

MATHEMATISCH CENTRUM

2e BOERHAAVESTRAAT 49

AMSTERDAM

STATISTISCHE AFDELING

Leiding: Prof. Dr D. van Dantzig

Chef van de Statistische Consultatie: Prof. Dr J. Hemelrijk

S 159

Statistical methods applied
to the mixing of solid particles, I.

by

J. Hemelrijk

1954

1. Introduction.

There is an extensive literature about the mixing of solid particles; concepts like "ideal mixing" or "complete mixing" and "degree of mixing" are often defined in different ways and reveal some confusion of ideas. A statistical approach seems to be the most appropriate one for the definition of these concepts, cf. LACEY [5], BUSLIK [2], BLUMBERG and MARITZ [1], and HERDAN [4].

These authors point out that it is hardly evitable to consider the problem statistically, moreover this has the advantage of providing statistical techniques for the investigation of mixtures. In this and subsequent reports some suggestions about this subject will be given and a number of statistical methods will be described, which can be applied to problems of practical importance concerning the mixing of solid particles.

A mathematical treatment of technical problems is always necessarily somewhat schematic. In this report only a very simple model will be considered; a mixture of two components consisting of particles of two types, A and B, equal in form and size, and only differing (for instance) in colour and possibly in weight. In practice this case will rarely occur; it is however useful for demonstrating the statistical character of the problems at hand and the statistical methods which may be used. This report is thus exclusively meant as a theoretical introduction to the statistical approach to the theory of mixtures and as an incomplete survey of the literature dealing with the subject. It is the intention to discuss more complicated mixtures in a later report.

In the following section a survey of the various definitions of complete mixing for this simple model, as stated by different authors, will be given and a discussion of these definitions is presented; then the concept "degree of mixing" and its statistical properties will be developed.

2. The concept of "complete mixing".

LACEY [5] quotes the following definitions of URE:

Definition a: A complete mixture is an entirely homogeneous mixture, i.e. a mixture of which every sample contains the same fraction of A-particles.

It is evident that one cannot adhere strictly to this definition, for, taking a sample consisting of one particle, this can only conform to the requirements stated in this definition

if the mixture consists of one component only and then it is no longer a mixture. Taking larger samples, it is still impossible to obtain always exactly the same fraction of A particles.

As a matter of fact this definition is only useful for mixtures of completely mixable fluids and the like and in the form given above it is not adaptable to solids (cf. NASKE [7] and LACEY [5]).

Definition b: A complete mixture is a mixture with the greatest possible regularity (cf. e.g. VALENTINE and MAC LEAN [9]).

This definition leads to a lattice for the particles, just as the ions in a crystal, as the ideal form of a mixture. Evidently this cannot be obtained with the usual mixing processes and this definition misses its aim (cf. also LACEY [5]).

Definition c: A complete mixture is a mixture which might have been obtained by distributing the particles of its components at random in the mixture.

This definition (given in a somewhat different form by LACEY [5] and BLUMBERG and MARITZ [1]) is based on a statistical approach and a small refinement makes it suitable for at least the above mentioned simple mixture of particles A and B, both components being exactly alike except e.g. in colour. This correction was given by BUSLIK [2].

In the first place it should be indicated that a mechanical mixing process, which is not influenced by the colour of the particles, can never lead systematically to a more satisfying result (if all particles have the same weight) than a random distribution of the A particles in the mixture. For the mixing machine cannot distinguish the colour of the particles, which makes a systematic distribution of the A grains in the mixture by means of the mixing procedure impossible. If such a systematic distribution is present at the beginning of the mixing process - e.g. if the A particles are put in the mixing vessel before the B particles - the purpose of the mixing is to destroy this systematic distribution. If the mixing is pursued until the original situation has no influence any more on the result, then there is no place in the mixture where we might expect to find an A particle more often than in any other place. If the particles of component A have a weight differing from those of B, then it would perhaps be possible to construct a mixing machine, which using this difference in weights meets the requirements of definition a or b as well as possible.

With the existing types of mixing processes however (stirring, shaking, turning etc.), a difference in weight can only have a disadvantageous effect on the degree of mixing by causing segregation. It is therefore a reason for contentment if one succeeds in constructing a mixing machine which does not discriminate between particles of different weight, i.e. a mixing machine which is equally efficient for components with different weights as it is for components with the same weight. Hence, for the usual mixing processes, our reasoning also holds for these kinds of mixtures, and we shall stick to the argument given.

There is however one other objection to the use of definition c as it stands. A random process of mixing may also produce very poor results, for instance even a mixture with all A particles at the bottom and all B particles at the top is a not impossible result. In strict accordance with definition c therefore, all mixtures should be called "ideal" or "complete", and this of course is not the intention. This difficulty can be avoided by applying the term "complete" to the mixing process itself instead of to its result (i.e. the mixture). This does not, of course, change the fact that a complete mixing process can still produce bad mixtures. Just as a true die can produce the number six a thousand times in a row, a complete mixing process can produce a result that does not deserve the name "mixture". In both cases however this will rarely occur. Nevertheless the term "complete" seems a little too strong as it suggests the impossibility of a poor result. For this reason we propose to replace this term by the statistical term "random", which describes the situation more realistically, and we shall use the term "random mixing process".

This leads to:

Definition d: A mixing process is called random if all particles are distributed independantly in the mixture in such a way, that for every component of the mixture the probability of finding a particle of this component in a given place is the same for all places in the mixture.

This definition on which BUSLIK [2], but also LACEY [5] and BLUMBERG and MARITZ [1], base their statistical considerations, may serve as the basis for the development of statistical methods.

A random mixing process differs from one not satisfying definition d in this respect, that the latter shows a preference

for some sort of system in the mixture. An example of this is a mixing process which has not been applied long enough. The result will then show resemblance to the initial situation.

Though this mixing method can produce the same mixtures as a random process, mixtures still resembling the initial situation will be more exceptional with the latter process than with the former one.

Furthermore statistical methods are now available for testing whether a mixing process is random against special well defined alternative possibilities, so that problems of practical importance can be investigated. We will return to this subject later on.

Definition d also leads to plausible definitions of the concept "degree of mixing", and it corresponds to a certain degree with definition a: it follows from the definition that samples with many particles will have approximately the same structure as the whole mixture, with only a small probability of large deviations. These questions will be treated in the next sections, which are partly based on the references given at the end of this paper.

3. A definition of "degree of mixing" and the statistical properties of this coefficient.

One of the most frequently occurring problems is the estimation of the degree of mixing of a given mixture on account of a number of samples taken from this mixture. To this end we introduce a coefficient closely connected with the variance of the composition of the samples, as has been done by LACEY [5], BUSLIK [2], a.o.

We suppose that the mixture described in the introduction is composed of N grains of type A and M grains of type B ; k samples are taken containing $n_1, n_2, \dots, n_k$ particles respectively. The samples taken together are supposed to form a small part of the mixture only. Only one condition is imposed with respect to the way of taking the samples: the probability of a particle to belong to a sample should not depend on whether it is an A or a B -particle. On the other hand the place of a particle in the mixture may influence this probability; the samples need not be taken at random.

Under these circumstances the samples, even if not taken at random themselves, have a random composition if the mixing process is a random one.

More precisely, if

$$(1) \quad \gamma = \frac{N}{N+M}, \quad \eta = \frac{M}{N+M} \quad (\gamma + \eta = 1),$$

and if the mixing is random, then an arbitrary particle of an arbitrary sample has probability γ to be of type A, and η to be of type B.

If now $\underline{x}_1, \dots, \underline{x}_k$ ¹⁾ respectively are the fractions of A particles in k samples of $n_1, \dots, n_k$ particles respectively, then we can summarise the results of the sampling in the following $2 \times k$ table:

		Sample number:				
		1	2	.	.	k
(2)	fraction A	$\underline{x}_1$	$\underline{x}_2$	.	.	$\underline{x}_k$
	fraction B	$1 - \underline{x}_1$	$1 - \underline{x}_2$	.	.	$1 - \underline{x}_k$

Or, with the numbers of particles instead of the fractions:

		Sample number:					
		1	2	.	.	k	total
(3)	number of	A	$n_1 \underline{x}_1$	$n_2 \underline{x}_2$	.	$n_k \underline{x}_k$	$\sum n_j \underline{x}_j$
	particles	B	$n_1 (1 - \underline{x}_1)$	$n_2 (1 - \underline{x}_2)$	.	$n_k (1 - \underline{x}_k)$	$\sum n_j (1 - \underline{x}_j)$
	total		n_1	n_2	.	n_k	$n = \sum n_j$

From the theory of the χ^2 distribution it now follows that under the hypothesis of random mixing the statistic

$$(4) \quad \underline{\chi}_1^2 = \sum_{j=1}^k n_j (\underline{x}_j - \gamma)^2 / \gamma \eta = (1/\gamma \eta) (\sum n_j \underline{x}_j^2 - 2\gamma \sum n_j \underline{x}_j + n \gamma^2)$$

has approximately a χ^2 distribution with k degrees of freedom.

The statistic

$$(5) \quad \underline{\chi}_2^2 = \sum_{j=1}^k n_j (\underline{x}_j - \underline{x}^*)^2 / \underline{x}^* (1 - \underline{x}^*) = \sum n_j \underline{x}_j^2 / \underline{x}^* (1 - \underline{x}^*) - n \underline{x}^* / (1 - \underline{x}^*)$$

$$(\underline{x}^* = \sum_{j=1}^k n_j \underline{x}_j / n)$$

1) An underlined symbol indicates a random variable, i.e. a variable possessing a probability distribution; the same symbol, not underlined, is used to denote values assumed by such a random variable.

is, under the same hypothesis, also distributed approximately according to a χ^2 distribution, with $k-1$ degrees of freedom. In this formula $\bar{x}^*$ represents the fraction of A grains in all k samples taken together.

This makes it possible, when the n_i are known and the $\bar{x}_i$ are observed, to test the randomness of the mixing process used.

When γ is known χ_1^2 should be used, otherwise χ_2^2 .

In both cases the large values of χ_a^2 ($a=1,2$) constitute the critical region, i.e. lead to rejection of the hypothesis of randomness. The tests are one-sided (to the right) as is nearly always the case with the χ^2 method. 2)

To introduce a mixing coefficient we try to find a variable which will assume a value approximately equal to 1 if the mixing is random. LACEY [5] proposes as such the coefficient $\sqrt{k/\chi_1^2}$ (for the case that γ is known) and one might analogously take $\sqrt{(k-1)/\chi_2^2}$ for unknown γ . These coefficients are chosen on the following grounds. A mixture is worse according as the value of χ_a^2 ($a=1,2$) is higher; it is therefore natural to put χ_a^2 in the denominator. Furthermore by taking the root we get a linear measure and finally, in the case of random mixing, the mathematical expectation of χ_1^2 equals k and that of χ_2^2 equals $n(k-1)/(n-1)$, which is approximately equal to $k-1$ for large n . It is, however, easy to prove that this coefficient proposed by LACEY has, in both cases, a mathematical expectation larger than 1 if the mixing is random; this is a consequence by the presence of χ_a^2 in the denominator. If one takes the value 1 to indicate randomness of the mixing process, this leads on the average to a flattered view as regards the quality of the mixing processes under investigation. This effect is of some importance for small values of k , but not for large ones.

Its extent can be investigated by using the above mentioned χ^2 approximations for the distribution of χ_1^2 and χ_2^2 . These approximations are quite good even for small values of k if the n_i are sufficiently large and this will usually be the case with samples from mixtures of small particles.

It can be proved, that the mathematical expectation of χ^{-1} with ν degrees of freedom is given by

2) BLUMBERG and MARITZ [1] give the same kind of test for known γ but they complicate things needlessly applying the transformation $x_i = 2 \lg \sin \sqrt{x_i}$.

$$(6) \quad \mathcal{E} \underline{\chi}^{-1} = \Gamma((\nu-1)/2) / \sqrt{2} \cdot \Gamma(\nu/2),$$

where Γ represents the complete Gammafunction. By means of this formula the expectation $\mathcal{E} \sqrt{\nu/\underline{\chi}^2}$ has been calculated and the results are given in Table I. Rather large deviations from 1 occur for $\nu < 10$.

However, substituting $\nu - 3/2$ for ν we find much smaller deviations. (The value $3/2$ has been found experimentally). For $\nu \rightarrow \infty$ the difference between ν and $\nu - 3/2$ vanishes and both expectations have the value 1 as their limit.

Table I

The mathematical expectation of two mixing-coefficients in the case of aselect mixing.

ν	$\mathcal{E} \sqrt{\nu/\underline{\chi}^2}$	$\mathcal{E} \sqrt{(\nu-3/2)/\underline{\chi}^2}$
2	1,77	0,886
3	1,38	0,977
4	1,25	0,991
5	1,19	0,995
6	1,15	0,997
7	1,13	0,998
8	1,11	0,999
9	1,09	0,999
10	1,08	0,999

In connection with this result the coefficients

$$(7) \quad \underline{M}_1 = \sqrt{(\nu-3/2)/\underline{\chi}_1^2} \quad \text{and} \quad \underline{M}_2 = \sqrt{(\nu-5/2)/\underline{\chi}_2^2}$$

respectively might be introduced for χ known and unknown respectively. In both cases the tests described above for the hypothesis of random mixing remain unchanged.

4. Mixtures of more than two components.

If a mixture consists of more than two components, all particles being of the same shape and size, differing for instance only in colour, it is possible to test for each of the components separately whether its particles are distributed at random over the mixture, by treating this component as type A of the foregoing sections and all the others together as one component of type B.

The above mentioned theory can then be applied without any changes. However it is also possible to test for all components simultaneously whether the mixing is random or not.

Denoting the components by $C_1, \dots, C_h$ we get, analogously to (3), a $n \times k$ table.

		Sample number:					total
		1	2	.	.	k	
number of particles	C_1	$n_{11} x_{11}$	$n_{12} x_{12}$	.	.	$n_{1k} x_{1k}$	$\sum_{j=1}^k n_{1j} x_{1j}$
	$\vdots$						
	C_h	$n_{h1} x_{h1}$	$n_{h2} x_{h2}$	.	.	$n_{hk} x_{hk}$	$\sum_{j=1}^k n_{hj} x_{hj}$
total		n_1	n_2	.	.	n_k	n

If the fractions of $C_1, \dots, C_h$ in the mixture are $\xi_1, \dots, \xi_h$ ($\sum \xi_i = 1$) then

$$(8) \quad \chi^2 = \sum_{i=1}^h \sum_{j=1}^k n_{ij} (x_{ij} - \xi_i)^2 / \xi_i$$

has approximately a χ^2 distribution with $k(h-1)$ degrees of freedom and

$$(9) \quad \chi^2 = \sum_{i=1}^h \sum_{j=1}^k n_{ij} (x_{ij} - \bar{x}_i^*)^2 / \bar{x}_i^* \quad (\bar{x}_i^* = n^{-1} \sum_{j=1}^k n_{ij} x_{ij})$$

a χ^2 distribution with $(k-1)(h-1)$ degrees of freedom. Large values of χ_a^2 ($a = 3, 4$) again are critical. The mixing-coefficients

$$(10) \quad \underline{M}_3 = \sqrt{\{k(h-1) - 3/2\} / \chi_3^2} \quad \text{and} \quad \underline{M}_4 = \sqrt{\{(k-1)(h-1) - 3/2\} / \chi_4^2}$$

can be introduced in analogy of the above mentioned mixing coefficients $\underline{M}_1$ and $\underline{M}_2$.

5. Other statistical methods.

The χ^2 coefficients dealt with in the preceding sections measure to a certain extent the character of the mixture, but they do not give a detailed impression of the mixture. If for instance the value of χ^2 , found for a certain set of samples, indicates deviations from randomness it does not give any indications of the source of these deviations; it may be that some of the samples differ strongly from the rest or all of them may be more different from one another than would be expected in the case of randomness. More detail may be found in a simple

graphical method, which will be described in this section. This method, which has been proposed by BUSLIK [2] for samples of equal size, will be given here for samples of different sizes too. For this method ξ is supposed to be known. Let the number of particles in the samples be n_j ($j=1, \dots, k$) and let the mixing be random, then the fractions x_j of A particles for large values of n_j are approximately normally distributed with mean ξ and variance $\xi\eta/n_j$ ($\eta=1-\xi$). The exact distribution is, for every j , a binomial one, which is asymptotically normal for $n_j \rightarrow \infty$. The quantities

$$(11) \quad \tilde{x}_j = (x_j - \xi) / \sqrt{\xi\eta/n_j} \quad (j=1, \dots, k),$$

called the reduced observations, are independently distributed and approximately $N(0,1)$ (i.e. normal with mean 0 and variance 1).

Let $F(\tilde{x})$ be the cumulative distribution function of the $N(0,1)$ distribution and $F_k(\tilde{x})$ the experimental cumulative distribution function of the $\tilde{x}_j$, i.e. let

$$(12) \quad F_k(\tilde{x}) = a/k$$

if there are a values among the

$$\tilde{x}_1, \dots, \tilde{x}_k \text{ which are } \leq x$$

then $F(\tilde{x})$ and $F_k(\tilde{x})$ can be plotted together in a graph. Cf. fig. 1.

This kind of graph shows more detail than just one number can do. In the case outlined in fig. 1, it is clear, for instance, that there were too many samples with too small values of $\tilde{x}$, i.e. too many samples with not enough particles of type A.

This method is closely connected with a test, called the KOLMOGOROV-SMIRNOV test, for the hypothesis of random mixing.

The test statistic of this test is

$$(13) \quad d = \text{Max} |F(\tilde{x}) - F_k(\tilde{x})|,$$

the largest distance in vertical direction between $F_k(\tilde{x})$ and $F(\tilde{x})$. This random variable d has, in the case of random mixing, a known probability distribution, extensively tabulated for large k (by means of an approximation)

by SMIRNOV [8], while for $k \leq 35$ some points of the exact distribution have been tabulated by MASSEY [6].

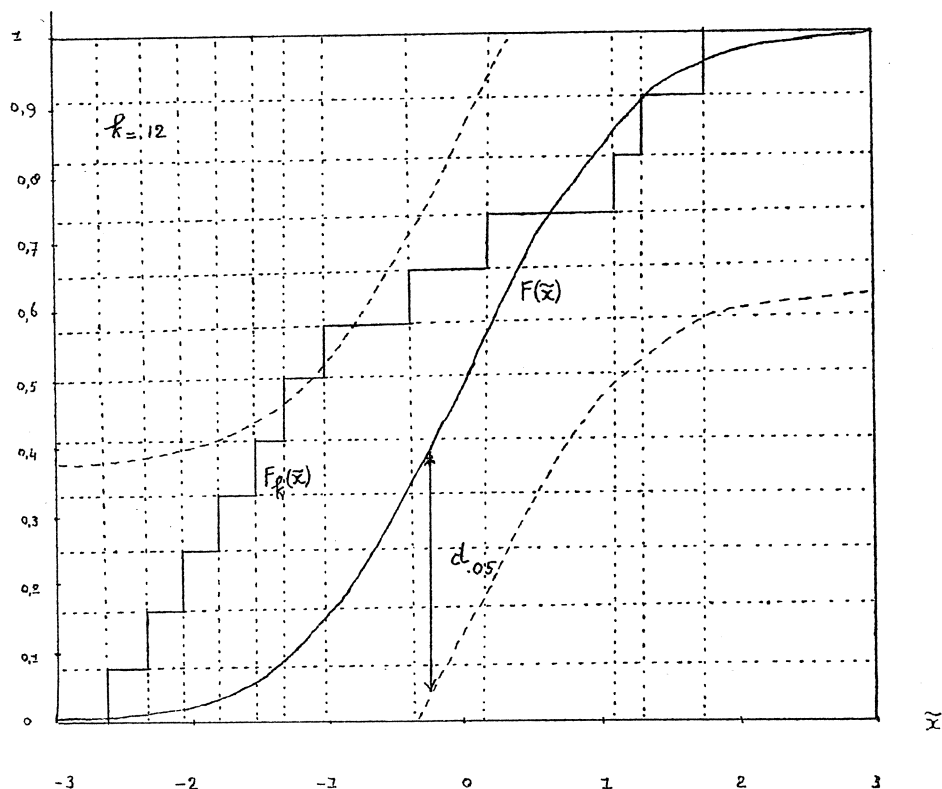


Fig. 1. An experimental and the theoretical cumulative distribution function of $\tilde{x}$ for $k=12$. The points on the $\tilde{x}$ scale represent the reduced observations $\tilde{x}_1, \dots, \tilde{x}_k$.

The latter writer also gives a clear summary of the method and discusses the power of the test. Denoting by d_α the critical value of d for the level of significance α , we can plot two lines on both sides of $F(\tilde{x})$, one situated d_α higher and the other d_α lower than $F(\tilde{x})$. This has been done in Fig.1. for $\alpha=0.05$ and $k=12$. Except for a probability α $F_k(\tilde{x})$ will be situated completely within this region. If this is not the case (as in fig.1.), then one can reject the hypothesis of random mixing, with respect to a level of significance α .

The statistic d can also be transformed into a mixing coefficient. As d itself takes values between 0 and 1, low values indicating a high degree of mixing and high values a low one, $1-d$ might be used as a mixing coefficient. The values found in

practice will then, however, only seldom be close to 0 or 1. Instead of $1-d$ we could also use for instance the tailprobability belonging to the value of d found. Approximate values of this tailprobability are tabulated, for not too small values of k , in the table of SMIRNOV [8]. This tailprobability only assumes values between 0 and 1, small values denoting bad mixing and larger ones good mixing. The mathematical expectation of this tailprobability is equal to $\frac{1}{2}$ in the case of random mixing. By taking twice the value of the tailprobability we get a mathematical expectation equal to 1.

More important than the definition of such a coefficient however is the knowledge of its probability distribution under certain hypotheses and the possibility of testing whether or not two values found for different mixtures indicate a systematic difference between these mixtures. In a later report we will see that the method just discussed and also the χ^2 methods dealt with in the previous sections do indeed procure tests of this kind.

A drawback of the above discussed method of KOLMOGOROV-SMIRNOV is that it can only be applied if γ is known.

6. Samples of equal size.

If all samples are of equal size, i.e. if

$$(14) \quad n_1 = n_2 = \dots = n_k \quad (= m \text{ say}),$$

the formulas (4) and (5) can be simplified to

$$(4') \quad \chi^2 = \gamma^{-1} \gamma^{-1} m \sum_{j=1}^k (\bar{x}_j - \gamma)^2$$

and

$$(5') \quad \chi^2 = \bar{x}^{-1} (1 - \bar{x})^{-1} m \sum_{j=1}^k (\bar{x}_j - \bar{x})^2 \quad (\bar{x} = k^{-1} \sum_{j=1}^k \bar{x}_j).$$

As a rule one will try to comply with (14). If (14) is only approximately fulfilled, one might take the mean of n_j for m and still use (4') and (5'). BLUMBERG and MARITZ [1] have shown that this inaccuracy has no large influence on the result. This can also be understood in a way slightly different from theirs, as follows. Let

$$(15) \quad m = k^{-1} \sum_{j=1}^k n_j \quad \text{and} \quad n'_j = n_j - m.$$

In the case of random mixing, we have:

$$\mathcal{E} (n_j \underline{x}_j - n_j \zeta)^2 = n_j \zeta \eta ,$$

and therefore

$$\mathcal{E} (\underline{x}_j - \zeta)^2 = \zeta \eta / n_j$$

and

$$\mathcal{E} \sum m (\underline{x}_j - \zeta)^2 = m \zeta \eta \sum 1/n_j ,$$

where

$$1/n_j = (m + n_j')^{-1} \approx m^{-1} (1 - m^{-1} n_j' + m^{-2} n_j'^2) .$$

Therefore

$$\mathcal{E} \sum m (\underline{x}_j - \zeta)^2 \approx k \zeta \eta + m^{-2} \zeta \eta \sum n_j'^2 .$$

Denoting

$$(16) \quad \sigma^2 \{n_j\} = k^{-1} \sum n_j'^2 ,$$

it follows, that

$$(17) \quad \mathcal{E} \zeta^{-1} \eta^{-1} m \sum (\underline{x}_j - \zeta)^2 = k \{1 + m^{-2} \sigma^2 \{n_j\}\} .$$

This expected value will exceed that of χ_1^2 only slightly if $\sigma^2 \{n_j\} / m^2$ is small. The same holds for χ_2^2 and the influence on the higher moments may be examined along similar lines.

The test of KOLMOGOROV-SMIRNOV is also more simple for equal values of n_i ; (11) then reduces to

$$(11') \quad \tilde{x}_i = (x_i - \zeta) / \sqrt{\zeta \eta / m}$$

and we can use x_i itself instead of $\tilde{x}_i$, the denominator now being the same for every i . The $\underline{x}_i$ are now approximately normally distributed with mean ζ and variance $\sqrt{\zeta \eta / m}$ and we can now compare $F_k(x)$ with $F(x)$ instead of $F_k(\tilde{x})$ with $F(\tilde{x})$, $F(x)$ representing the cumulative distribution function of the normal distribution with mean ζ and variance $\sqrt{\zeta \eta / m}$.

7. Samples of unknown size.

The supposition, made up to now, that the amounts n_i are known, is, especially for fine grained powders rather unrealistic

Often one will have a vague idea about the order of magnitude of n_i , without being able to determine these values exactly.

If the ratio of the n_i is known more or less exactly, for instance from the ratio of the weights or of the volumes of the samples, and if

$$(18) \quad n_1 : n_2 : \dots : n_k = n'_1 : n'_2 : \dots : n'_k$$

with n'_i known, then we can put

$$(19) \quad n_j = a n'_j \quad (j = 1, \dots, k)$$

and then an upper confidence limit for a , under the hypothesis of random mixing, with a given confidence coefficient $1 - \alpha$, may be determined.

If $\bar{x}$ is known, one can proceed as follows.

Take, from a table of the χ^2 distribution with k degrees of freedom, the critical value χ^2_α corresponding with level of significance α (for instance $\alpha = 0.05$). In the case of random mixing we thus have

$$P[\chi^2 \leq \chi^2_\alpha] = 1 - \alpha.$$

By means of (14) and (19) this expression gives

$$P[\bar{x}^{-1} n^{-1} a \sum_{j=1}^k n'_j (x_j - \bar{x})^2 \leq \chi^2_\alpha] = 1 - \alpha.$$

If the observed values of $x_1, \dots, x_k$ are $x_1, \dots, x_k$ it follows that

$$(20) \quad a \leq \bar{x} \chi^2_\alpha / \sum_{j=1}^k n'_j (x_j - \bar{x})^2 \quad (= a_1 \text{ say})$$

with level of significance α (or equivalently: with confidence coefficient $1 - \alpha$).

If $\bar{x}$ is unknown, we use χ^2_2 instead of χ^2_1 . We then find the χ^2_α , corresponding to the level of significance α , of a χ^2 distribution with $k-1$ degrees of freedom and the final result is:

$$(21) \quad a \leq x^*(1-x^*) \chi^2_\alpha / \sum_{j=1}^k n'_j (x_j - x^*)^2 \quad (= a_2 \text{ say}).$$

Though the n_j are not known exactly, one can nevertheless judge in many cases whether the upper limit $a_1 n_j$ (or $a_2 n_j$ respectively), which have been found, is reasonable or not. If these

upper limits are too small, the hypothesis of random mixing must be rejected.

If the ratios n_j' of the numbers of particles in the samples are also unknown, but all samples have about the same size, the formulas of section 6 can be used and an upper limit for the mean m may be determined. The formulas (20) and (21) then become

$$(20') \quad m \equiv \zeta \chi^2_{\alpha} / \sum_{j=1}^k (x_j - \zeta)^2 \quad (= m_1 \text{ say})$$

and

$$(21') \quad m \equiv x^*(1-x^*) \chi^2_{\alpha} / \sum_{j=1}^k (x_j - x^*)^2 \quad (= m_2 \text{ say}).$$

Upper limits like these may, if ζ is known, also be determined graphically with the help of the method of KOLMOGOROV-SMIRNOV as described in section 5. To that purpose $n_j = a n_j'$ (or $n_j = m$ respectively) is substituted in (11) or (11') respectively, and the graph of $F_k(\bar{x})$ is drawn for some value a_0 (or m_0 respectively) of a (or m respectively). Multiplying a_0 (or m_0 respectively) by a factor C amounts to multiplying the values $\bar{x}_j$ by $\sqrt{C}$. The largest value of C for which $F_k(\bar{x})$ is still completely above the lower bound for $F(\bar{x})$, corresponding to a level of significance α , can then easily be determined. The upper limit for a (or m respectively) is then ca_0 (or cm_0 respectively).

Lower confidence limits could be determined analogously, but these are of less interest and will not be described here.

8. The size and the place of the samples.

The sizes (n_j) of the samples used have a large influence on the results. Some investigators (for instance LACEY [5]) consider this a serious disadvantage for the methods described, but from a practical viewpoint this should be considered inaccurate. In practice usually samples up to a certain size are used from a certain mixture and these samples should have a reasonably constant composition, while for still smaller samples the composition does not matter very much.

A remark of this kind may be found in DANCKWERTS [3]. He emphasises the impossibility to examine whether a mixture should be labelled "good" or "bad" before the requirements necessary for a "good" mixture are determined. One of the items to be determined is the size of the samples, which is considered of prac-

tical importance. A second item in this connection are the places where the samples are taken

The theory described leaves this point still open. The choice of these places depends on the alternative hypotheses against which one wishes to test the hypothesis of random mixing. If one wishes to examine the homogeneity of the whole mixture, it is best to take the samples systematically from all parts of the mixture. If on the other hand one has the idea that the mixing is not random in certain places in the mixture, one will take the samples exactly in those places. For still more specified alternative hypotheses, for instance a trend of the amount of the component A in a special direction in the mixture, or if special parts of the mixture may contain less of component A than other parts, it is better to use methods especially designed to detect defects of a specified kind; such methods will be described later.

As an example of a special alternative, for which the methods described here may be used, we will deal with a strongly simplified model of clotting. We suppose that the particles of component A adhere to each other but not to those of component B, while the particles of component B also adhere to each other. Furthermore we suppose that the lumps of A and B all have about the same size and contain an average of ν grains (ν unknown). If the mixing of these lumps is random, such that only the clotting will lead to a deviation from complete randomness, then the methods described in section 7 will give an upper limit for the number of lumps in the samples; for instead of the number of particles the number of lumps will now appear in the formulas. If the number of particles in a sample is known, it is possible to compute at once from this upper limit for the number of lumps per sample a lower limit for ν , the amount of particles per lump. In the case of samples of equal size for instance, formula (20') determines the upper limit m , for the number of lumps, while m_1 , the number of particles per sample, is known.

We then have, with a level of significance α ,

$$(22) \quad \nu \geq m/m_1.$$

A similar method has been used by BUSLIK [2] to give an estimate of ν . He finds (23) $\nu = \chi^2/k$ or χ^2/k_1 , respectively. This estimate follows from the preceding theory in the same way as (22).

Literatuurlijst

- [1] R. Blumberg and J.S. Maritz, Mixing of solid particles, Chem.Engin. Sc. 2 (1953) 240-246.
- [2] D. Buslik, Mixing and sampling with special reference to multi-sized granular material, A.S.T. M. Bulletin no. 165 (1950) (TP 92) 66-73.
- [3] P.V. Danckwerts, Theory of mixtures and mixing, Research 6 (1953) 355-361.
- [4] G. Herdan, Small particle statistics, Elsevier, Amsterdam-Houston-N.Y.-Paris(1953).
- [5] P.M.C. Lacey, The mixing of solid particles, Transactions of the Institution of Chemical Engineering, 21 (1943) 53-59.
- [6] F.J. Massey Jr., The Kolmogorov-Smirnov test for goodness of fit, Jrn.Am.Stat.Ass. 46 (1951) 68-78.
- [7] C. Naske, Mischen Fester Stoffe, Der Chemie-Ingenieur, I, 137.
- [8] N.Smirnov, Table for estimating the goodness of fit of empirical distributions, Ann. Math. Stat. 19 (1948) 279-281.
- [9] K.S. Valentine and G. Mac Lean, Mixing, Perry's Chemical Engineers Handbook 3th ed. (1950) section 17.
-