained in the plants also have characteristics outside these limits. Such a situation is a consequence of the fact that experiments cannot be caried out in due time, for uniformly covering the variation intervals of the characteristics; as a matter of fact, they can be directed only to uniformly cover amounts' variation intervals.

3°. From the study of the statistical indexes (Table 1) and of the deviation percents, situated wthin STAS-limits, models A3, A4 and B1 must be retained, which have to be verified (as concerns the accuracy), with the help of 20 laboratory experiments (these 20 experiments should be also used to check the other models).

Verification will lead to selecting the model which will bring about the desired mixture, with the highest probability. The extension of the model to CO 90 will be done by completing the experiments with other ones, which would lead to the uniform covering of interval 85—95, and which will be subsequently processed according to the same methodology. Relations obtained in this manner will make up the quality restrictions of the mixtures' general model.

4°. In the case when, by the obtained linear models, the desired accuracy is not attained, nonlinear models, obtained by processing the same experimental data, or otherwise shaped models, met with in the speciality literature, will be tested.

REFERENCES

1. AUSLÄNDER G.: *The Properties of Mixtures.* Brit. Chem. Eng. 9, 610, 1964.
2. BERLINE CH., WELS, J.: *Etude des problèmes économiques du raffinage à l'aide de la théorie des programmes linéaires et des calculateurs électroniques.* (World Oil Congress, New York, 1969).
3. DUMITRU V. *O aplicație a programării liniare în problema amestecurilor de benzine* (An Application of Linear Programming on the Petrol Mixtures Problem). Revista de Statistică, 10, 1964.
4. DUȚĂ D., FABIAN Cs., MARINOIU V.: *Determinarea pe cale statistică a ecuațiilor proceselor industriale complexe (II).* (Statistical Determination of Equations of Complex Industrial Processes). Studii și cercetări de calcul economic și cibernetică economică, nr. 1, 1971.
5. DUMITRU V., IONESCU V., BĂRBATU I. GH.: *A non-Linear Model of Optimization for Gasoline Blends with Tetraethyl Lead Additive.* Economic Computation and Economic Cybernetics Studies and Research, 3, 1968.
6. MĂNESCU M., DUMITRU V.: *Some Applications of Linear Programming in Problems of the Petroleum Industry.* Studii și cercetări economice, 4, 1965.
7. MĂNESCU M., DUMITRU V., BĂRBATU I. GH.: *Programarea matematică în industria petrolieră.* (Mathematical Programming in the Oil Industry). Editura Academiei R. S. România, 1970.
8. MĂNESCU M.: *Orientări și tendințe în cibernetica economică.* (Orientation and Trends in the Economic Cybernetics). Studii și cercetări de calcul economic și cibernetică economică, 1, 1968.
9. MAURIN H.: *Programmation linéaire appliquée.* Editions Techniques, Paris, 1967.
10. YULE G. U., KENDALL M. C.: *Introducere în teoria statisticii* (Introduction in the Theory of Statistics). Editura Științifică, 1969.

L'UTILISATION DES METHODES STATISTIQUES MATHEMATIQUES POUR ELABORER LES FONCTIONS DE QUALITE DANS LES PROBLEMES DE MELANGE

L'un des problème autrement important de l'industrie pétrolière est celui des mélanges de produits pétroliers, à savoir la réalisation du modèle physique du mélange, c'est-à-dire la détermination des relations fonctionnelles entre la caractéristique d'un mélange et les caractéristiques des composants.

Dans une série de travaux de spécialité, on prend comme point de départ l'activité des caractéristiques principales des essences, ce qui a permis de présenter une caractéristique du mélange comme une moyenne pondérée des caractéristiques des composants. Le plus souvent une telle relation aboutit à l'obtention de mélanges ayant des caractéristiques qui s'écartent de la valeur imposée.

Le présent article essaie de déterminer quelques relations semblables pour les principales caractéristiques des essences, utilisant les méthodes statistiques mathématiques et leur application à une grande raffinerie. Le cas réel étudié se rapporte à huit composants desquels on doit obtenir, en tant que produit final, l'essence ayant 90 pour chiffre octanique.

ПРИМЕНЕНИЕ СТАТИСТИКО-МАТЕМАТИЧЕСКИХ МЕТОДОВ ПРИ РАЗРАБОТКЕ ФУНКЦИЙ КАЧЕСТВА В ЗАДАЧАХ СМЕСИ

Одним из важных вопросов нефтяной промышленности является вопрос смесей нефтепродуктов, а именно осуществление физической модели смеси, то-есть определение функциональных отношений между характеристикой смеси и характеристиками компонент. В ряде специальных работ исходили из активности основных характеристик бензинов, что позволило выразить характеристику смеси как средневзвешенную характеристик компонент. В большинстве случаев такое отношение приводит к получению смесей с характеристиками, отклоняющимися от заданной величины.

В настоящей статье делается попытка определить такие отношения основных характеристик бензинов при помощи статистико-математических методов с применением их на крупном нефтезаводе. Рассматриваемый конкретный случай касается восьми компонент, их которых нужно получить в качестве конечного продукта бензин с октановым числом 90.