FACTOR ANALYSIS OF A COMPLETE EXPERIMENT

By

J. BORJÁN—Z. FEKETE—O. OROSZI

Department of Building Materials, Technical University, Budapest

Received: November 5, 1980

Presented by Associate Prof. Dr. GY. BALÁZS, Head of Department

Empirical relationships are advisably established from results of planned tests. In planning the experiment, a system of factor levels is acting on the test object, hence adjustments are planned such that yield solutions of adequate accuracy [1].

A complete experimental adjustment is that involving all possible combinations of assumed factors and factor levels [2].

The classic test of factor effects is the variance analysis.

Here a new method for analyzing the effect and interaction of factors, different from variance analysis, will be presented that is simpler and has a wider range of applications. Its essential is to express the effect and interaction of factors in terms of the deviations of empirical distribution functions and quantile functions [3].

1. Fundamentals

Test object is the measurable quantity the test is made for. Test object may be one or more variables or their interrelation.

Factor is the measurable variable quantity taking a determined value under given circumstances. The factor suits to adjust the test object.

Factor level is the adjustable value of the factor achievable under determined, real circumstances.

Adjustment. In experiments, several factors act simultaneously. Every factor is given a determined or planned value. An actual system of factor levels is the adjustment.

Empirical relationships are functional relationships of two or more variables permitting to mutually deduce each other's values.

Two or more stochastic variables can only be related by *stochastic relationships*. Also in their occurrence, mathematical methods are applied to find an unambiguous function possibly best expressing the character of the relationship and deducing from the measured value of one variable the non-measured values of the other variable with as little error as possible.

Regression analysis is a mathematical model suiting to relate stochastic variables under determined conditions. Its initial condition is that values of the one variable are exact, and those of the other are elements picked out at random of a random variable of determined distribution.

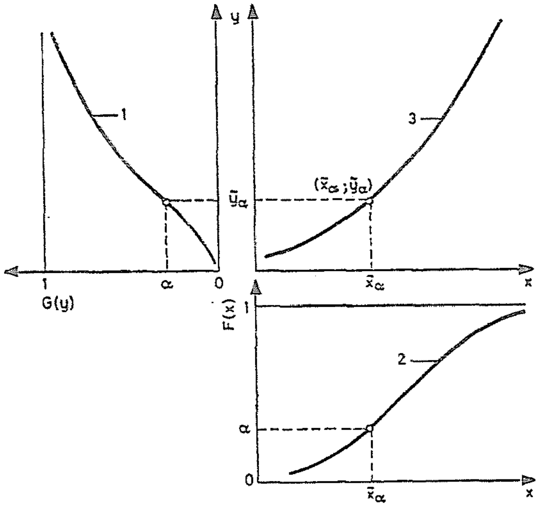

Fig. 1. Principle of plotting quantile curves: 1 — distribution function of variable y; 2 — distribution function of variable x; 3 — quantile curve

Quantile function. For known, monotonously increasing distribution functions of random variables, and known overall distribution of the variables, points determined by identical probability levels of distribution functions (related quantile points) lie on the quantile function [4]. Regression function deviates from the quantile function. Principle of plotting quantile curves is seen in Fig. 1.

2. Expression of factor effects

In variance analysis, factor effects are expressed as difference of measurement results on the test object due to the change of factor level. In multi-level, repeated tests, significance of the differences is examined.

In the following, the factor effect will be expressed as deviation between variable distribution functions and empirical quantile functions as seen in Fig. 2.

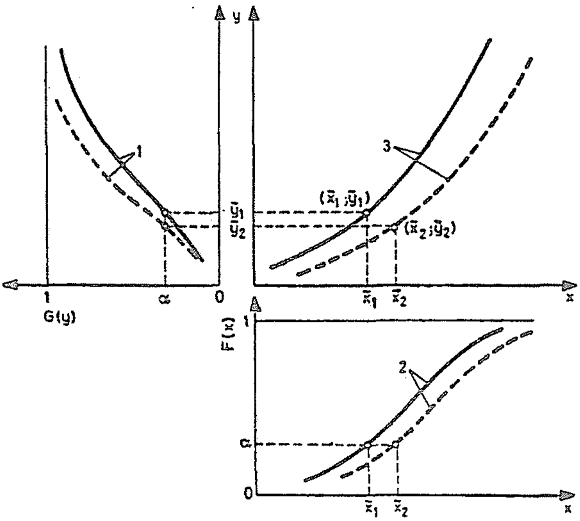

Fig. 2. Representation of factor effects: 1 — deviation of distribution functions of variable y; 2 — deviation of variable x; 3 — deviation of quantile curves

Distribution functions will be plotted by sorting the overall test results according to levels of the just examined factor and plotting separate distribution functions as well as quantile functions for part sets [5].

In speaking of a factor effect, it has to be indicated whether effect of the given factor on the distribution of variable y or x or on their correlation function is examined.

A factor may have different effects on a variable [6]:

- If the factor is irrelevant to the variable, the distribution curves coincide.
- If the factor has a constant effect, the distribution curves are parallel shifted.
- If the factor has a varying effect, then the slope of the distribution curves, hence also the statistic characteristics of standard deviation change.
- Even the distribution type may change.

A factor may have different effects on the correlation between two variables:

- If the factor is irrelevant to the relationships, the quantile curves coincide. This may be the case even if the tested factor has different effects on each distribution curve, a case represented in Fig. 3.

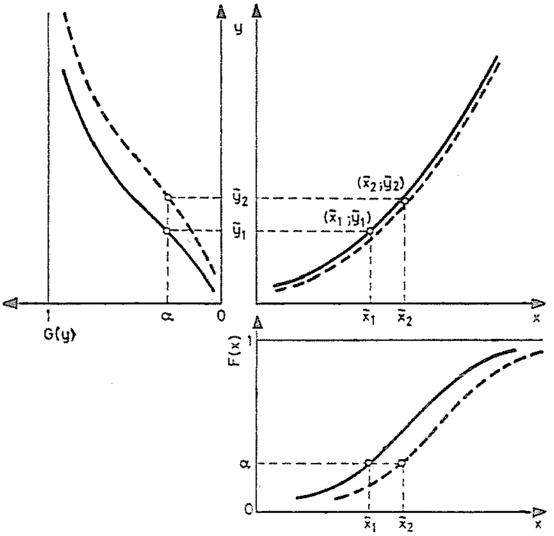

Fig. 3. The tested factor affects distribution curves rather than quantile curves

— Upon a constant factor effect the quantile curves are parallel shifted.
— A uniformly varying factor effect causes the quantile curves to proportionately diverge.
— If the factor has a cyclic varying effect, the quantile curves may be parallel over a section, and diverge over the other.

Variance analysis does not suit to demonstrate such differences in factor effects.

3. Expression of the interaction between factors

Interaction is the term for the phenomenon where the effect of one factor at one level of the other factor differs from that on its other level. The interaction is expressed by the calculated deviation of differences (effects) measured at each level. In variance analysis, significance of deviations is examined in multi-level, repeated tests.

Here the interaction is suggested to be expressed in terms of deviations between distribution functions and empirical quantile functions of the variables.

Distributions functions are plotted by sorting overall test results according to level combinations of the examined factors, and plotting separately both distribution and quantile functions for each part set. Overall test results will be sorted to as many part sets as the product of the numbers of examined factor levels. The principle of examining the interaction between factors is illustrated in Fig. 4.

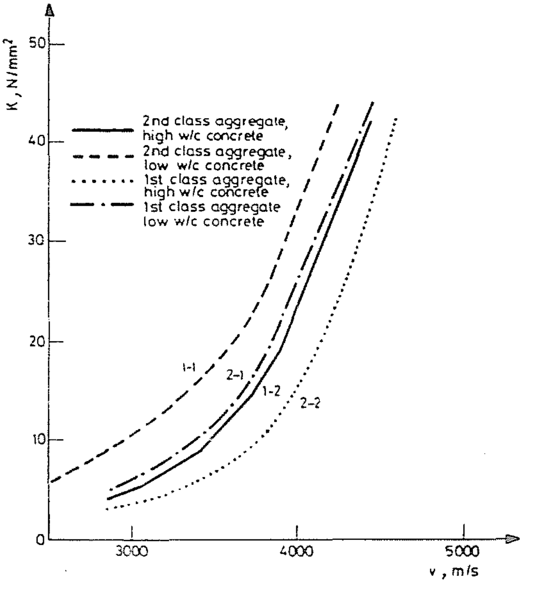

Fig. 4. Representation of factor interactions
1-1 — both factors are at level 1; 1-2 — one factor is at level 1 and the other at level 2; 2-1 — one factor is at level 2 and the other at level 1; 2-2 — both factors are at level 2

In examining multiple interactions, distribution curves will be plotted by sorting the overall test results according to level combinations of the actually tested (at least three) factors, and separately plotting distribution and quantile functions for each part set.

4. Plotting of function fields

Measurement results sorted according to all, or at least great many, factor level combinations suit plotting of a multitude of distribution and quantile curves permitting general description of the phenomenon. Combined representation of distribution functions, quantile functions for the entire test, as well as of curves of effects, interactions, multiple interactions of factors yields a curve field to be processed by mathematical-statistical methods [7]. An ultrasonic strength assessment function field is seen in Fig. 5 as an example.

Analysis of function field elements yields valuable data. For instance, variation of the strength distribution functions shows how the standard deviation of strength values depends on the mean value.

Variable transformation may ease general description of function fields. For instance, the function field for ultrasonic strength assessment is well approximated by a set of straight radii, after one variable (cube strength) has been logarithmized (Fig. 6). The only parameter of the set of radii is the line slope.

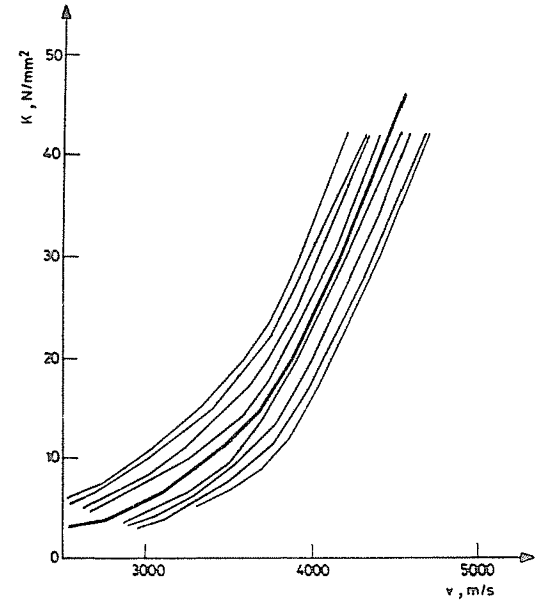

Fig. 5. Empirical curve field

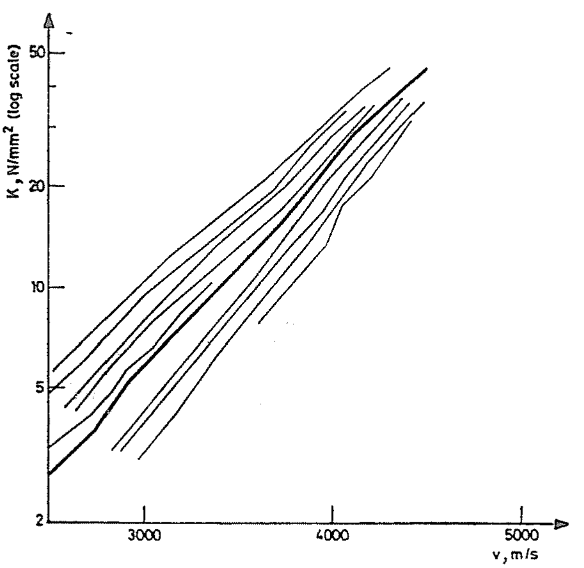

Fig. 6. Transformation of an ultrasonic strength assessment function field

Distribution function of this parameter as a random variable seen in Fig. 7 delivers probability levels of the empirical function field, presented, with its probability levels, by Fig. 8.

Function of the strength assessment function fields is, in general form:

$$\lg K = 2{,}407 - a_v \cdot 10^{-4} (5760 - v)$$

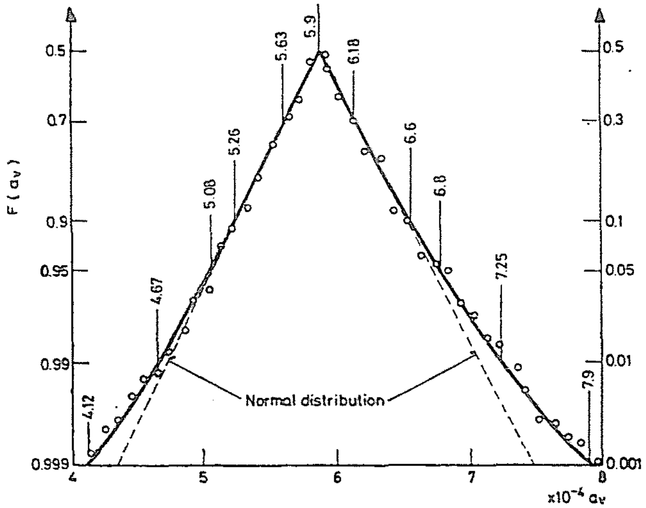

Fig. 7. Slope distribution of quantile radii

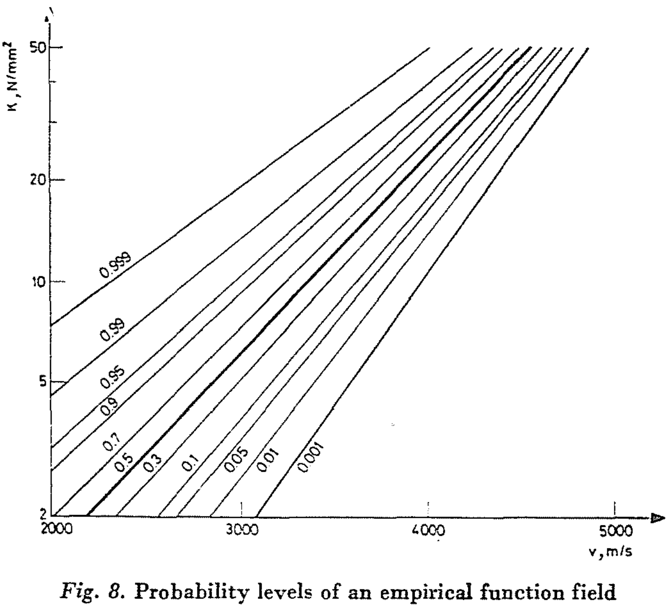

Fig. 8. Probability levels of an empirical function field

where:

- K — concrete cube strength, N/mm^2;
- a_v — constant variable for one element of the function field, and random variable of the entire function field;
- v — ultrasonic velocity, m/sec.

a_v values have been compiled in Table 1.

Table 1

Parameters of the strength assessment function field

$F(x)$	$a_v \cdot 10^{-4}$
0.001	7.90
0.01	7.25
0.05	6.80
0.1	6.60
0.3	6.18
0.5	5.90
0.7	5.63
0.9	5.26
0.95	5.08
0.99	4.67
0.999	4.12

5. Practical application

The general equation of the strength assessment function field can be fed into a programmable pocket calculator. Settling one parameter of the function field, any strength assessment function can be produced, permitting, in turn, to immediately calculate assessed strength values corrected according to concrete technology factors in field tests [7].

6. Plotting empirical functions — differently*

Students sharing the Department's research work on nondestructive concrete tests have developed a more exact and simpler method for plotting empirical relationships between stochastic variables, in particular, linear or linearizable functional relationships.

* Abridgment of a paper submitted to a Scientific Students' Circle Competition

Initial conditions of regression analysis for expressing stochastic relationships as functions are not always met, inducing researchers to plot two regression functions (Fig. 9).

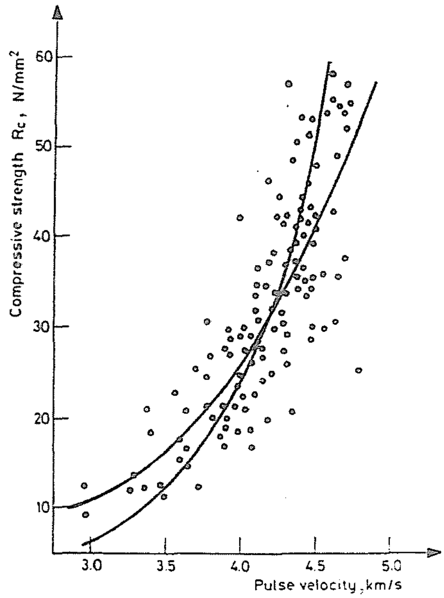

Fig. 9. Regression curves

Reimann suggests to express stochastic relationships by quantile functions [4].

For the discrete approximation of the quantile function, Borján [3] applies discrete distribution functions plotted from ordered samples of the variables, assigning identical quantiles of the ordered samples (Fig. 10).

Bun Chamroun Sok [9] fits regression functions to empirical quantile points by the least squares' method (Fig. 11), better approximating theoretical quantile functions than do regression functions for unique results, since discrete quantile points result from simultaneous minimization with respect to both variables. Difficulty of this method resides in ordering the variables.

Approximation of the theoretical quantile function without previously ordering the variables was attempted by finding statistics utilizing original data to directly deliver the theoretical quantile function.

The wanted function, optimizing along both x and y — if any — is assumed to proceed between regression lines fitted to quantile points and can be derived from them (using lines $y_x = a_x x + b_x$ and $y_y = a_y x \times b_y$) just

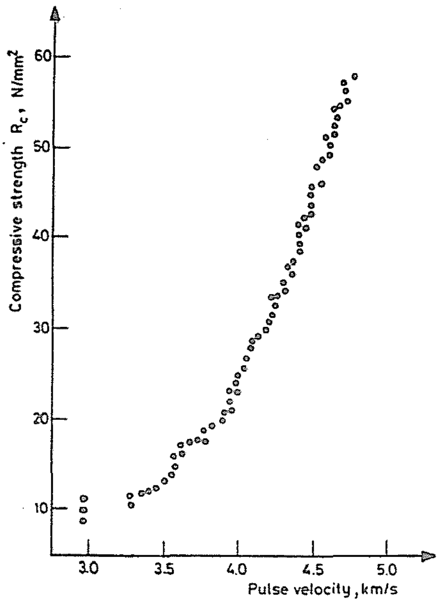

Fig. 10. Quantile curve points

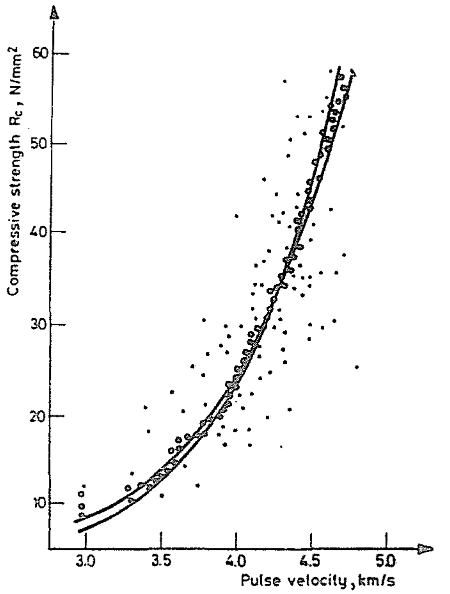

Fig. 11. Quantile point regression

as from regression line or the original set, (using lines $y_x^0 = a_x^0 + b_x^0$ and $y_y^0 = a_y^0 + b_y^0$). This condition is met by line

$$y_{xy} = \sqrt{m_x^0 \cdot m_y^0}\,(x - \bar{x}) + \bar{y}$$

where

- y_{xy} wanted function value at $x - \bar{x}$;
- m_x^0 slope of the regression line of y with respect to x;
- $x - \bar{x}$ distance of x from the mean set value along x;
- $\bar{y}$ mean value of the set with respect to variable y.

This line always proceeds between regression lines fitted to quantile points, and can be computed either from quantile points or from the original points, namely:

$$m_x m_y = m_x^0 m_y^0 = m_{xy}^2$$

where m_{xy} is the slope of the wanted line.

Substituting the known formulae of m_x^0 and m_y^0 into

$$m_{xy} = \sqrt{m_x^0 m_y^0}$$

yields a formula for direct calculation:

$$m_{xy} = \frac{\sqrt{\Sigma y^2 - \frac{(\Sigma y)^2}{n}}}{\sqrt{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}}$$

involving no product sum xy so that x and y values can be separately fed into the computer. The line slope (increasing or decreasing trend of the function) is decided by a separate analysis. Identical built-up of numerator and denominator is advantageous in programming.

The resulting line was found to simultaneously minimize with respect to both variables just as REIMANN demonstrated it for quantile functions. This line, however, minimized the product sum $(\Sigma|x| \cdot |y|)$ of absolute variable values in proceeding between the points.

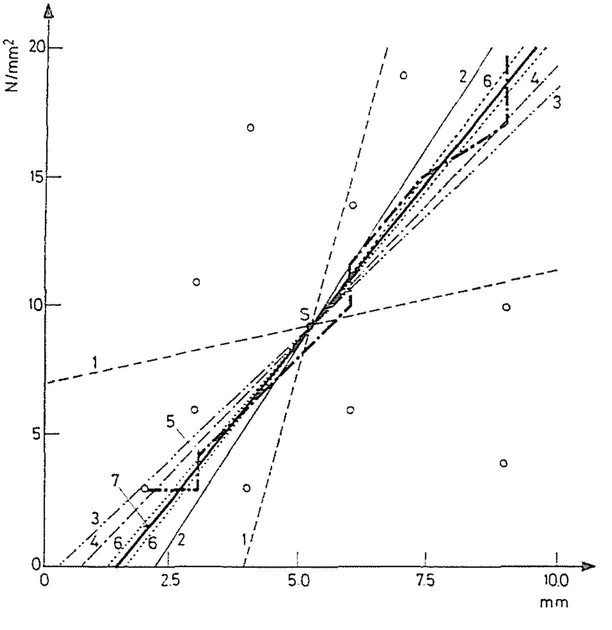

Fig. 12. Comparison of function plotting methods: 1 — regression lines fitted to original points by the least squares' method; 2 — line featuring the mean slope of regression lines; 3 — regression line bisector; 4 — main axis of inertia of points in the set; 5 — broken line of straight sections passing through quantile points; 6 — regression lines fitted to quantile points; 7 — line obtained by the suggested method

Figure 12 is a comparison between lines obtained by this method and by other, published methods, concerning a set of very loose correlation.

This method has been applied in practice for processing test results at the Department of Building Materials. Relationships for nondestructive characteristics and strength values were logarithmized, while relationships between concrete temperature and weather data involved original values [10].

Summary

A new method of factor analysis of a test with factors of complete experimental arrangement has been presented. Effects, interactions, multiple interactions of factors are expressed by deviations between empirical distribution functions and quantile functions of the variables. Quantile function is a set of geometrical loci of points defined by identical probability levels of variables. Function field determined by quantile functions is represented by statistical methods. The function field is regularly applied in practice in nondestructive strength assessments. A simplified method of plotting empirical functions is presented.

References

1. ADLER, YU. P., MARKOVA, E. V., GRANOVSKI, YU. V.: Planning Experiments for Determining Optimum Conditions.* Műszaki Könyvkiadó, Budapest, Mir Publishers, Moscow, 1977.
2. FELIX, M.—BLAHA, K.: Mathematical Statistics in the Chemical Industry.* Műszaki Könyvkiadó, Budapest, 1964.
3. BORJÁN, J.: Suggested Nondestructive Method for Strength Assessment. Per. Pol. Civ. Eng. Budapest, Vol. 22. (1978), No. 1—2.
4. REIMANN, J.: Mathematical-Statistical Analysis of Flood Characteristics.* Hidrológiai Közlöny, 4. 1975.
5. BORJÁN, J.: Tentative Standard Specifications for the Nondestructive Concrete Testing and Evaluation.* Manuscript, 1978.
6. BORJÁN, J.: Inhomogeneity in Structural Concrete and its Nondestructive Determination. International Symposium on Testing in-Situ of Concrete Structures. Budapest, 1977.
7. BORJÁN, J.: Schätzung der Druckfestigkeit des Betons aufgrund unterschiedlicher Kennwerte. Wissenschaftliche Berichte der Technischen Hochschule Leipzig. Heft. 10. 1979.
8. TALABÉR, J., BORJÁN, J., JÓZSA, ZS.: Effect of Concrete Technology Parameters on Non-Destructive Strength Assessment Relationships.* Scientific Publications No. 29 of the Department of Building Materials, Technical University, Budapest, 1979.
9. CHAMROUN, B. S.: Comparison of Test Results by Schmidt Hammer Types P and N.* Research Report, Department of Building Materials, Technical University, Budapest, 1976.
10. BORJÁN, J.: Nondestructive Testing of Concrete.* Műszaki Könyvkiadó, Budapest, 1981.

Senior Assistant Dr. József BORJÁN
Zoltán FEKETE, Olga OROSZI — students in architecture
H-1521, Budapest

* In Hungarian