KONINKLIJK NEDERLANDS
METEOROLOGISCH INSTITUUT

Wetenschappelijk Rapport
W.R. 62-2 (III-270)
U.D.C.: 551.508.99:
551.510.72:
519.2

Dr. C. Levert

Statistical aspects concerning the measurement of the specific artificial atmospheric radioactivity.

.-.-.-.-.

CONTENTS

ADDENDUM B

Statistical aspects concerning the estimates of the "age" and "the zero day"-value of the pure filteractivity on the ground of the measurements of the filteractivity on two or more days after the sampling day.

.-.-.-.-.

1. The principle of the measurement

1.1 Some general remarks

The specific artificial atmospheric radioactivity can be measured according to the following principle.

a. A sample Vm^3 of atmospheric air is sucked through a filter (e.g. $V = 40$, 100, 1000 $m^3/24$ h).

b. Consequently this filter bears radioactive particles of dust from the air (natural and artificial radioactivity). One waits a few days (T_1 days; e.g. 4) before counting the filteractivity (corrected for background) for the first time. The true value may be β_1 imp/min. In this short period the natural atmospheric radioactivity has decayed (almost) completely, but in the same time also the artificial radioactivity has decreased. Since it is the purpose to find the zero day's value of the artificial radioactivity, say β_o, (zero day = day of sampling), it is necessary to make at least one additional counting of the filteractivity (true value β_2), say on day T_2, if it is assumed that the decrease of the artificial radioactivity follows an exponential law with constant and numerically known exponent . The values β_1 and β_2 enable us to compute β_o. Since V is also known, the zero day's value of the specific filteractivity $\alpha = \beta_o : V$, the true mean number of impulses per minute and per m^3 of air, can be calculated.

c. Next the true specific artificial/air radioactivity, say R, can be found by means of the relation:

$$R = C.\alpha \qquad pC/m^3$$

Here C, a proportionality constant between R and α, is a function of a number of quantities, characterised by the measuring apparatus and the character of the radioactive substance. Even if all these characteristics were known (in reality they are not), still their numerical values cannot be known exactly. Hence generally a value $\hat{C}$ is used which may differ more or less from the true unknown value C. For two instruments the values of C may differ, C_1 and C_2. Information on this difference (without knowledge of C_1 and C_2 separately) may be obtained by measuring with these instruments the same true, but unknown, value of the artificial atmospheric radioactivity in such a way that the samples of air are "identical" 1). Such samples must refer to the same place and the same time. If they could be drawn, then the values α_1 and α_2 could be calculated and next $\hat{R}_1 = \hat{C}_1.\alpha_1$ and $\hat{R}_2 = \hat{C}_2.\alpha_2$. Generally R_1 and R_2 will differ. Still there exists 2) only one well defined specific artificial atmospheric radioactivity , sothat one would expect $R_1 = R_2$. Consequently, at least one of the values $\hat{C}_1$ and $\hat{C}_2$ cannot be correct. In spite of the fact that the true values C_1 and C_2 are not known, the ratio $q = R_1/R_2$ may be used as a reduction factor, the interpretation of which would be as follows: as soon as these instruments are installed at different places and simultaneous measurements would give the values W_1 and W_2, then the values W_1 and $W'_2 = q.W_2$ are "better comparable" then W_1 and W_2.

1.2 Complicating factors

The above considerations are too simple. There are some further complications.

Complication 1.

The filteractivity is counted during a finite time (in the Netherlands about 15 minutes) and hence the measured mean numbers of counts per minute, written $\hat{\beta}_1$ on day T_1 and $\hat{\beta}_2$ on day T_2, are more or less accurate estimates of the true mean values, β_1 resp. β_2. The value obtained for $\hat{\beta}_1$ in a given quarter of an hour, is more or less a "random value". If it would be possible to measure β_1 once more, then certainly a quite different value for $\hat{\beta}_1$ would be obtained. This fact follows from the statistical nature of the production of the impulses, which takes place in time according to the poissonian law. Consequently, both β_1 and β_2 are measured inaccurately ($\underline{\hat{\beta}}_1$ and $\underline{\hat{\beta}}_2$)3). Therefore, also β_0 is known inaccurately ($\underline{\hat{\beta}}_0$), usually even far more inaccurate than $\underline{\hat{\beta}}_1$ and $\underline{\hat{\beta}}_2$ separately (For more details see Addendum A). This means that even

notes 1) A statistical definition of "identical" should be given.

2) Of course one should define the meaning of the "true specific artificial atmospheric radioactivity" at a given place and at a given time.

3) Underlining ($\underline{\hat{\beta}}_1$ instead of $\hat{\beta}_1$) means that $\underline{\hat{\beta}}_1$ is stochastically distributed. Would one be able to repeat the measurement of the same true value β_1 an infinite number of times, then a population of $\underline{\hat{\beta}}$-values would be obtained, varying around β_1 with a certain standarddeviation $\sigma(\underline{\hat{\beta}}_1)$.

if C_1 and C_2 would be known exactly, then R_1 and R_2 could be unequal, since usually $\hat{\alpha}_1 \neq \hat{\alpha}_2$, in spite of the fact that $\alpha_1 = \alpha_2$ for two adjacent apparates[1] (equal zero day's values).

Complication 2.

It is impossible to draw two samples on exactly the same place and in the same time. Of course the distance between both filters should be made as small as permissible. Even then, the samples of air cannot be completely "identical" (which assumption should be fulfilled in the present reasoning), since both samples do not contain exactly equal total numbers of particles of dust; this is especially the case for those particles which carry large quantities of radioactivity ("hot spots"). The filters, after having sucked through the "same" air, may bear 5 and 5 hot spots as well as 2 and 8 or 0 and 10. What happens is more or less a question of chance. The one filter may possess far more radioactivity than the other, notwithstanding the fact that both instruments sampled the "same" air. In this way the so called "spatial random error" of measurement is introduced.

Nevertheless, when the instruments are brought together at the same place with their filters close to each other, it is certain that:

a. systematical differences between the true specific atmospheric radioactivity of the samples of air are absent or neglectably small;

b. random differences between the specific air radioactivity of the samples, as mentioned above, are as small as possible.

Complication 3.

The constant C may be written as:

$$C = \frac{1}{a}\,\frac{D}{D'}\,\frac{1}{g}\,\frac{10}{60 \times 3.7}\,\frac{1}{e}$$

a = fraction of the radiation absorbed in the window of the Geiger Müller counter

D = true known activity of the radioactive standard (e.g. U_3O_8)

D'= measured value of D

g = geometrical factor, relating to the position of the filter with regard to the counter tube

---

note 1) For the sake of simplicity V_1 and V_2 (the samples of air) are supposed to be known "exactly". In practice the maximal error in the measurement of V is about 1%.

e = efficiency of the suction; e.g. e = 0.70, which means that 70% of all dust particles are cought by the filter.

In the Bilt C = 1.5, in Belgium C = 2.0, but these values 1.5 or 2.0 are only constant as long as the numerical values of a, D, D', g, e are constant (for this reason C is called a semi constant). Since these numerical values are not known exactly, only an estimate $\hat{C}$ of C can be calculated. Moreover C will depend on the nature of the radioactivity itself; this relation may not be known.

Complication 4.

Some remarks have already been made on the so called "comparability" of two or more measurements. Apart from the fact that a fully satisfactory definition of "comparability" is still missing, it is felt that the concept of comparability is not only related to the systematical errors (questions of reduction), but also to the random errors. Suppose for instance, that the systematical error would be known and hence, when measuring the values W_1 and W_2 with two instruments at different places, one should correct W_2 to $W_2' = qW_2$, where q is the factor mentioned above. Then W_2' and W_1 are better "comparable" than W_2 and W_1. However, also the random errors should be taken into consideration. If they follow distributions with standard deviations σ_1 and σ_2, the true values R_1 and R_2, of which W_1 and W_2 are point estimates, are situated ("with 0.95 probability") within the confidence regions $W_1-2\sigma$ to $W_1+2\sigma$, and $W_2'-2\sigma_2$ to $W_2'+2\sigma_2$. The σ's may be so large and hence these regions so broad, that one may even doubt whether a correction for systematical errors makes any sense. We will treat this aspect in detail later on.

1.3 Formulation of the purpose of the study

In the following sections we will consider in detail, theoretically and numerically, all statistical aspects concerning the computation of systematical and random errors.

When sometimes these considerations give the impression of being highly academical and of only little practical importance, one should always bear in mind, that only an analysis of the whole procedure of measuring and computing the artificial radioactivity and a critical study of each step in this procedure, in particular with regard to the underlying assumptions, will give an idea of the intrinsic value of the ultimately published figures (pC/m^3) which are used by the various users (public health services; meteorologists, etc.). This knowledge may assist in drawing conclusions in a statistically more justified way. It will in any case be advisable to be fully aware of any special assumption notwithstanding the difficulty of verifying its validity e.g. as a result of scarcity of data.

2. Mathematical attack; symbols and definitions

Let k instruments (of which it is not known whether they are "identical" or not)be installed close to each other at one at the same place. The wording "close to each other" refers to the sucking, sampling, units of the instruments. Let on each of n days ($i = 1, 2, 3 \ldots, n$) simultaneous measurements be made with these apparates ($j = 1, 2 \ldots, k$).

Of course the final results $z_1, z_2 \ldots z_k$ (expressed in pC/m^3) will differ on each of these days. We will examine the origin of these differences.

The following symbols and definitions are introduced.

C_j = true unknown reduction or proportionality constant for apparatus j. C_j possesses the same value for each day, but it is possible, that $C_1 \neq C_2 \ldots \neq C_k$.

$\hat{C}_j$ = used value, i.e. the value substituted for C_j; this value may be different for the different instruments, but it is supposed to be constant during the n days.

$\hat{C}_j - C_j$ is mentioned "systematical error" (s.e.) of C_j.

$(\hat{C}_j - C_j) : C_j$ is the percentual s.e. of C_j.

α_{ij} = true mean number of impulses per min. and per m^3 air for the artificial radioactivity of the filter of apparatus j on the sampling day i.

$\hat{\alpha}_{ij}$ = estimate of α_{ij}.

$\hat{\alpha}_{ij} - \alpha_{ij}$ = random counting error.

$(\hat{\alpha}_{ij} - \alpha_{ij}) : \alpha_{ij}$ = percentual counting error.

u_i = true unknown value of the specific artificial air radioactivity on day i at the place where all apparates are measuring. All instruments $j = 1, 2 \ldots k$ are measuring the same value u_i, which usually will vary day for day.

$\underline{\hat{u}}_{ij} = C_j \alpha_{ij}$ = true value of the specific artificial air radioactivity in the sample drawn by instrument j on day i. This $\underline{\hat{u}}_{ij}$ is an estimate of u_i and differs from u_i at random.

$(\underline{\hat{u}}_{ij} - u_i)$ = the random spatial sampling error.

$(\underline{\hat{u}}_{ij} - u_i) : u_i$ = the percentual value of the random spatial error.

$\underline{z}_{ij} = \hat{C}_j \underline{\hat{\alpha}}_{ij}$ = specific artificial air radioactivity as measured by instrument j on day i.

$A_j \equiv \lg(\hat{C}_j : C_j) \cong (\hat{C}_j - C_j) : C_j$, if the absolute value of the right hand side is sufficiently small[1].

$\underline{B}_{ij} \equiv \lg(\underline{\hat{\alpha}}_{ij} : \alpha_{ij}) \cong (\underline{\hat{\alpha}}_{ij} - \alpha_{ij}) : \alpha_{ij}$, if the absolute value of the right hand side is sufficiently small.

$\underline{f}_{ij} \equiv C_j \alpha_{ij} - u_i \equiv \underline{\hat{u}}_{ij} - u_i$.

$\underline{e}_{ij} \equiv \lg(\underline{\hat{u}}_{ij} : u_i) \cong (\underline{\hat{u}}_{ij} - u_i) : u_i \equiv \underline{f}_{ij} : u_i$, if $|\underline{f}_{ij}/u_i| \ll 1$.

The random nature implies

$\mathcal{E}\underline{B}_{ij} = 0$, $\mathcal{E}\underline{e}_{ij} = 0$ for each i and each j.

We state: if it would be possible to remeasure with the same instrument on the same day i the same true value u_i and if C_j and α_{ij} would be known exactly, then $\underline{\hat{u}}_{ij} = C_i \alpha_{ij}$ would vary around u_i with a certain standard deviation. Now α_{ij} is not known, an estimate $\underline{\hat{\alpha}}_{ij}$ is found, and hence $C_j \underline{\hat{\alpha}}_{ij}$ will be distributed around $\underline{\hat{u}}_{ij} = C_j \alpha_{ij}$ with certain standard deviation. Now C_j is not known, but an (constant) estimate $\hat{C}_j$ is used; hence, finally the value $\underline{z}_{ij} = \hat{C}_j . \underline{\hat{\alpha}}_{ij}$ will vary around u_i with certain standarddeviation.

(1)
$$\underline{z}_{ij} \equiv \hat{C}_j \underline{\hat{\alpha}}_{ij} = C_j \alpha_{ij} \left(\frac{\hat{C}_{ij}}{C_j}\right)\left(\frac{\underline{\hat{\alpha}}_{ij}}{\alpha_{ij}}\right) = \underline{\hat{u}}_{ij} \left(\frac{\hat{C}_j}{C_j}\right)\left(\frac{\underline{\hat{\alpha}}_{ij}}{\alpha_{ij}}\right)$$
$$= u_i \left(\frac{\underline{\hat{u}}_{ij}}{u_i}\right)\left(\frac{\hat{C}_j}{C_j}\right)\left(\frac{\underline{\hat{\alpha}}_{ij}}{\alpha_{ij}}\right) = u_i \exp.(\underline{e}_{ij} + A_j + \underline{B}_{ij}) = u_i \exp.(\underline{E}_{ij} + A_j)$$

where $\underline{E}_{ij} \equiv \underline{e}_{ij} + \underline{B}_{ij}$.

Note[1]) Here the theorem is used: $\lg(1+m) \cong m$ for small m.
The approximation is as follows: $\left|\{\lg(1+m) - m\} : m\right| < 0.05$ for $|m| < 0.10$.
Take for instance $m = (\hat{C}_j - C_j) : C_j$.

Take the natural logarithm and we obtain:

(2)
$$\underline{y}_{ij} = x_i + \underline{E}_{ij} + A_j$$

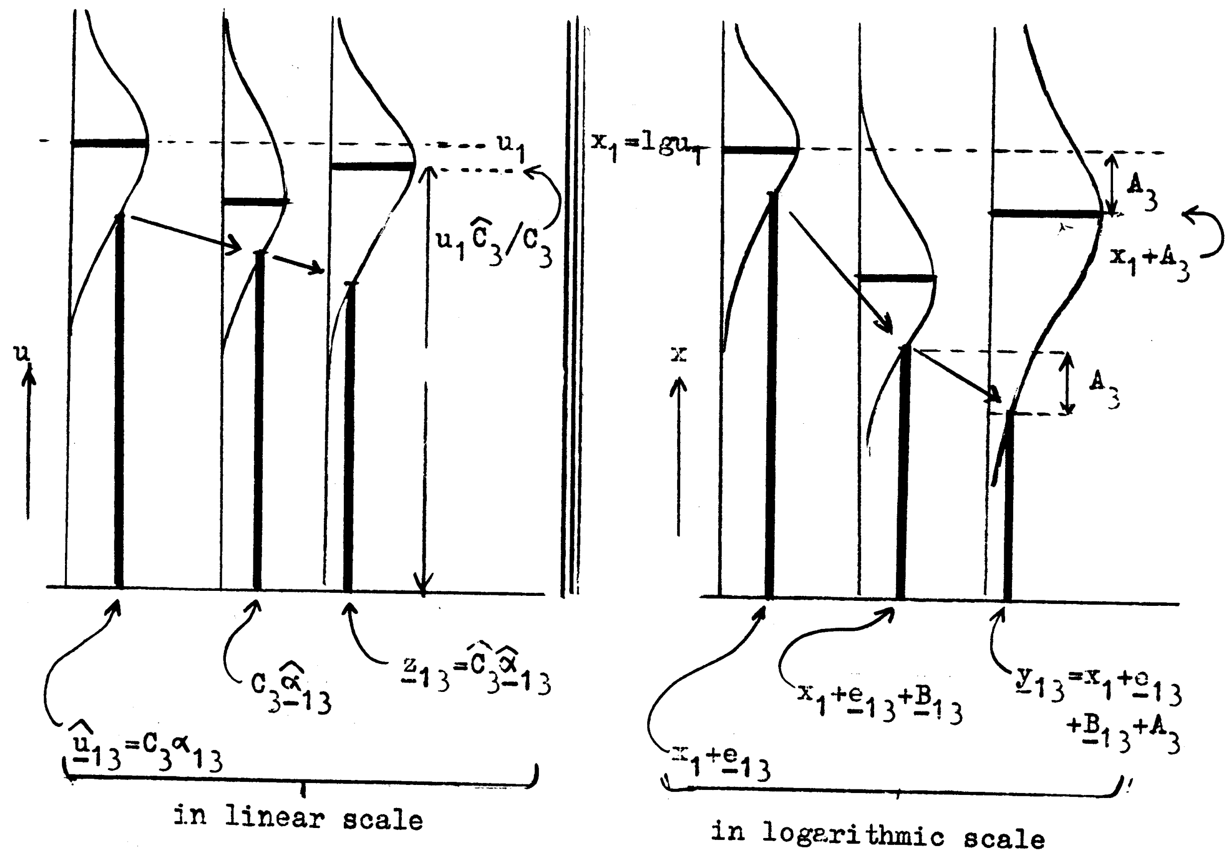
where $x = \lg u$; $\underline{y} = \lg \underline{z}$

$\underline{E}$ and A are dimensionless.

Here, as said already above, $\underline{e}$ represents the spatial random error, $\underline{B}$ represents the counting random error and A represents the constant systematical error.

Hence, since $\underline{z}_{ij}$ is distributed around $u_i \frac{\hat{C}_j}{C_j}$ in a certain way as explained above (suppose we would be able to measure the same true value u_i an infinite number of times with the same apparatus j), also $\underline{y}_{ij}$ follows a random distribution, the mean value of which is approximately lg (mean value of $\underline{z}_{ij}$) and its standarddeviation $\sigma(\underline{E}_{ij}) \equiv \{\sigma^2(\underline{e}_{ij}) + \sigma^2(\underline{B}_{ij})\}^{\frac{1}{2}}$. Since the mean value of $\underline{z}_{ij}$, written $\mathcal{E}\underline{z}_{ij}$, is $u_i \hat{C}_j : C_j$, the mean value of $\underline{y}_{ij}$, written $\mathcal{E}\underline{y}_{ij}$, is nearly $x_i + A_j$. In this way we have passed from (1) with a multiplicative character to (2) with an additive character, better suitable for the application of statistical methods[1].

The following sketch illustrates the mathematical attack given above for day i = 1 and apparatus j = 3.



Note [1] - See following page.

N.B. One can also proceed in an additive way, as follows.

Let be $\hat{\underline{u}}_{13} = u_1 + \underline{\Delta u}_{13}$, with $\mathcal{E}\underline{\Delta u}_{13} = 0$;

$\hat{\underline{\alpha}}_{13} = \alpha_{13} + \underline{\Delta\alpha}_{13}$, with $\mathcal{E}\underline{\Delta\alpha}_{13} = 0$

and $\hat{C}_3 = C_3 + \Delta C_3$

and $\hat{u}_{13} = C_3\alpha_{13}$ on day i = 1 with apparatus j = 3.

Then
$$\underline{z}_{13} = \hat{C}_3 \cdot \hat{\underline{\alpha}}_{13} = (C_3 + \Delta C_3)(\alpha_{13} + \underline{\Delta\alpha}_{13}) = C_3\alpha_{13} + C_3\underline{\Delta\alpha}_{13} + \alpha_{13}\Delta C_3 + (\Delta C_3)(\underline{\Delta\alpha}_{13}) = (u_1 + \underline{\Delta u}_{13}) + C_3\underline{\Delta\alpha}_{13} + C_3\alpha_{13}\frac{\Delta C_3}{C_3} + (\Delta C_3)(\underline{\Delta\alpha}_{13}) = (u_1 + \underline{\Delta u}_{13})(1 + \frac{\Delta C_3}{C_3}) + C_3\alpha_3\frac{\underline{\Delta\alpha}_{13}}{\alpha_3} + (\Delta C_3)(\underline{\Delta\alpha}_{13}) = (u_1 + \underline{\Delta u}_{13})(\frac{\hat{C}_3}{C_3} + \frac{\underline{\Delta\alpha}_{13}}{\alpha_3}) + (\Delta C_3)(\underline{\Delta\alpha}_{13}).$$

Now taking the mean value $\mathcal{E}$ (supposing one would be able to measure the same u_1 many times with instrument 3) one again obtains

$$u_1 \cdot \frac{\hat{C}_3}{C_3}.$$

Attention: $\underline{\Delta u}_{13}$ and $\underline{\Delta\alpha}_{13}$ are supposed to be distributed independently.

3. Systematical errors

3.1 Simultaneous measurements on only one day

3.1.1 Estimate

Consider only one specified day, sothat we may drop the index i in (1) and (2). The k instruments, installed close to each other at one and the same place, give the values $z_1, z_2, \ldots z_k$ pC/m^3 and the

---

Note[1]) In this connection the following theorem is brought to the attention of the reader. When a statistical variable $\underline{v}$ follows a gaussian distribution with a mean value ξ and a standarddeviation σ_v (say $\delta = \sigma/\xi$), then $\underline{w} = \lg \underline{v}$ follows a non normal distribution with a mean value $\eta \cong \lg\xi - \frac{1}{2}\delta^2$ and a standarddeviation $\sigma_w = \delta$. The approximation for η is the better, the smaller δ. In the complete expression $\eta = \lg\xi - \frac{1}{2}\delta^2 - T$, the positive term T is smaller than $\frac{3}{4}\delta^2 : (1-\delta^2)$. Reversely, when a sample of w values gives the mean value $\bar{w}$ and the standarddeviation s, then $\bar{w}$ and s are estimates of η and σ_w. Then ξ and σ_v can be estimated by $\hat{\xi} = \exp.(\bar{w} + \frac{1}{2}s^2) \cong \exp.\bar{w}$, if s is sufficiently small, and $\hat{\sigma} = s\hat{\xi}$.

k values

(3) $y_j = x + E_j + A_j$; $j = 1, 2, \ldots k$.

These k expressions contain the same unknown value x. They will differ because usually the values of E_j and A_j will differ. If we want to find $A_1, A_2, \ldots A_k$, we should bear in mind that there are k equations with k + 1 unknowns x, $E_1+A_1, \ldots E_k+A_k$.

Each constant A_j is combined with a random value E_j. The number of unknowns can be decreased with one in two ways:

i) either we consider the k-1 differences A_j-A_1, $j = 2, 3, \ldots k$, if instrument No.2 would be chosen as reference ("reference method"). Then each constant A_j-A_1 is combined with a random value E_j-E_1;

ii) or we consider the k-1 differences $A_j-\bar{A}$, where $\bar{A} \equiv \frac{1}{k}\sum_1^k A_j$. Then each constant $A_j-\bar{A}$ is combined with a random value $E_j-\bar{E}$. Now $\bar{A}$ refers to an "imaginary" instrument, say the "mean" apparatus ("mean instrument method"). The larger k, the less $\bar{A}$ will differ from zero, provided that the k instruments are not completely identical. Said with other words: the larger the number of different instruments, the better will the overal mean measured value approach to the true unknown value.

More details as to these methods i) and ii) will be discussed in section 3.3.

Now define: $\bar{E} = \frac{1}{k}\sum_1^k E_j$; $\bar{e} = \frac{1}{k}\sum_1^k e_j$; $\bar{B} = \frac{1}{k}\sum_1^k B_j$;

$$\bar{y} = \frac{1}{k}\sum_1^k y_j \; ; \quad \Delta E_j = E_j - \bar{E} \; ; \quad \Delta A_j = A_j - \bar{A} \, .$$

Next compute the differences $\Delta y_j = y_j - \bar{y}$, and we see that:

(4) $\Delta y_j = \Delta E_j + \Delta A_j$. $j = 1, 2, \ldots k$.

Hence Δy_j "almost equals" ΔA_j. The difference is of random nature. Therefore Δy_j may be considered as an estimate, say $\widehat{\Delta A}_j$, of ΔA_j. Write:

(5) $\widehat{\Delta A}_j = \Delta y_j$. $j = 1, 2, \ldots k$.

The estimate is unbiassed for each j. The statistical explanation is as follows: if one would be able to measure the same true unknown value u on a given day with each of all instruments again and again, then a population of values e_j, B_j, E_j $(j = 1, 2, \ldots k)$ would result and hence a population of values of y_j, $\widehat{\Delta A}_j$ for instrument N^{o}j would result. For this reason it would be better to underline the statistical variables and to write:

(6) $$\underline{y}_j = x + \underline{E}_j + A_j \qquad \text{and} \qquad \underline{E}_j = \underline{e}_j + \underline{B}_j .$$

Now, as said already, $\mathcal{E}\underline{B}_j = \mathcal{E}\underline{e}_j = 0$, and hence $\mathcal{E}\underline{E}_j = 0$, for each j. The population of $\underline{\widehat{\Delta A}}_j$ values (j given) possesses a mean value, written as $\mathcal{E}\underline{\widehat{\Delta A}}$. If this mean value equals exactly the unknown value ΔA_j which we desire to find, then it is "unbiassed". Now $\mathcal{E}\underline{\widehat{\Delta A}}_j = \mathcal{E}(\Delta\underline{y}_j) = \mathcal{E}(\Delta\underline{E}_j + A_j) = \mathcal{E}\Delta\underline{E}_j = \mathcal{E}(\underline{E}_j - \underline{\bar{E}}) = 0$ because $\mathcal{E}\underline{E}_j = 0$ for each j.

3.1.2 Accuracy

The hypothetical population of $\underline{\widehat{\Delta A}}_j$ values mentioned above also possesses a standarddeviation, say $\sigma(\underline{\widehat{\Delta A}}_j)$, which may be considered as the accuracy of the estimate $\underline{\widehat{\Delta A}}_j$. The differences $(\underline{\widehat{\Delta A}}_j - A_j)$ may be considered as an estimation error.

Referring to (4) we have:

$$\sigma^2(\Delta\underline{y}_j) = \sigma^2(\Delta\underline{E}_j + \Delta A_j) = \sigma^2(\underline{E}_1 - \underline{\bar{E}}) = \sigma^2\left\{(1 - \tfrac{1}{k})\underline{E}_j + \tfrac{1}{k}\sum_{\substack{1 \\ m\neq j}}^{k} \underline{E}_m\right\}$$

or:

(7) $$\boxed{\sigma^2(\underline{\widehat{\Delta A}}_j) = \left(1 - \tfrac{1}{k}\right)^2 \sigma^2(\underline{E}_j) + \tfrac{1}{k}\sum_{\substack{1 \\ m\neq j}}^{k} \sigma^2(\underline{E}_m).} \qquad j = 1, 2, \ldots k.$$

Here $\sigma(\underline{E}_j)$, shorter σ_j, denotes the standarddeviation of the population of $\underline{E}_j$ values, described in 3.1.1.

The underlying assumption is that $\underline{E}_p$ is not correlated with $\underline{E}_q$; $p \neq q$; $p,q = 1, 2, \ldots k$. This implies: the chance to obtain a value $E_p > 0$ is independent of the fact that $E_q > 0$ or $E_q < 0$.
If $\sigma_1 = \sigma_2 = \ldots \sigma_k$, say , then:

(8) $\sigma^2(\widehat{\underline{\Delta A}}_j) = \frac{k-1}{k}\sigma^2 = \sigma^2$ for each j, and for sufficiently large k.

$k = 2 \quad \sigma(\widehat{\underline{\Delta A}}) = 0.71\,\sigma$

$= 10 \qquad\quad = 0.95\,\sigma$ etc.

The second assumption is that $\underline{e}$ and $\underline{B}$ are not correlated. Of course this hypothesis is correct, because the spatial sampling error has nothing to do with the counting error.

Then:

(9) $$\sigma_j^2 = \sigma^2(\underline{E}_j) = \sigma^2(\underline{e}_j) + \sigma^2(\underline{B}_j) \qquad j = 1, 2, \ldots k.$$

This means: if one wishes to study the differences between the spatial random errors of two or more instruments by means of the $\sigma(\underline{E})$ values, their $\sigma(\underline{B})$ values should be so small, that $\sigma(\underline{E}) \cong \sigma(e)$, or that $\sigma(B)/\sigma(e) \ll 1$. If in practice this cannot be achieved, it will be necessary to compute $\sigma(\underline{B})$ in each case separately. For details regarding this computation, see Addendum A.

3.2 Simultaneous measurements on several days

3.2.1 Estimate

Of course it is better to make simultaneous measurements on more than one day. Let k simultaneous measurements (with k instruments; $j = 1, 2, \ldots k$) be made at one and the same place on each of n days, $i = 1, 2, \ldots n$. The groups of measurements will be $z_{i1}, z_{i2}, \ldots z_{ik}$, $i = 1, 2, \ldots n$, giving

$y_{ij}, y_{i2}, \ldots y_{ik}$, with $y = \lg z$

where

(10) $$\underline{y}_{ij} = \underline{x}_i + \underline{E}_{ij} + A_j, \qquad \text{with } \underline{E}_{ij} = \underline{e}_{ij} + \underline{B}_{ij}$$

$i = 1, 2, \ldots n$; $\qquad j = 1, 2, \ldots k.$

The variable x has been underlined ($\underline{x}$), since it will vary from day to day.

Further $\underline{E}$, $\underline{e}$, $\underline{B}$ will depend both on the day and the instrument; however A depends only on the instrument.

As an illustration, let us take $k = 3$, $n = 8$. The table of z values will be:

i \ j	j=1	2	3
1	z_{11}	z_{12}	z_{13}
2	z_{21}	z_{22}	z_{23}
etc. 8	z_{81}	z_{82}	z_{83}

The y-table becomes

i \ j	1	2	3	mean
1	y_{11}	y_{12}	y_{13}	$\bar{y}_1$ or $y_{1.}$
2	y_{21}	y_{22}	y_{23}	$\bar{y}_2$ or $y_{2.}$
etc. 8	y_{81}	y_{82}	y_{83}	$\bar{y}_8$ or $y_{8.}$
mean	$y_{.1}$	$y_{.2}$	$y_{.3}$	$y_{..}$

By computing the differences $\Delta y_{11} = y_{11} - y_{1.}$; $\Delta y_{12} = y_{12} - y_{1.}$ etc. we obtain the following table

	Δy_{11}	Δy_{12}	Δy_{13}	mean 0
	Δy_{21}	Δy_{22}	Δy_{23}	" 0
	⋮			
	Δy_{81}	Δy_{82}	Δy_{82}	" 0
mean	$\Delta y_{.1}$	$\Delta y_{.2}$	$\Delta y_{.3}$	0

with $\Delta y_{.1} = y_{.1} - y_{..}$.

On each day, that is in each horizontal line of the table, an estimate $\widehat{\Delta A}_j$ of ΔA_j can be made by means of $\widehat{\Delta A}_{ij} = y_{ij} - y_{i.} = \Delta y_{ij}$. Averaging these n estimates, one obtains as the overall best estimate:

$$\widehat{\Delta A}_j = \frac{1}{n}\sum_{i=1}^{n} \widehat{\Delta A}_{ij} = \frac{1}{n}\sum y_{ij} - \frac{1}{n}\sum y_i = y_{.j} - y_{..} = \Delta y_{.j} \qquad (11)$$

N.B. Even in the case that on one or more days one or more measurements fail, equation (11) can be used. The procedure is: calculate the average values of the columns and the overall mean $y_{..}$; the differences then give the $\widehat{\Delta A}_j$ values. Of course $\sum_1^k \widehat{\Delta A}_j = 0$.

Now let us give the interpretation. Referring to (1) and taking e.g. $i = 1$ and $j = 3$, one gets:

$$\underline{z}_{13} = u_1 \varepsilon^{\underline{e}_{13}} \varepsilon^{A_3} \varepsilon^{\underline{B}_3} \quad , \text{ with } \Delta A_3 = A_3 - \bar{A} \quad \text{or } A_3 = \Delta A_3 + \bar{A}$$

(ε = base of natural logarithm). Rewrite:

$$\underline{z}_{13} = u_1 \varepsilon^{\underline{E}_3} \varepsilon^{\Delta A_3} \varepsilon^{\bar{A}} \cong u_1 \; (1 + \underline{E}_{13})(1 + \Delta A_3)(1 + \bar{A}) \; .$$

Now ΔA_3 is estimated as $\widehat{\Delta A}_3$, but $\bar{A}$ remains unknown. Hence, when $\underline{z}_{13}$ is measured (with the purpose to find u_1), we may conclude that the unknown u_1 is situated within a 0.95 confidence interval of broadth $4\sigma_3(\underline{E})$ around a central value $\underline{z}_{13}(1 - \widehat{\Delta A}_3)(1 - \bar{A})$. In the same way on day $i=2$ the true unknown value u_2 lies within the interval of broadth $4\sigma_3(\underline{E})$ around the central value $\underline{z}_{23}(1 - \widehat{\Delta A}_3)(1 - \bar{A})$, whereas the two σ_3 values may be different. Although the true value of $\bar{A}$ is not known, this does not matter, because all central values are affected with the same factor $1 - \bar{A}$ on each of all days. This fact does not influence the "comparability" of the figures.

3.2.2 Accuracy

Consider instrument 3. For each of the n days the value of $\widehat{\Delta A}_{i3}$ is calculated according to (5) and its accuracy $\sigma(\widehat{\Delta A}_{i3})$ according to (7). The accuracy of the overall mean $\widehat{\Delta A}_3$, see (11), follows from the statistical theorem that if $\underline{y} = \underline{x}_1 + \underline{x}_2 + \cdot \cdot \cdot \underline{x}_m$, then $\sigma^2_y = \sigma^2_1 + \sigma^2_2 + \cdot \cdot \sigma^2_m$ in case the variables x are uncorrelated. Application of this theorem here gives:

(12) $$\sigma^2(\widehat{\Delta A}_3) = \frac{1}{n^2} \sum_{i=1}^{n} \sigma^2(\widehat{\Delta A}_{i3}) = \frac{1}{n^2} \sum_{i=1}^{n} \left[(1 - \frac{1}{k})\sigma^2_{i3} + \frac{1}{k^2} \sum_{\substack{j=1 \\ \neq 3}}^{k} \sigma^2_{ij} \right]$$

In case on one or more days one or more measurements would fail, then one should bear in mind that the summation within the brackets in (12) does not comprise $k - 1$ terms. Again σ_{ij} stands for $\sigma(E_{ij})$. Next it is desirable, but not necessary to make the following assumption: σ_{ij} does not depend on the day, that is σ_j is independent of the level of radioactivity, sothat the suffix i may be dropped. Then (12) may be rewritten as:

(13) $\boxed{\sigma^2(\widehat{\Delta A}_3)} = \frac{1}{n}[(1-\frac{1}{k})^2\sigma_3^2 + \frac{1}{k^2}\sum_{\substack{1\\ j\neq 3}}^{k} \sigma_j^2] = \boxed{\frac{1}{nk}[\overline{\sigma^2} + (k-2)\sigma_3^2]}$

The same applies for $j = 1, 2, 4, \ldots k$; here $\overline{\sigma^2} = \frac{1}{k}\sum_1^k \sigma_j^2$

If moreover $\sigma_1 = \sigma_2 = \ldots \sigma_k$, say σ, then

(14) $\sigma^2(\widehat{\Delta A}) = \frac{k-1}{nk}\sigma^2$ or $\boxed{\sigma(\widehat{\Delta A}) = \sigma\sqrt{\frac{k-1}{nk}}}$ for each j

(14a) When k is large, then $\boxed{\sigma(\widehat{\Delta A}) \cong \sigma/\sqrt{n}}$

Here we like to draw attention to the theorem in the footnote on page 8. For instance, for j = 3 we have $\sigma_{i3}^2 = \sigma^2(\underline{e}_{i3}) + \sigma^2(\underline{E}_{i3})$, $i = 1, 2, \ldots n$. We are able to arrange the measurements of the filteractivities in such a way that for each level of radioactivity the $\sigma(\underline{B})$ has the same value. If so, then the hypothesis $\sigma_{13} = \sigma_{23} = \ldots = \sigma_{n3}$ implies the requirement that $\sigma(\underline{e}_{i3})$ is independent of the level. But $\underline{e}_{i3} \cong \lg(\hat{\underline{u}}_{i3} : u_i)$ and $\sigma(e_{i3}) = \sigma(\hat{\underline{u}}_{i3})/u_i$. Consequently, the assumption mentioned above would imply that the percentual standarddeviation of the random spatial error possesses the same value at all levels of radioactivity. Of course we should verify this assumption by means of a statistical test and (or) try to examine its validity by means of physical considerations.

3.3 The "mean instrument" and the "reference" method

Reference is made to the methods i) and ii) treated under 3.1.1. As already said we prefer to choose the "mean instrument" method. The reasons are:

a) Suppose we would choose instrument No. 1 as the reference instrument. Then $\Delta y_2 = y_2 - y_1 = \Delta E_2 + \Delta A_2 \ldots \Delta y_k = y_k - y_1 = \Delta E_k + \Delta A_k$, with $\Delta E_j = E_j - E_1$; $\Delta A_j = A_j - A_1$; $j = 2, 3, \ldots k$. Again Δy_j may be considered as an estimate $\widehat{\Delta A}_j$ of ΔA_j and again this estimate is unbiassed. But what is its accuracy? Let us state for simplicity $\sigma_1 = \sigma_2 \ldots \sigma_k$, say σ, then when measuring simultaneously during n days, each $\widehat{\Delta A}$ $(j = 2, \ldots k)$ possesses the same accuracy $\sigma\sqrt{2/n}$. In the "mean instrument" method, this is $\sigma\sqrt{\frac{k-1}{nk}}$. Since $\frac{k-1}{k} < 1$, the <u>"mean instrument" method may be called better than the "reference" method</u>. For large k, the accuracy of each $\widehat{\Delta A}$ is nearly $\sqrt{2}$ as large.

b) The second reason has to do with the verification of new instruments.

Suppose simultaneous measurements during n_k days have been made with k non-identical instruments 1, 2, 3 . . .k at one and the same place. These measurements have given the estimates $(\widehat{\Delta A_1})_k$, $(\widehat{\Delta A_2})_k$ . . . $(\widehat{\Delta A_k})_k$. Next each instrument is sent back to the station of origine. Now on certain day a new type of instrument, say N$^{\underline{o}}$ k+1, is introduced; there are reasons to suppose that its systematical error differs from the other ones. How to estimate ΔA_{k+1}? Of course it is unpractical to re-install now all k+1 apparates at one and the same place and to perform again simultaneous measurements, say during n_{k+1} days. It is also unnecessary. Suppose one decides to make simultaneous measurements during n_2 days with the apparates 2 and k+1 installed at one and the same place. Let us study how things turn out when applying both methods: I:"mean instrument" method end II: "reference" method.

I. The n_2 duplomeasurements (instrument 2 and k+1) give $(\widehat{\Delta A_{k+1}})_2$ as an estimate of ΔA_{k+1}. If it would have been possible (but it is not) to install all k+1 instruments at one and the same place, then the n_{k+1} simultaneous measurements would have given $(\widehat{\Delta A_{k+1}})_{k+1}$ as an estimate of ΔA_{k+1}. Hence the values $(\widehat{\Delta A_2})_k$ and $(\widehat{\Delta A_{k+1}})_2$ are available and we would like to know $(\widehat{\Delta A_{k+1}})_{k+1}$. The question arises: how to make the best linear combination of $(\widehat{\Delta A_2})_k$ and $(\widehat{\Delta A_{k+1}})_2$? As said already $(\widehat{\Delta A_2})_k$ is an estimate of $(\Delta A_2)_k$ defined by $(\Delta A_2)_k = A_2 - \overline{A_1 + A_2 + \ldots A_k}$. Likewise $(\widehat{\Delta A_{k+1}})_2$ estimates $(\Delta A_{k+1})_2 = A_{k+1} - \overline{A_{k+1} + A_2}$ and $(\widehat{\Delta A_{k+1}})_{k+1}$ estimates $(\Delta A_{k+1})_{k+1} = A_{k+1} - \overline{A_1 + A_2 + \ldots A_k + A_{k+1}}$. It turns out that $(\Delta A_{k+1})_{k+1} = (\frac{2k}{k+1})(\Delta A_{k+1})_2 + (\frac{k}{k+1})(\Delta A_2)_k$.

For this reason we take the linear combination $(\frac{2k}{k+1})(\widehat{\Delta A_{k+1}})_2 + (\frac{k}{k+1})(\widehat{\Delta A_2})_k$ as the best estimate of ΔA_{k+1} .

The estimate is unbiassed since its components are unbiassed.

Next we are interested in the accuracy, say $\sigma_0(I)$, of this linear combination. Take for the sake of simplicity $\sigma_1 = \sigma_2 = \ldots \sigma_k = \sigma_{k+1}$, say σ, then we obtain:

(15) $$\left(\sigma_0(I)\right)^2 = \left[\left(\frac{2k}{k+1}\right)^2 \frac{1}{2n_2} + \left(\frac{k}{k+1}\right)^2 \left(\frac{k-1}{n_k k}\right) \right] \sigma^2$$

In case $n_2 = n_k$, say n, then

(16) $$\sigma_0^2(I) = \frac{k(3k-1)}{n(k+1)^2} \xrightarrow[k \to \infty]{} \frac{3}{n}\sigma^2$$

II. Next the second method, with instrument N° 1 as reference instrument, will be discussed. We would obtain:

$(\widehat{\Delta A_2})_k$ as an estimate of $(\Delta A_2)_k = A_2 - A_1$ (n_k days) and
$(\widehat{\Delta A_{k+1}})_2$ " " " " $(\Delta A_{k+1})_2 = A_{k+1} - A_2$ (n_2 ") and
$(\widehat{\Delta A_{k+1}})_{k+1}$ " " " " $(\Delta A_{k+1})_{k+1} = A_{k+1} - A_1$ (n_{k+1} ").

The values of $(\widehat{\Delta A_2})_k$ and $(\widehat{\Delta A_{k+1}})_2$ are available and we wish to estimate ΔA_{k+1}. This means: there have been made simultaneous measurements during n_k days with the k instruments 1, 2, 3 . . . k, and n_2 simultaneous measurements with 2 instruments N° 2 and k+1. We see $(\Delta A_{k+1})_{k+1} = (\Delta A_2)_k + (\Delta A_{k+1})_2$, sothat we combine $(\widehat{\Delta A_2})_k$ and $(\widehat{\Delta A_{k+1}})_2$ in the same linear way as an estimate of ΔA_{k+1}. This estimate is unbiassed, since its components are unbiassed. Its variance becomes:

(17) $$\sigma_0^2(II) = \left(\frac{2}{n_2} + \frac{2}{n_k}\right)\sigma^2$$ (does not contain k), if again $\sigma_1 = \sigma_2 = \sigma_{k+1}$, say σ,

and

(18) $$\sigma_0^2(II) = \frac{4}{n}\sigma^2, \text{ if } n_2 = n_k, \text{ say } n.$$

We now compare (17) with (15). Since $\sigma_0(I) < \sigma_0(II)$ for each k and each n_2 and n_k we prefer <u>the "mean instrument" method over the "reference" method</u>.

4. <u>Random errors</u>

In the foregoing chapter it has been shown that numerical values of the accuracies of the estimates $\widehat{\Delta A_j}$ of the systematical errors ΔA_j can be computed only if the standarddeviations $\sigma_1, \sigma_2, \ldots \sigma_k$ of the total random errors are known. Thus, the question arises how to find these standarddeviations ?

Another reason to pay attention to these random errors is that a better comparability between the final figures can only be obtained if both systematical and random errors are known.

We shall consider two cases:

a) Simultaneous measurements between k instruments installed at one and the same place are carried out;

b) Simultaneous measurements between k instruments installed at the stations of a given network are carried out.

4.1 <u>Simultaneous measurements at one and the same place</u>

4.1.1 <u>Estimates</u>

4.1.1.1 <u>Duplo measurements; k = 2</u>

Let u_i be the true air radioactivity on day i; i = 1, 2, . . n. Per definition $x_i = \lg u_i$. The instruments give z_{i1}, z_{i2} pC/m^3 and y_{i1}, y_{i2}. The instruments are installed so close to each other that they measure the same true unknown value u_i, sothat:

(19) $$\underline{y}_{i1} = \underline{x}_i + \underline{E}_{i1} + A_1 \quad \text{and} \quad \underline{y}_{i2} = \underline{x}_i + \underline{E}_{i2} + A_2$$

For the sake of simplicity we will drop the index i.
Each day gives a difference $\underline{\Delta} = \underline{y}_1 - \underline{y}_2 = (\underline{E}_1 - \underline{E}_2) + (A_1 - A_2)$. These n differences possess a variance, say s_Δ^2. We have

(20) $$s_\Delta^2 = s_1^2 + s_2^2 - 2rs_1s_2,$$ and if $\underline{E}_1$ and $\underline{E}_2$ are not correlated, then

(20a) $$\sigma_\Delta^2 = \sigma_1^2 + \sigma_2^2 \quad (\text{for } n \rightarrow \infty \; ; \; r \rightarrow \rho = 0).$$

Here s_1^2 is the variance of the n values of y_1 ($n \rightarrow \infty$; $s_1 \rightarrow \sigma_1$) and s_2^2 the variance for y_2, etc.

There are two cases:

Case a): $\sigma_1 = \sigma_2$, say σ ("identical" instruments with respect to the random errors). Then $\sigma_\Delta^2 = 2\sigma^2$ and the best estimate, say $\hat{\sigma}$, of σ is found by means of

(21) $$\boxed{\hat{\sigma} = \tfrac{1}{2}\, s_\Delta \sqrt{2}}$$

Case b): $\sigma_1 \neq \sigma_2$. It can be proved that the only linear combination of s_1^2, s_2^2, s_Δ^2 which gives an unbiassed estimate of σ_1^2 has the form

(22) $\widehat{\sigma_1^2} = \frac{1}{2}[s_\Delta^2 + s_1^2 - s_2^2]$ and in analogy $\widehat{\sigma_2^2} = \frac{1}{2}[s_\Delta^2 + s_2^2 - s_1^2]$.

Attention: one of these σ^2 values lies above, the other below $\frac{1}{2}s_\Delta^2$.

If on the basis of physical or instrumental reasoning one is "certain" that $\sigma_1 = \sigma_2$, we can use (21) immediately. If the equality seems uncertain, it is necessary to apply a statistical test in order to examine whether $\widehat{\sigma_1^2}$ and $\widehat{\sigma_2^2}$, see (22), differ statistically significant or not. Then also the accuracies of these estimates are needed. For this detail see 4.1.2.

Sometimes it is interesting to test the assumption $\rho(\underline{E}_1, \underline{E}_2) = 0$, that is the hypothesis that $\underline{E}_1$ and $\underline{E}_2$ are not correlated, whence (20a) would follow. Therefore we rewrite (22) as follows:

(22a)
$$\widehat{\sigma_1^2} = s_1 s_2 \left(\frac{s_1}{s_2} - r\right)$$
$$\widehat{\sigma_2^2} = s_1 s_2 \left(\frac{s_2}{s_1} - r\right)$$

Here r is the correlation coefficient between the n pairs y_1, y_2 ($n \to \infty$, $r \to \rho$). Generally even in case $\rho = 0$, the value of r will differ from zero.

Attention: $\rho = 0$ does not imply $\sigma_1 = \sigma_2$!

4.1.1.2 Triplo measurements; k = 3

We proceed in the same way:

$\underline{y}_1 = \underline{x} + \underline{E}_1 + A_1$; $\underline{y}_2 = \underline{x} + \underline{E}_2 + A_2$; $\underline{y}_3 = \underline{x} + \underline{E}_3 + A_3$

Again there are two cases:

Case a): $\sigma_1 = \sigma_2 = \sigma_3$, say σ. When defining $s_{12}^2 \equiv s^2(y_1 - y_2)$ = variance of the n values $y_1 - y_2$, etc. we obtain

(23)
$$\boxed{\widehat{\sigma^2} = \frac{1}{6}\{s^2(y_1-y_2) + s^2(y_2-y_3) + s^2(y_3-y_1)\} \equiv \frac{1}{6}(s_{12}^2 + s_{23}^2 + s_{31}^2)}$$

Case b): $\sigma_1 \leqslant \sigma_2 \leqslant \sigma_3$ (at least one < sign).

It can be proved that the best estimates $\widehat{\sigma_1^2}$, $\widehat{\sigma_2^2}$, $\widehat{\sigma_3^2}$ of σ_1^2, σ_2^2 and σ_3^2 are:

$$(24)\qquad \widehat{\sigma_1} = \tfrac{1}{2}\,(s_{12}^2 - s_{23}^2 + s_{13}^2)$$
$$\widehat{\sigma_2} = \tfrac{1}{2}\,(s_{23}^2 - s_{31}^2 + s_{12}^2)$$
$$\widehat{\sigma_3} = \tfrac{1}{2}\,(s_{31}^2 - s_{12}^2 + s_{23}^2)$$

As soon as $\sigma_1 = \sigma_2 = \sigma_3$, say σ, then (24) gives three estimates of the same (σ^2) and averaging these estimates gives expression (23).

It is possible to rewrite (24) in such a way that the correlation coefficients $r_{12} \equiv r(y_1, y_2)$, r_{31} and r_{12} (which are not related) appear. We get

$$(25)\qquad \widehat{\sigma_1^2} = s_1^2\left(\frac{r_{12}-r_{13}}{r_{23}}\right); \quad \widehat{\sigma_2^2} = s_2^2\left(1 - \frac{r_{23}r_{12}}{r_{13}}\right); \quad \widehat{\sigma_3^2} = s_3^2\left(1 - \frac{r_{13}r_{23}}{r_{12}}\right)$$

4.1.1.3 Four simultaneous measurements; $k = 4$

Now $\underline{y}_1 = \underline{x} + \underline{E}_1 + A_1 \; . \; . \; . \; \underline{y}_4 = \underline{x} + \underline{E}_4 + A_4$. Again there are two cases:

Case a): $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4$, say σ. One obtains:

$$(26)\qquad \boxed{\widehat{\sigma^2} = \frac{1}{12}\,(s_{12}^2 + s_{13}^2 + s_{14}^2 + s_{23}^2 + s_{24}^2 + s_{34}^2).}$$

For the proof, see in 4.1.1.4.

Case b): $\sigma_1 \leqslant \sigma_2 \leqslant \sigma_3 \leqslant \sigma_4$ (at least one $<$ sign). Then:

$$(27)\qquad \widehat{\sigma_1^2} = \frac{1}{3}\,(s_{12}^2 + s_{13}^2 + s_{14}^2) - \frac{1}{6}\,(s_{23}^2 + s_{24}^2 + s_{34}^2)$$
$$\widehat{\sigma_2^2} = \frac{1}{3}\,(s_{23}^2 + s_{24}^2 + s_{12}^2) - \frac{1}{6}\,(s_{13}^2 + s_{14}^2 + s_{34}^2)$$
$$\widehat{\sigma_3^2} = \frac{1}{3}\,(s_{34}^2 + s_{31}^2 + s_{32}^2) - \frac{1}{6}\,(s_{12}^2 + s_{14}^2 + s_{24}^2)$$
$$\widehat{\sigma_4^2} = \frac{1}{3}\,(s_{41}^2 + s_{42}^2 + s_{43}^2) - \frac{1}{6}\,(s_{12}^2 + s_{13}^2 + s_{23}^2)$$

When trying to rewrite (27) by introducing the correlation coefficients $r_{12}, r_{13}, r_{14}, r_{23}, r_{24}, r_{34}$ we should bear in mind that as soon as $k > 3$ these correlation coefficients are interrelated. This can easily be seen as follows. Since $y_1 = x + \underline{E}_1 + A_1 \; . \; . \; . \; y_4 = x + \underline{E}_4 + A_4$ and because $\underline{E}_r$ and $\underline{E}_t$ are independent ($r \neq t$), we have:

$r_{12} = s_x^2/s_1s_2$; $r_{13} = s_x^2/s_1s_3$; $r_{14} = s_x^2/s_1s_4$ etc. Defining $d_1 = s_x/s_1$ etc. we get 6 equations with 4 unknowns d_1, d_2, d_3, d_4, namely $r_{12} = d_1d_2$; $r_{13} = d_1d_3$; $r_{14} = d_1d_4$; $r_{23} = d_2d_3$; $r_{24} = d_2d_4$; $r_{34} = d_3d_4$. Consequently there must be interrelations. They are: $r_{24} = r_{14}r_{23}/r_{13}$ and $r_{34} = r_{14}r_{23}/r_{12}$.

Introducing these r's in (27) we obtain:

(28) $$\widehat{\sigma_1^2} = s_1^2\left(1-\frac{r_{12}r_{13}}{r_{23}}\right);\quad \widehat{\sigma_2^2} = s_2^2\left(1-\frac{r_{23}r_{24}}{r_{34}}\right)$$
$$\widehat{\sigma_3^2} = s_3^2\left(1-\frac{r_{34}r_{31}}{r_{41}}\right);\quad \widehat{\sigma_4^2} = s_4^2\left(1-\frac{r_{41}r_{42}}{r_{12}}\right)$$

<u>N.B.</u> As soon as $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4$, say σ, then (27) gives 4 estimates of the same σ^2. Averaging gives exactly (26).

4.1.1.4 <u>Simultaneous measurements with k instruments</u>

<u>General formulas</u>

Now $y_1 = \underline{x} + \underline{E}_1 + A_1 \dots \underline{y}_k = \underline{x} + \underline{E}_k + A_k$

Case a): $\sigma_1 = \sigma_2 = \dots \sigma_k$, say σ.

(29) $$\boxed{\widehat{\sigma^2} = \frac{1}{k(k-1)}\quad s_{12}^2 + s_{13}^2 + \dots s_{1k}^2 + s_{23}^2 + \dots s_{2k}^2 + \dots s_{k-1,k}^2}$$

This may be proved as follows.

The variance of the values y_{i1}, y_{i2}, . . y_{ik} on day i is defined by

(30) $$\sigma_i^2 = \frac{1}{k-1}\sum_{j=1}^{k}(y_{ij} - y_{i.})^2\ ,\ \text{where}\ y_{i.} = \frac{1}{k}\sum_{j=1}^{k} y_{ij}$$

On each of the n days (i=1, 2, . . n) an expression as (30) can be calculated. Because of the assumption that σ does not depend on the level, the overall best estimate $\widehat{\sigma^2}$ of σ^2 is made by pooling these n values (30). Pooling[1]) gives :

Note[1]) On one or more days one or more measurements may drop out. Then pooling should be carried out in the correct way, that is by taking into consideration the correct number of degrees of freedom: k_i ($\leqslant k$) instruments on day i, then $d_i = k_i - 1$ = number of degrees of freedom on day i and $\widehat{\sigma^2} = \sum_1^n s_i^2\, d_i \; : \sum_1^n d_i$. When each $k_i = k$ then expression (31) again holds.

$$(31)\quad \widehat{\sigma^2} = \frac{(k-1)s_1^2 + (k-1)s_2^2 + \ldots (k-1)s_n^2}{n(k-1)} = \sum_{i=1}^{n}\sum_{j=1}^{k}(y_{ij}-y_{i.})^2 \Big/ n(k-1)$$

Next the following mathematical theorem is used:

$$(32)\quad \sum_{j=1}^{k}(y_{ij}-y_{i.})^2 = \frac{1}{k}\sum_{\substack{r,t=1\\ t>r}}^{k}(y_{ir}-y_{it})^2$$

Substituting for instance k=3 (32) furnishes

$$(y_{i1}-y_{i.})^2 + (y_{i2}-y_{i.})^2 + (y_{i3}-y_{i.})^2 = \frac{1}{3}\ (y_{i1}-y_{i2})^2 + (y_{i2}-y_{i3})^2 + (y_{i3}-y_{i1})^2$$

Using this theorem in (31) we obtain:

$$(33)\quad \boxed{\widehat{\sigma^2} = \frac{1}{k(k-1)}\sum_{\substack{r,t=1\\ t>r}}^{k} s_{r,t}^2} \quad \text{if } s_{r,t}^2 = s^2(y_r-y_t)$$

Substituting k=3 one obtains (23)

Case b): $\sigma_1 \leqslant \sigma_2 \leqslant \ldots \leqslant \sigma_k$ (at least one $<$ sign)

$$(34)\quad \boxed{\widehat{\sigma_1^2} = \frac{1}{k-1}\left[\sum_{j=2}^{k} s_{1j}^2 - \frac{1}{k-2}\sum_{\substack{r,t=2\\ t>r}}^{k} s_{rt}^2\right]}$$

In general form $(1<j<k)$

$$(34a)\quad \boxed{\widehat{\sigma_j^2} = \frac{1}{k-1}\left[\sum_{\substack{v=1\\ v\neq j}}^{k} s_{jv}^2 - \frac{1}{k-2}\sum_{\substack{r,t=1\\ r,t\neq j\\ t>r}}^{k} s_{rt}^2\right]}$$

4.1.2 Accuracies

One should not be content with only the estimates $\widehat{\sigma^2}$ of σ^2 without knowing the accuracies of these estimates. An extremely inaccurate $\widehat{\sigma^2}$ is almost valueless. Moreover when verifying whether two or more estimates $\widehat{\sigma_j^2}$ are statistically different, the application of a statistical test needs knowledge of the accuracies of these variances.

The case $\sigma_1 = \sigma_2 = \ldots \sigma_k$ is the simplest one. We will start with treating this situation and refer to some well-known statistical formulas.

When one draws random samples of m elements from any population and computes the variance s^2 in each of the samples of constant size, then the value $\underline{s}^2$ is a stochastical variable with a mean value:

(35) $$\mathcal{E}(\underline{s}^2) = \sigma^2$$

and a variance:

(36) $$\sigma^2(\underline{s}^2) = \frac{1}{m}\left(\mu_4 - \frac{m-3}{m-1}\sigma^4\right).$$

Here $s^2 = \frac{1}{m-1}\sum_1^m (x_i - \bar{x})$; $\bar{x} = \frac{1}{m}\sum_1^m x_i$;
σ^2 = variance = 2^{th} moment in the x population:
s^2 is an estimate of σ^2; $\mu_4 = 4^{th}$ moment in the x population. In particular for normal populations we have:

(37) $$\sigma^2(\underline{s}^2) = \frac{2}{m-1}\sigma^4 \quad \text{or} \quad \sigma(\underline{s}^2) = \sigma^2\sqrt{\frac{2}{m-1}}$$

In the same way the standarddeviation s, defined by $\sqrt{s^2}$, is a stochastical variable, with

(38) $$\mathcal{E}(\underline{s}) \cong \sigma\sqrt{1 - \frac{1}{2(m-1)}} \cong \sigma\sqrt{1 - \frac{1}{2m}} \quad \text{and}$$

(39) $$\sigma^2(\underline{s}) \cong \sigma^2/2(m-1) \cong \sigma^2/2m \quad \text{approximately for large m.}$$

Referring to (31), where the estimate $\widehat{\sigma^2}$ is based on n(k-1) degrees of freedom, we replace m by n(k-1) and s^2 by $\widehat{\sigma^2}$, so that (37) and (39) give

(40) $$\sigma(\widehat{\sigma^2}) \cong \sigma^2\sqrt{2/n(k-1)} \qquad \text{for } n(k-1) \gg 1$$

(41) $$\sigma(\widehat{\sigma}) \cong \sigma\sqrt{1/2n(k-1)}$$

Since σ^2 is known, but estimated as $\widehat{\sigma^2}$, this estimate $\widehat{\sigma^2}$ must be substited for σ^2 in (40) and $\sqrt{\widehat{\sigma^2}}$ must be substituted for σ in (41). Then

(42) $$\widehat{\sigma}(\widehat{\sigma^2}) \cong \widehat{\sigma^2}\sqrt{2/n(k-1)} \text{ and } \widehat{\sigma}(\widehat{\sigma}) \cong \widehat{\sigma}\sqrt{1/2n(k-1)}, \text{ with } \widehat{\sigma} = \sqrt{\widehat{\sigma^2}}$$

Interpretation (although not quite correct): the 95% confidence region for the unknown σ^2 lies between

(43) $$\widehat{\sigma^2} - 2\widehat{\sigma}(\widehat{\sigma^2}) \quad \text{and} \quad \widehat{\sigma^2} + 2\widehat{\sigma}(\widehat{\sigma^2})$$

and (only approximatively) the interval for σ lies between

(44) $\sqrt{\widehat{\sigma^2} - 2\hat{\sigma}(\widehat{\sigma^2})}$ and $\sqrt{\widehat{\sigma^2} + 2\hat{\sigma}(\widehat{\sigma^2})}$ or, better, $\hat{\sigma} - 2\hat{\sigma}(\hat{\sigma})$ and $\hat{\sigma} + 2\hat{\sigma}(\hat{\sigma})$

4.1.2.1 Duplo measurements; $k = 2$[1)]

There ar two cases a) and b).

Case a): $\sigma_1 = \sigma_2$, say σ

Substituting k = 2 in (42) we obtain:

(45) $$\boxed{\hat{\sigma}(\widehat{\sigma^2}) = \widehat{\sigma^2}\sqrt{2/n}}$$

Case b): $\sigma_1 \neq \sigma_2$

It can be proved that the variance of the value $\widehat{\sigma_1^2}$, which is a statistical variable, estimated by means of the n pairs y_1, y_2 via (22), is given by

(46) $$\sigma^2(\widehat{\sigma_1^2}) = \frac{2}{n-1}\sigma_1^4 + \frac{1}{n-1}\{\sigma_x^2(\sigma_1^2 + \sigma_2^2) + \sigma_1^2\sigma_2^2\}$$ and likewise

$$\sigma^2(\widehat{\sigma_2^2}) = \frac{2}{n-1}\sigma_2^4 + \frac{1}{n-1}\{\sigma_x^2(\sigma_1^2 + \sigma_2^2 + \sigma_1^2\sigma_2^2\}$$

The values σ_1^2, σ_2^2, σ_x^2 are the variances of the values y_1, y_2, x if $n \to \infty$. Here the variance of the true unknown value x (varying from day to day) appears. For $k \geqslant 3$ σ_x^2 does not enter into the expressions.

In practice we know neither σ_1, nor σ_2, nor σ_x. Of course we substitute $\hat{\sigma}_1$ and $\hat{\sigma}_2$ from (22) for σ_1 resp. σ_2, but what about σ_x ?

We have $\sigma^2(y_1) = \sigma_x^2 + \sigma_1^2$ and $\sigma^2(y_2) = \sigma_x^2 + \sigma_2^2$ or $\sigma_x^2 = \sigma^2(y_1) - \sigma^2 = \sigma^2(y_2) - \sigma^2$, so that it is obvious to substitute for $s_x^2 \equiv \widehat{\sigma_x^2}$ the expression

(47) $$\tfrac{1}{2}\{(s^2(y_1) - \widehat{\sigma_1^2}) + (s^2(y_2) - \widehat{\sigma_2^2}\} = \tfrac{1}{2}\{s^2(y_1) + s^2(y_2) - s_\Delta^2\}$$

Substituting (47) for σ_x^2 and $\widehat{\sigma_1^2}$ resp. $\widehat{\sigma_2^2}$ from (22) for σ_1^2 and σ_2^2 in (46) we finally obtain:

Note 1) See also the paper "On estimating precision of measuring instruments and product variability" by F.E.Grubbs in Journ.Am. Stat. Ass. 43 243 1948.

(48)
$$\widehat{\sigma^2}(\widehat{\sigma_1^2}) = \frac{1}{n-1}\left[\ \widehat{\sigma_1^4} - s_\Delta^2 s^2(y_1)\right]$$
$$\widehat{\sigma^2}(\widehat{\sigma_2^2}) = \frac{1}{n-1}\left[\ \widehat{\sigma_2^4} - s_\Delta^2 s^2(y_2)\right]$$

, if $\sigma_x \neq 0$. Here $\widehat{\sigma_1^4} \cong [\widehat{\sigma_1^2}]^2$

If one is certain (on non-statistical grounds) that x does not vary during the n days of simultaneous measurements ($\sigma_x = 0$), then (46) may be rewritten with the help of (22) as

(49)
$$\widehat{\sigma^2}(\widehat{\sigma_1^2}) = \frac{\widehat{\sigma_1^2}}{n-1}\ (\widehat{\sigma_1^2} + s_\Delta^2)$$
$$\widehat{\sigma^2}(\widehat{\sigma_2^2}) = \frac{\widehat{\sigma_2^2}}{n-1}\ (\widehat{\sigma_2^2} + s_\Delta^2)$$

if $\sigma_x = 0$

These expressions may be used when it is known that during the simultaneous measurements the level was constant day after day.

4.1.2.2 Triplo measurements; $k = 3$

Case a): $\sigma_1 = \sigma_2 = \sigma_3$, say σ. Then (42) gives (k=3)

(50)
$$\widehat{\sigma}(\widehat{\sigma^2}) = \widehat{\sigma^2}\sqrt{1/n}$$

Case b): $\sigma_1 \leqslant \sigma_2 \leqslant \sigma_3$ (at least one <sign)

It can be shown that the variances of $\widehat{\sigma_1^2}$, $\widehat{\sigma_2^2}$, $\widehat{\sigma_3^2}$ are

(51)
$$\sigma^2(\widehat{\sigma_1^2}) = \frac{2}{n-1}\sigma_1^4 + \frac{1}{n-1}\left[\sigma_1^2\sigma_2^2 + \sigma_2^2\sigma_3^2 + \sigma_3^2\sigma_1^2\right]$$
$$\sigma^2(\widehat{\sigma_2^2}) = \frac{2}{n-1}\sigma_2^4 + \frac{1}{n-1}\left[\sigma_2^2\sigma_3^2 + \sigma_3^2\sigma_1^2 + \sigma_1^2\sigma_2^2\right]$$
$$\sigma^2(\widehat{\sigma_3^2}) = \frac{2}{n-1}\sigma_3^4 + \frac{1}{n-1}\left[\sigma_3^2\sigma_1^2 + \sigma_1^2\sigma_2^2 + \sigma_2^2\sigma_3^2\right]$$

In case a statistical test would lead to the conclusion that $\sigma_1, \sigma_2, \sigma_3$ do not differ, it is not correct to average the three expressions in (51), but (50) should be used. Averaging would furnish $\sigma^2(\widehat{\sigma^2}) = \frac{5}{n-1}\sigma^4$, but (50) gives $\frac{1}{n}\sigma^4$.

Substituting the estimates (28) in (51) and $s_{12}^2 \equiv$ var. $(y_1 - y_2)$ etc. we obtain as estimates:

(52)
$$\widehat{\sigma^2}(\widehat{\sigma_1^2}) = \frac{1}{n-1}\,[\widehat{\sigma_1^4} + s_{12}^2 s_{13}^2]$$
$$\widehat{\sigma^2}(\widehat{\sigma_2^2}) = \frac{1}{n-1}\,[\widehat{\sigma_2^4} + s_{23}^2 s_{21}^2]$$
$$\widehat{\sigma^2}(\widehat{\sigma_3^2}) = \frac{1}{n-1}\,[\widehat{\sigma_3^4} + s_{31}^2 s_{32}^2]$$

Such expressions would facilitate the computations. Firstly one makes the difference y_1-y_2, y_2-y_3, y_3-y_1. Next one computes the variance s_{12}^2 of the n values y_1-y_2, and likewise s_{23}^2, s_{31}^2. Then with the help of (28) the estimates of the accuracies are calculated. Finally the estimates of the accuracies of these estimates can be calculated, see (52).

Accuracy of $\widehat{\sigma_1^2}$ is defined by $\sqrt{\widehat{\sigma^2}(\widehat{\sigma_1^2})}$ etc.

4.1.2.3 Four simultaneous measurements; k = 4

Case a): $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4$, say σ. See (42), we obtain

(53)
$$\hat{\sigma}(\widehat{\sigma^2}) = \widehat{\sigma^2}\sqrt{2/3n}$$

Case b): $\sigma_1 \leqslant \sigma_2 \leqslant \sigma_3 \leqslant \sigma_4$ (at least one $<$ sign). We obtain

$$\sigma^2(\widehat{\sigma_1^2}) = \frac{2\sigma_1^4}{n-1} + \frac{1}{n-1}\,[\frac{4}{3^2}(\sigma_1^2\sigma_2^2 + \sigma_1^2\sigma_3^2 + \sigma_1^2\sigma_4^2) + \frac{4}{3^2 2^2}(\sigma_2^2\sigma_3^2 + \sigma_2^2\sigma_4^2 + \sigma_3^2\sigma_4^2)]$$

(54) etc.

$$\sigma^2(\widehat{\sigma_4^2}) = \frac{2\sigma_4^4}{n-1} + \frac{1}{n-1}\,[\frac{4}{3^2}(\sigma_4^2\sigma_1^2 + \sigma_4^2\sigma_2^2 + \sigma_4^2\sigma_3^2) + \frac{4}{3^2 2^2}(\sigma_1^2\sigma_2^2 + \sigma_1^2\sigma_3^2 + \sigma_2^2\sigma_3^2)]$$

When substituting the estimates $\widehat{\sigma_j^2}$ from (32) in (54) it will be possible to simplify the expressions; we did not work out this substitution in this report.

4.1.2.4 General case; k instruments

Case a): $\sigma_1 = \sigma_2 = \ldots\ldots \sigma_k$, say σ.

(55) $$\hat{\sigma}(\widehat{\sigma^2}) = \widehat{\sigma^2}\sqrt{2/n(k-1)}$$

Case b): $\sigma_1 \leqslant \sigma_2 \leqslant \ldots\ldots \leqslant \sigma_k$ (at least one $<$ sign).

(56) $$\sigma^2(\widehat{\sigma_1^2}) = \frac{2}{n-1}\sigma_1^4 + \frac{1}{n-1}\left[\frac{4\sigma_1^2}{(k-1)^2}\sum_2^k \sigma_j^2 + \frac{4}{(k-1)^2(k-2)^2}\sum_{\substack{r,t=2\\ t>r}}^k \sigma_r^2\sigma_t^2\right]$$

Analoguously for $j = 2,3 \ldots\ldots, k$.

N.B. We have always distinguished two cases a): $\sigma_1 = \sigma_2 = \ldots\ldots \sigma_k$ and b): $\sigma_1 \leqslant \sigma_2 \leqslant \ldots\ldots \leqslant \sigma_k$ (at least one $<$ sign). The second case should be understood as follows: we know nothing a priori about the mutual relations between the σ_j values. It may be that $\sigma_1 = \sigma_2 = \sigma_3 < \sigma_4 < \sigma_5$ (k=5), but also that $\sigma_1 < \sigma_2 < \sigma_3 = \sigma_4 = \sigma_5$ etc. If we knew the relation, we could use this information and quite different expressions for $\widehat{\sigma_j^2}$ and $\sigma^2(\widehat{\sigma_j^2})$ would result. We did not treat such a case in the present report.

4.1.3 How to find a firm estimate of σ ?

Suppose we know for certain that some instruments possess the same standarddeviation σ of the random errors, but σ itself is unknown. We want to find a firm estimate of σ. How long should we measure with these instruments simultaneously at one and the same place and is it possible to decrease the minimum duration of these measurements by placing more identical instruments close to each other?

The value of $\hat{\sigma}$ which can be considered as a firm estimate of σ can not a priori be indicated; it is more or less a question of taste. Of course one should consider the accuracy $\sigma(\hat{\sigma})$ with which σ is found. The requirement, which is to be satisfied, will be of the type

$$\sigma(\hat{\sigma})/\hat{\sigma} \leqslant t \quad ; \quad \text{e.g. } t = 0.05;\ 0.1.$$

Then (41), when substituting $\hat{\sigma}$ for σ, gives

(57) $$n \geqslant [\,2(k-1)t^2]^{-1}$$ days of simultaneous measurements.

If t = 0.1 then $n \geqslant 50/(k-1)$. Thus, with 2 identical instruments one should make at least 50 simultaneous measurements; with 3 at least 25; with 4 at least 17; with 11 at least 5; with 51 at least 1. At first sight the decrease of the minimum n with increasing k looks strange, but it should be realized that the basic assumption is that σ does not depend on the level of radioactivity. Then one could make simultaneous measurements with a few identical instruments on each of many days as well as simultaneous measurements with many identical instruments on each of a few days.

4.2 Simultaneous measurements in a network of stations

The statistical considerations with regard to the estimate of the standarddeviation of the total random errors when performing simultaneous measurements in the stations of a given network instead of at one place (per station one instrument; k stations) have been developed by Drs. W.J.A. Kuipers. Since these considerations would fall outside the scope of the present report (matrices algebra, etc.) we only give here the line of thought. The reader is referred to a separate report of Kuipers for details.

If the true values u_i of the specific artificial atmospheric radioactivity on day i possess the same value at all k stations, then this situation is completely similar to the situation that all instruments have been installed close to each other at one at the same place. How to know whether all u_i values are equal? Sometimes on physical grounds the assumption may be considered as true. If not, then the next simple hypothesis would be that the field values u_{ij} on day i are linear functions of the positions of the stations. If u_{ij} were plotted in a vertical direction then the u_i surface would be flat. If ξ_j, η_j are the cartesian coordinates of station j and if $x_{ij} = \lg u_{ij}$ on day i at station j, then we suppose

(58) $$x_{ij} = \mu_{i1} + \mu_{i2}\,\xi_j + \mu_{i3}\,\eta_j \;.$$

The number of days is n, the number of stations k; i = 1,2,....n; j = 1,2,....k.

The true values of the coefficients μ are unknown. Since the position of the plane surface u_i will vary from day to day, the true values $\mu_{i1}, \mu_{i2}, \mu_{i3}$ will also vary from day to day.

(59) Next write $y_{ij} = x_{ij} + E_{ij}$; E_{ij} = random error: $\mathcal{E}\, E_{ij} = 0$ for each i and each j.

We suppose that systematical errors are absent or that corrections for systematical errors have been applied. On each of the n days the values of $\mu_{i1}, \mu_{i2}, \mu_{i3}$ should be estimated, say $\widehat{\mu}_{i1}, \widehat{\mu}_{i2}, \widehat{\mu}_{i3}$, by means of the values $y_{i1}, y_{i2} \ldots\ldots y_{ik}$ (i = 1,2 .... n) on each day separately. With the help of these estimates $\widehat{\mu}_{i1}, \widehat{\mu}_{i2}, \widehat{\mu}_{i3}$ we can calculate on each day new estimates $\widehat{y}_1, \widehat{y}_2 \ldots\ldots \widehat{y}_k$ of $x_1, x_2 \ldots\ldots x_2$ using (58) by substituting $\widehat{\mu}_1, \widehat{\mu}_2, \widehat{\mu}_3$ for μ_1, μ_2, μ_3. Consequently, we have two estimates of the same unknown x_{ij}, namely y_{ij} and $\widehat{y}_{ij}$ for each day.

(60) Let be $\widehat{E}_{ij} = y_{ij} - \widehat{y}_{ij}$.

Let be again $\sigma_1, \sigma_2 \ldots\ldots \sigma_k$ the standarddeviations of the total random errors at the k stations (independent of the level of radioactivity!). It is possible to derive the following expressions

(61)
$$\sigma^2(\widehat{E}_1) = D_{11}^2\sigma_1^2 + \ldots\ldots\ D_{k1}^2\sigma_k^2$$
$$\sigma^2(\widehat{E}_2) = D_{21}^2\sigma_1^2 + \ldots\ldots\ D_{k2}^2\sigma_k^2$$
etc.
$$\sigma^2(\widehat{E}_k) = D_{k1}^2\sigma_1^2 + \ldots\ldots\ D_{kk}^2\sigma_k^2$$

Here the D's are functions of the coordinates of the stations in the network. They are the elements of a determinant, the value of which is identically zero.

The measurements on the n days give estimates of $\sigma^2(\widehat{E}_1) \ldots\ldots \sigma^2(\widehat{E}_k)$. Although it is not necessary to assume $\sigma_1 = \sigma_2 = \ldots\ldots \sigma_k$, say σ, this assumption simplifies the calculation considerably. The best estimate $\widehat{\sigma^2}$ of σ^2 becomes

(62)
$$\widehat{\sigma^2} = \frac{s^2(E_1)}{\sum_1^k D_{j1}^2} = \frac{s^2(E_2)}{\sum_1^k D_{j2}^2} = \ldots\ldots \frac{s^2(E_k)}{\sum_1^k D_{jk}^2}$$

This expression is so simple only as a consequence of the assumption that the u field would be linear in all directions. As soon as this assumption is not valid, expressions of degree two, three etc. in the cartesian coordinates should be used instead of (58). Then the expression for $\widehat{\sigma^2}$ would become very complicate. For this reason it is highly desirable to examine whether the mentioned assumption is to be rejected or not. Such a statistical test needs knowledge of the accuracies of the estimates $\widehat{\mu_1}$, $\widehat{\mu_2}$, $\widehat{\mu_3}$ on each of the n days. It is possible to derive rather simple expressions for these accuracies. Once knowing these accuracies it is also possible to select all the days on which the true values u, although belonging to different stations, are equal or "almost equal" (then we may speak of pseudo simultaneous measurements).

5. The proposition of Grandjean

5.1 Introduction

At the Euratom meeting of 26 November 1959 the Belgium delegate Dr. J. Grandjean (Royal Meteorological Institute) proposed to make a thorough comparison between the instruments, which measure the atmospheric artificial radioactivity in the various countries. When studying the figures given by the stations of the European network, especially when they refer to one and the same day, it seems obvious that they are not sufficiently "comparable" and that they should be corrected for systematical errors, for some stations even in a marked degree. A way to find the necessary reduction factors would be to install all typical instruments at one and the same central place and then to make here a series of simultaneous measurements. If the instruments are installed close to each other, the differences between the daily figures can only be of systematical and random nature since it is permitted to say that they are all measuring the same true but unknown atmospheric radioactivity. Prof. Bleeker, the Netherlands' delegate at the meeting, expressed the desirability to examine all statistical aspects beforehand and to try to estimate the minimum duration of such an experiment. Such a study could also give some indications about measurements to be performed by each of the participating countries prior to the comparison.

Grandjean's proposition

Survey of formulae

		A		B	
		$\sigma_1, \sigma_2, \ldots \sigma_k$ are <u>exactly known</u> or at least have been measured very accurately beforehand		$\sigma_1, \sigma_2 \ldots \sigma_k$ are <u>not known</u>. Estimates $\widehat{\sigma}, \widehat{\sigma_2} \ldots$ are made during the Grandjean exp.	
$\sigma_1 = \sigma_2 = \ldots = \sigma_k$, say σ	Duration c	$n = \frac{k-1}{k} f^2$	$f = \sigma/d$; requirement $\sigma(\widehat{\Delta A}) = \text{const}.d$	$n \approx \frac{k-1}{k} f^{*2}$	$f^* = fg \quad g = 1 + \frac{1}{f : \sqrt{2/k}} > 1$
	Selection c	$\lvert\widehat{\Delta A}_j\rvert > 4\sigma c = 4\sigma(1 + \frac{1}{2f})$ each j	$c = 1 + \frac{1}{2}\sqrt{\frac{k-1}{kn}} = 1 + \frac{1}{2f}$	$\lvert\widehat{\Delta A}_j\rvert > 4\widehat{\sigma} c c' = 4\widehat{\sigma}(1 + \frac{1}{2f^*})(1 + \frac{\sqrt{2}}{k-1}\frac{1}{f^*})$ each j	$c' = 1 + \frac{\sqrt{2}}{\sqrt{n(k-1)}}$
$\sigma_1 \leqslant \sigma_2 \leqslant \ldots \leqslant \sigma_k$ not all equal	Duration c	$n = \frac{k-1}{k} f^2 T^2$	$f = \bar{\sigma}/d$; $T = F_k/\bar{\sigma}$; requirement $\sigma(\widehat{\Delta A}_j) \leqslant d$, $j = 1, \ldots k$; $F_j = \sqrt{\{\bar{\sigma}^2 + (k-2)\sigma_j^2\} : (k-1)}$; $\bar{\sigma} = \frac{1}{k}\Sigma\sigma_j$; $\bar{\sigma^2} = \frac{1}{k}\Sigma\sigma_j^2$	Try to find an interdependence prior to the Grandjean experiments $\sigma_1 : \sigma_2 : \ldots \sigma_k = a_1 : a_2 : \ldots a_k$ Then F_k can be expressed in σ_1 and n can be estimated	
	Selection criterion	$\lvert\widehat{\Delta A}_j\rvert > 4\sigma_j + 2F_j\sqrt{\frac{k-1}{nk}} = 4\bar{\sigma}(q_j + \frac{2}{f} t_j)$	$q_j = \sigma_j/\bar{\sigma} \lesseqgtr 1$; $t_j = F_j/F_k \leqslant 1$	$\lvert\widehat{\Delta A}_j\rvert > 4\tau_j + 2F_j(\tau)\sqrt{\frac{k-1}{nk}}$ n estimated above	$\tau_j = \widehat{\sigma_j} + 2\widehat{\sigma}(\widehat{\sigma_j})$; $\widehat{\sigma}(\widehat{\sigma_j}) = \frac{\widehat{\sigma_j}}{\sqrt{2n(k-1)}}$ only approx.

Basic principles:

1) Require that the percentual accuracy of each estimated systematical error decreases with increasing absolute accuracy. This requirement can be fulfilled 1) in case $\sigma_1 = \ldots \sigma_k$, by putting $\sigma(\widehat{\Delta A}) = \text{const} = d$; each j.
2) in case σ's are unequal, putting $\sigma(\widehat{\Delta A_j}) \leqslant d$; $j = 1, \ldots k$
2) Apply a correction for systematical error only if the whole 0.95 confidence region for the unknown true value around the not corrected measurement lies completely outside of the analoguous one around the corrected measurement. In case the σ's are not known exactly, but are to be estimated by means of the Grandjean experiments, then take account of the inaccuracies of the σ values and hence of the $\widehat{\Delta A}_j$ values.
3) In the last instance the numerical values of both the <u>minimum duration</u> and the <u>selection level</u> depend on one and the same parameter f, the numerical value of which should be chosen by the user of the data.

The principal purpose of the proposed simultaneous measurements, suggested by Grandjean, is to compute the so called systematical error (s.e.) with which the measurements should be corrected so that the corrected values are better comparable than the original ones.

A theoretical statistical attack of the question immediately leads to asking three questions:

a) Since the Grandjean experiments will be performed during a finite (preferably a small) number of days, the resulting systematical errors will be inaccurate. Which accuracies are required for practical use?

b) It seems senseless to correct for very small s.e. Which level should be surpassed before corrections are applied?

c) The performance of the Grandjean experiments will cost time and money (the instruments must be brought to a central point; simultaneous measurements are to be made very carefully; the results will have to be analyzed statistically, and so on). We like to know, even before the experiments will start, during how many days simultaneous experiments should be made. How to estimate this minimum duration?

This leads to the development of a so called duration- and selection criterion.

The duration criterion D.C. aims at giving a minimum number of days of simultaneous measurements with the instruments of the participating countries at one central point in order to fulfil specified conditions.

The selection criterion S.C. After having made n simultaneous measurements (n given by the D.C.), the s.e.'s can be calculated; $\widehat{\Delta A}_j$ $j = 1 \ldots k$. There may be small and large ones (in a relative sense). The S.C. is intended to indicate the level $|\widehat{\Delta A}_j|$ for each of the instruments beyond which the measurements made with the instrument j should be corrected both in the simultaneous experiments and in future as soon as the instrument is brought back to the country of origin.

Of course it is desirable to start formulating criteria as general and as objective as possible, but in the last instance, when numerical values are to be decided upon, the preference of the user is decisive.

We consider the following situations.

A. The standarddeviations σ_1, σ_2, ... σ_k of the random errors are exactly known or highly accurately measured a priori.

A.1. $\sigma_1 = \sigma_2 = ... \sigma_k$, say σ.

A.2. The equality does not hold; say $\sigma_1 \leqslant \sigma_2 \leqslant .... \leqslant \sigma_k$ (at least one $<$ sign).

B. The σ_1 ...., σ_k are not known a priori, but are to be measured during the Grandjean experiments.

B.1. $\sigma_1 = \sigma_2 = ... \sigma_k$, say σ.

B.2. The equality does not hold; say $\sigma_1 \leqslant \sigma_2 \leqslant .... \leqslant \sigma_k$ (at least one $<$ sign).

It is easier to deal with case A than with case B. A good a priori knowledge of the σ's will be necessary. The way in which σ can be estimated before the simultaneous experiments will start[1] has been described fully in chapter 4.

5.2 The standarddeviations σ_1, σ_2 .... σ_k are known exactly

5.2.1 The standarddeviations are equal; $\sigma_1 = \sigma_2 = ... \sigma_k$, say σ.

5.2.1.1 Duration criterion

It seems reasonable to continue the simultaneous measurements so long that the resulting $\widehat{\Delta A_j}$ value is a "sufficiently accurate" estimate of the unknown true value ΔA_j, for each of the k instruments.

Now the standarddeviation $\sigma(\widehat{\Delta A_j})$ of $\widehat{\Delta A_j}$ is a statistical measure of this accuracy. Consequently the requirement may be translated as follows: Let the series of simultaneous measurements be

---

Note [1]) It goes without saying that in all following considerations the hypothesis is made that each of the σ_j and each of the ΔA_j values is constant in time, and does not depend on the level of atmospheric radioactivity. Of course this hypothesis should be verified. Even if it is suspected not to be true on strictly theoretical grounds, this hypothesis may act as an working hypothesis provided that it is not obviously rejected by the numerical facts.

sufficiently long so that the ratio $\sigma(\widehat{\Delta A_j}) : |\widehat{\Delta A_j}|$ "sufficiently small" for each j (= 1, 2, 3 .... k). On first thought one would choose for this ratio a prescribed value (which itself is to be decided by the user). On second thought, however, it does not seem practical that both a value e.g. $\widehat{\Delta A_1} = 0.02$ and a value $\widehat{\Delta A_2} = 0.50$ should be calculated (measured) with the same percentual accuracy of for instance 5%, in which case the 95% reliability region for ΔA_1 would be 0.018 to 0.022 and for ΔA_2 0.045 to 0.055. It is obvious that very small ΔA values may possess a very large absolute inaccuracy; in other words a very broad 95% reliability region as 0.00 to 0.05 for the unknown ΔA=0.01 is permissible, but an analoguous percentual accuracy for a large systematical error, say 0.80, is not permissible. Consequently we should not fix the ratio $\sigma(\widehat{\Delta A_j}) : (\widehat{\Delta A_j})$ once and for all beforehand, but let the ratio depend on the value of ΔA_j itself. This seems impossible because the ΔA_j values are unknown; we wish to calculate them by means of the experiments.

Nevertheless we could continue as follows, in any case satisfying the requirement that <u>the percentual accuracy decreases with increasing absolute accuracy</u>.

How to do this? One of several possibilities is to require that a value $m\widehat{\Delta A}$ (with $m>1$) is measured with a percentual standard-deviation which is m times as small as that of $\widehat{\Delta A}$. Formulated mathematically

(63) $$\sigma(m\widehat{\Delta A}) : |m\widehat{\Delta A}| = \frac{1}{m}\,\sigma(\widehat{\Delta A}) : |\Delta A|$$ or shorter

(64) $$R(m\widehat{\Delta A}) = R(\widehat{\Delta A})/m$$

(65) when introducing $R(\widehat{\Delta A}) \equiv \frac{\sigma(\widehat{\Delta A})}{|\widehat{\Delta A}|}$ = relative accuracy of the systematical error.

Condition (64) means

(66) $$\sigma(\widehat{\Delta A_j}) = \text{constant, say } d$$

It is easily satisfied just because all σ's are equal.

(67) Since $\sigma(\widehat{\Delta A}) = \sigma\sqrt{\frac{k-1}{kn}}$ for each j,

from (66) follows:

(68) $$\boxed{n = \frac{k-1}{k} f^2}$$ with $f = \sigma/d$.

This is the duration criterion in a very simple form. For a sufficiently large number of instruments it only depends on the ratio of the standarddeviation of the random error and the accuracy of the systematical error.

It seems obvious that one wishes to know the systematical error much more accurate than the random error (to know the random error implies to know the standarddeviation of it probability distribution). In other words: we wish $d \ll \sigma$ or $f \equiv \sigma/d \gg 1$. How much f must surpass 1 depends on the preference of the user.

There are two possibilities (attention: the σ is given a priori):

I. d is fixed beforehand, then f depends on σ; e.g. $d = 0.01$; $f = 100\sigma$

II. f is fixed beforehand, then d depends on σ; e.g. $f = 5$; $d = 0.2\,\sigma$

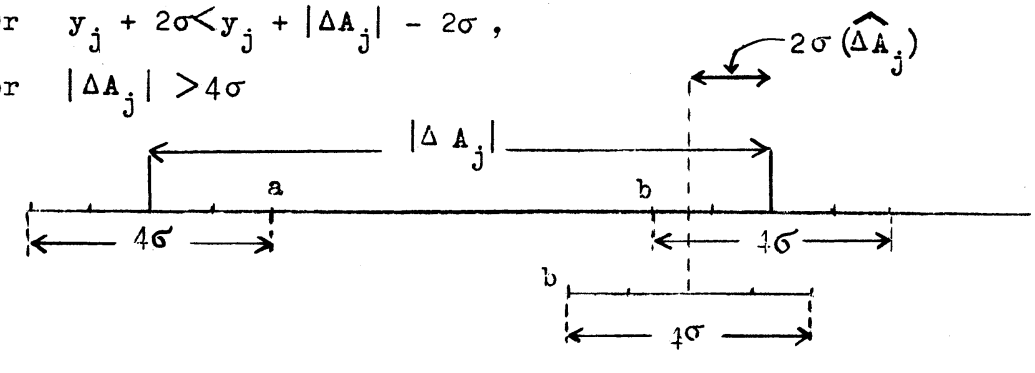
5.2.1.2 Selection criterion

It seems reasonable to correct for s.e. only if the difference between the corrected value and the not corrected value is "sufficiently high". In defining what "sufficiently high" means, one should take into account the existence of r.e.

In the following the value $|\Delta A_j|$ is chosen so large that the whole 95% confidence region, laid around $y_j + |\Delta A_j|$, lies outside the 0.95 region around the not corrected value y_j. We propose to speak then of a "sufficiently large" difference between the not corrected value y_j and the corrected value $y_j + |\Delta A_j|$. We therefore require that point b in the graph underneath representing $y_j + |\Delta A_j| - 2\sigma$ is always to the right of point a, representing $y_j + 2\sigma$.

(69) Or $y_j + 2\sigma < y_j + |\Delta A_j| - 2\sigma$,

(70) or $|\Delta A_j| > 4\sigma$



In this reasoning the true ΔA_j is used. However, this true value is unknown; the estimate is $\widehat{\Delta A_j}$. For $\sigma(\widehat{\Delta A_j})$ see (67).

Consequently it should be so arranged that point b' is to the right of point a, or formulated mathematically:

(71) $y_j + 2\sigma < y_j + \{|\widehat{\Delta A}_j| - 2\sigma(\widehat{\Delta A}_j)\} - 2\sigma$, which gives

(72) $\boxed{\widehat{\Delta A}_j > 4\sigma c}$ with $c = 1 + \frac{1}{2}\sqrt{\frac{k-1}{kn}} > 1$ and

$c = 1 + \frac{1}{2}\sqrt{\frac{1}{n}}$ for large k.

In this way even for the lower boundary of the reliability region of ΔA_j the inequality (69) applies.

Next we substitute the n value for (68) into (72) and obtain

(73) $\boxed{|\widehat{\Delta A}_j| > 4\,\sigma(1 + \frac{1}{2f})}$

In this way, provided that $\sigma_1, \sigma_2, \ldots \sigma_k$ are equal and given beforehand, the duration and selection criterium have been derived, see (68) and (73).

Both depend on the same, more or less arbitrary, parameter f.

5.2.2 The standarddeviation $\sigma_1, \sigma_2, \ldots \sigma_k$ are not all equal

5.2.2.1 Duration criterion

Let be $\sigma_1 \leqslant \sigma_2 \leqslant \sigma_3 \ldots \leqslant \sigma_k$ (at least one $<$ sign).

(74) Since $\sigma(\Delta A_j) = \frac{1}{\sqrt{nk}}\sqrt{\overline{\sigma^2} + (k-2)\,\sigma_j^2}$

where $\bar{\sigma} = \frac{1}{k}\sum\sigma_j$ and $\overline{\sigma^2} = \frac{1}{k}\sum\sigma_j^2$

we also have

(75) $\sigma\,\widehat{\Delta A}_1) \leqslant (\widehat{\Delta A}_2) \ldots\ldots \leqslant \sigma(\widehat{\Delta A}_k)$.

Here we particulary should bear in mind, that generally σ is not related to Δ A, i.e. that any dependence between systematical and random errors is absent. It may be that $\Delta A_1 = 10\,\Delta A_5$ and $\sigma_1 = \frac{1}{2}\sigma_5$ and so on. Nevertheless we wish to derive an expression analoguous to (68) without any a priori knowledge about the ΔA values.

Referring to section 5.2.1 it is now impossible to require that (64) is fulfilled just because (66) is not true now; reference is made to the essential inequality of the $\widehat{\Delta A}_j$ values in (74) and (75). We propose, however, to maintain the "requirement" $\sigma(\widehat{\Delta A}) = d$ "with a

(76) small changement", and we require that $\sigma(\widehat{\Delta A}_j) \leqslant$ constant $= d$ for $j = 1, 2 \ldots k$.

(77) Introducing $F_j \equiv [\{\overline{\sigma^2} + (k-2)\,\sigma_1^2\} : (k-1)]^{\frac{1}{2}}$

(78) so that $\sigma(\widehat{\Delta A}_j) = F_j \sqrt{\frac{k-1}{kn}}$, $\quad j = 1, 2 \ldots k$

(79) with $F_1 \leqslant F_2 \leqslant F_3 \ldots \leqslant F_k$, then

from (78) and (76) follows

(80) $n \geqslant \frac{1}{d^2}\;\frac{k-1}{k}\;F_j^2 \qquad j = 1, 2 \ldots k.$

With regard to (79) the minimum duration should be

(81) $$\boxed{n = \frac{1}{d^2}\;\frac{k-1}{k}\;F_k^2}$$

Written in a different way

(82) $$\boxed{n = \frac{k-1}{k}\,f^2 T^2 = \frac{k-1}{k}\,(f^*)^2}$$ where

$$T = F_k/\overline{\sigma}\;;\quad f^* = f.T\;;\quad f = \overline{\sigma}/d$$

As soon as the given σ's are unequal (in the sense that $\sigma_1 = \sigma_2 = \ldots = \sigma_k$ does not hold true), then $T > 1$ and $f^* > f$ 1).

The proof of the statement $T > 1$ for $\sigma_1 < \sigma_2 \ldots < \sigma_k$ runs as follows:

$$T = \{\overline{\sigma^2} + (k-2)\,\sigma_k^2\} : (k-1)(\overline{\sigma})^2 > 1 \quad \text{would imply}$$

$$\overline{\sigma^2} + (k-2)\,\sigma_k^2 > (k-1)(\overline{\sigma})^2 = (k-2)(\overline{\sigma})^2 + (\overline{\sigma})^2 .$$

Note 1) The following theorem is of importance. Given $x_1, x_2 \ldots x_k$ not all being equal and defining

$\varphi(t) = \left[\frac{1}{k}\sum_1^k x_i^t\right]^{1/t}$, then $\varphi(1) < \varphi(2) < \varphi(3)$ etc.

Or

$\overline{x} < (\overline{x^2})^{\frac{1}{2}} < (\overline{x^3})^{\frac{1}{3}}$ etc.

Now certainly $\overline{\sigma^2} > (\bar{\sigma})^2$
and certainly $(k-2)\sigma_k^2 > (k-2)(\bar{\sigma})^2$ because σ_k is the largest of all σ_j; thus we have proved what was to be proved.

Of course (76) must reappear as soon as the value of n from (81) is substituted in the expression (78) for $\sigma(\widehat{\Delta A}_j)$. We obtain:

(83) $$\sigma(\widehat{\Delta A}_j) = d. \quad {}^{F_j}/_{F_k} \; ; \; j = 1, 2 \dots k$$

and indeed (76) reappears as a consequence of (79). From j=1 to k (with a not decreasing order of σs) the $\sigma(\widehat{\Delta A}_j)$ values increase to d. Attention: the smallest $\sigma(\widehat{\Delta A})$ (here j = 1) may belong to the instrument with the smallest ΔA value as well as to the apparatus with the largest one.

(84) In (82) we also read $n = \frac{k-1}{k} \frac{F^2k}{d^2} = \frac{(k-1)\{\overline{\sigma^2} + (k-2)\sigma_k^2\}/(k-1)}{k \qquad d^2}$

Now compare with (68), where $\sigma_1 = \sigma_2 = \dots \sigma_k$, say σ. We see σ^2 is replaced by $\{\overline{\sigma^2} + (k-2)\sigma_k^2\} : (k-1)$, which is always larger than $(\bar{\sigma})^2$, as soon as $\sigma_1 = \sigma_2 \dots = \sigma_k$ does not hold and σ_k is the largest of all σ's. This fact suggests a second "solution" of the problem: maintain the expression (68) but replace the σ by $\bar{\sigma}$. ($\leq\sqrt{\overline{\sigma^2}}$). Then the duration criterium would become:

(84a) $$n = \frac{k-1}{k}\left(\frac{\bar{\sigma}}{d}\right)^2 = \frac{k-1}{d} f^2 \quad \text{with } f = \bar{\sigma}/d$$

It is difficult to conclude whether (82) or (84a) is "better", because the definition of "better" in this connection has not been formulated.

Once more stress is laid on the differences between the consequences when starting from (82) or from (84a).

In the first case all $\sigma(\widehat{\Delta A}_j) \leqslant d$, whatever may be the j.

In the second case $\sigma(\widehat{\Delta A}_j) = d\, F_j/\bar{\sigma}$, and this value grows, going with j from 1 to k, from a value smaller than d to a value larger than d. It seems preferable to require that all $\sigma(\widehat{\Delta A}_j)$ values should be smaller than d.

It is obvious that any inequality of σ's increases the minimum duration n of the Grandjean-experiments. Moreover, it can easily be seen that the "more unequal these σ's", the larger the increase of n. Instead of proving this mathematically[1], we only give some numerical examples.

Numerical examples:

Let us fix d as 0.05. We will compare the following cases.

1) $\sigma_1 = \sigma_2 = \sigma_3$ ($=\sigma$) = 0.30. Then (68) gives $n_1 = \frac{2}{3}\left(\frac{0.30}{0.05}\right)^2 =$ 24 days.

2) $\sigma_1 = 0.10$ $\sigma_2 = 0.30$ $\sigma_3 = 0.50$, so that $\bar{\sigma} = 0.30$; $\overline{\sigma^2} = 0.117$; $F_3 = 0.429$. Then (82) gives $n_2 = \frac{2}{3}\left(\frac{0.30}{0.05}\right)^2 \left(\frac{0.429}{0.30}\right)^2 = 24 \times 2.0 =$ 48 days.

3) $\sigma_1 = \sigma_2 = 0.10$ $\sigma_3 = 0.70$, so that $\bar{\sigma} = 0.30$; $\overline{\sigma^2} = 0.170$ and $F_3 = 0.575$. Then (82) gives $n_3 = \frac{2}{3}\left(\frac{0.30}{0.05}\right)^2 \left(\frac{0.575}{0.30}\right)^2 = 24 \times 3.69 =$ 84 days.

These three cases possess the same $\bar{\sigma} = 0.30$.
The result is $n_3 > n_2 > n_1$, which illustrates the general effect, that the larger the inequality between the σ's, the larger the minimum duration of the Grandjean-experiments. Roughly spoken the increase depends on the ratio of the largest σ_k of all σ's to the overall mean value $\bar{\sigma}$.

5.2.2.2. Selection criterion

Suppose that the value of n is calculated by means of (82) (the σ's being given !) and that during n days simultaneous measurements have been made. Next the values of $\widehat{\Delta A_j}$ are computed and again the question arises which instrument should be corrected for systematical errors.

Referring to (71), where now the $\sigma(\widehat{\Delta A_j})$ values are different for different j values (see (74)), we obtain:

$$|\widehat{\Delta A_j}| > 4\sigma_j + 2\,F_j\sqrt{\frac{k-1}{kn}} \quad ; \quad j = 1, 2 \ldots k \tag{85}$$

Note[1] An exact mathematical proof would require the formulation of the popular wording "the more unequal". It is too academical to elaborate on this aspect.

The value of n is computed from (82), so that for each j the right hand side (the selection level) can be calculated. This gives k levels. As soon as the measured $|\widehat{\Delta A}_j|$ surpasses its corresponding level (and only then), the measurements with this instrument j deserve correction for systematical errors.
When (82) is substituted in (83) we obtain:

(85a) $$\boxed{|\widehat{\Delta A}_j| > 4\bar{\sigma}\left(q_j + \frac{2}{f}\, t_j\right)}$$, where q_j stands for $\sigma_j/\bar{\sigma}$, t_j for F_j/F_k and f for $\bar{\sigma}/$ d.

It is interesting to compare (85a) with (73). Whereas in (73) there is only one selection level, the value of which is a factor $(1 + \frac{1}{2f})$ as large as the σ, now, in the case that $\sigma_1 = \sigma_2 = \ldots \sigma_k$ does not hold, there are as many selection levels as instruments. Even if σ is replaced by $\bar{\sigma}$, all these levels need not be larger than $4\bar{\sigma}$, in consequence of the fact that q_j grows from smaller to larger than one and t_j grows to one, for growing j (1, 2, . . . k), so that certainly for j = k the selection level is larger than $4\bar{\sigma}$, but for small j values (j = 1) the level may be smaller as well as larger than $4\bar{\sigma}$, according to the value of d.

<u>Numerical example:</u>
Let be $k = 4$; $\sigma_1 = 0.05$; $\sigma_2 = 0.10$; $\sigma_3 = 0.15$; $\sigma_4 = 0.20$ so that $\bar{\sigma} = 0.125$; $\overline{\sigma^2} = 0.01875$ and $F_4 = 0.1810$. Let be chosen d = 0.05.
Then the minimum duration, see (82), becomes $\frac{3}{4}\left(\frac{0.125}{0.05}\right)^2 \left(\frac{0.1810}{0.105}\right)^2 =$ $4.7 \times 2.1 = 10$ days (this duration is so small because of the large d value).
See the following table for more details.

Instrument N=j	1	2	3	4	remark
let be given σ	0.05	0.10	0.15	0.20	
F is	0.093	0.113	0.169	0.181	see (77)
q	0.4	0.8	1.2	1.6	
t	0.52	0.62	0.94	1.00	
let be measured after n days $\widehat{\Delta A}$	0.70	-0.80	0.60	-0.50	sum=0
the selection level	0.41	0.65	0.98	1.20	see (83)
instr. needs correction for s.e.	yes	yes	no	no	
because	0.70>0.41	0.80>0.65	0.60<0.98	0.50<1.20	

5.3 <u>The standard deviations $\sigma_1, \sigma_2 \ldots \sigma_k$ are not known a priori.</u>

In this case the Grandjean experiments are also used to calculate the σ's. This complicates the analysis. We again distinguish between two cases. First case: $\sigma_1 = \sigma_2 = \ldots \sigma_k$, say σ (the equality is known, but the value of σ is unknown). Second case: the σ's are not all equal.

5.3.1 <u>Standard deviations are equal; $\sigma_1 = \sigma_2 = \ldots \sigma_k$, say σ.</u>

5.3.1.1 <u>Duration criterion</u>

Referring to (67) we should take into consideration that σ is not known. The n simultaneous measurements furnish an estimate $\hat{\sigma}$, which possesses an accuracy given by

(86) $$\sigma(\hat{\sigma}) = \frac{\sigma}{\sqrt{2\,n\,(k-1)}} \quad \text{and measured as} \quad \hat{\sigma}(\hat{\sigma}) = \frac{\hat{\sigma}}{\sqrt{2\,n\,(k-1)}}$$

This means that for each of the k instruments the 95% reliability region for the unknown σ lies between $\hat{\sigma} - 2\hat{\sigma}(\hat{\sigma})$ and $\hat{\sigma} + 2\hat{\sigma}(\hat{\sigma})$. In case σ were exactly known, the requirement would be written as:

$$\sigma(\widehat{\Delta A}) = \sigma\sqrt{\frac{k-1}{kn}} = d$$, see (66) and (67), but now we write analogously:

(87) $$\{\hat{\sigma} + 2\hat{\sigma}(\hat{\sigma})\}\sqrt{\frac{k-1}{kn}} = d$$ from which follows

(88) $$\boxed{n = \frac{k-1}{k} f^2 \quad (1 + \sqrt{\frac{2}{n\ (k-1)}}\ )^2} \qquad \text{with } f = \hat{\sigma}/d$$

This equality contains n implicitely. We wish to find n as an expression only in k and f and, if possible, in the form

$$n = \frac{k-1}{k} (f^*)^2 \quad \text{with } f^* = f.g.$$

where g is an function of k and f only.

The equation is of the second degree in $\sqrt{n}$. An approximative solution (see Appendix) is

(88a)

$n \cong \frac{k-1}{k} (f^*)^2$	$f^* = f.g.$	$g = 1 + \frac{1}{f}\sqrt{\frac{2}{k}} > 1)$

Comparing (88a) with (68) we see that f has been increased to f^*, but the increase is the smaller the larger f and (or) the larger k. The fact that the absence of any knowledge about σ has led to an increase of the minimum duration is easily understood, because the same data are used for two purposes, namely for the calculation of the systematical errors as well as for the calculation of the random errors.

<u>Numerical example:</u>

$\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \sigma_5 = \sigma_6$, say σ. Take f = 5.

1) σ known. Minimum duration $n = \frac{5}{6} \cdot 5^2 = 21$ days.

2) σ unknown. Minimum duration $n = \frac{5}{6} \cdot 5^2 (1 + \frac{1}{5}\sqrt{\frac{2}{6}})^2 = 26$ days.

5.3.1.2 Selection criterion

Reference is made to the graph under 5.2.1.2. We again require that the whole 95% confidence interval for the unknown true value which is measured as y, lies outside the 0.95 confidence region of the corrected value, but we should now take into account both the uncertainty in σ and in ΔA. Consequently we should $y_j+2\sigma$ (see (71)) replace by $y_j + 2\{\hat{\sigma} + 2\hat{\sigma}(\hat{\sigma})\}$, whereas $\sigma\sqrt{\frac{k-1}{kn}}$ should be replaced by $\{\hat{\sigma} + 2\hat{\sigma}(\hat{\sigma})\}\sqrt{\frac{k-1}{kn}}$.
As a result, (71) becomes:

(89) $$y_j + 2\{\hat{\sigma} + 2\hat{\sigma}(\hat{\sigma})\} < y_j + [\,|\widehat{\Delta A_j}| - 2\{\hat{\sigma} + 2\hat{\sigma}(\hat{\sigma})\}\sqrt{\frac{k-1}{kn}}\,] - 2\{\hat{\sigma} + 2\hat{\sigma}(\hat{\sigma})\}$$

with

(90) $$\hat{\sigma}(\hat{\sigma}) = \frac{\hat{\sigma}}{\sqrt{2\ n\ (k-1)}}$$

It then follows that

(91) $$\boxed{|\widehat{\Delta A}_j| > 4\hat{\sigma}c.c'}$$ with $c = 1 + \frac{1}{2}\sqrt{\frac{k-1}{kn}} > 1$; if $k \to \infty$ then $c \to 1 + \frac{1}{2}\sqrt{\frac{1}{n}}$

$$c' = 1 + \frac{\sqrt{2k}}{\sqrt{n(k-1)}} > 1; \text{ if } k \to \infty \text{ then } c' \to 1 + 2\sqrt{\frac{1}{nk}}$$

The right hand side of (91) represents the selection level. Of course, since σ's are equal, although unknown, there is only one selection level. As soon as the measured $|\widehat{\Delta A}_j|$ surpasses this level, instrument j needs correction for the s.e.

The absence of knowledge about σ has increased the selection level; compare (91) with (72). The influence of the factor c' is the smaller, the larger the number of participating instruments. Next we substitute n from (88a) in (91) and we obtain:

(91a) $$\boxed{|\widehat{\Delta A}_j| > 4\hat{\sigma}\left(1 + \frac{1}{2\, f^*}\right)\left(1 + \frac{\sqrt{2}}{k-1}\,\frac{1}{f^*}\right)}$$ with $f^* = f.g.$

We see that the correction level decreases both with increasing f and increasing k, but remains larger than $4\hat{\sigma}$.

5.3.2 Standard deviations are not all equal; $\sigma_1 \leqslant \sigma_2 \ldots \leqslant \sigma_k$.

5.3.2.1 Duration criterion

It is easily understood that as in section 5.2.2 the degree of inequality of the σ's must affect the minimum duration. In the former case the σ's are known, so that the inequality of the σ's enters into the n expressions by means of F_k, see (81) and (77). If any knowledge even with regard to the interdependence of the σ's is absent, it seems impossible to attack the problem. Now in practice we usually posses a rough idea about the mutual dependence between σ values, although the σvalues themselves are unknown. This idea may be based for instance on the sampled volumes of air. Such a rough idea should be applied as follows.

<u>Numerical example:</u>

Suppose $\sigma_1 : \sigma_2 : \ldots \sigma_6$ is approximately 1:1:2:2:2:4, but σ_1 is unknown. Then $\bar{\sigma} = 2\sigma_1$; $\overline{\sigma^2} = 5\sigma_1^2$ and, see (77),

$F_6 = \sqrt{\frac{1}{5}(5\sigma_1^2 + 4 \times 16\sigma_1^2)} = 3.72\sigma_1$, so that $T = F_6/\bar{\sigma} = 1.86$ and, using (82), $n = \frac{5}{6} f.^2\ 3.5$

If $f = 5$, then $n = 73$ days. In this way an estimate of n has been made; some knowledge on an interdependence of the σ values was necessary. Sometimes this knowledge is very rough, hence also n is estimated very roughly, but even a rough estimate is always better than no estimate at all. In case the exact value of the interdependence of the σ values is considered of particular theoretical or practical value, it is of course desirable to test whether the assumption made on this value is valid, as soon as the σ values have been estimated.

5.3.2.2 Selection criterion

It seems impossible to continue our line of thought in the most general way.

Analogously to (71), but now giving indices j to σ (since the σ's, are not all equal), and at the same time each σ_j replacing by its upper reliability limit $\widehat{\sigma_j} + 2\widehat{\sigma}(\widehat{\sigma_j})$, say τ_j, with

(92) $\widehat{\sigma}(\widehat{\sigma_j}) = \widehat{\sigma_j} / \sqrt{2n(k-1)}$, we obtain the requirement

(93) $y_j + 2\tau_j < y_j + \left[|\widehat{\Delta A_j}| + 2F_j(\tau)\sqrt{\frac{k-1}{kn}} \right] - 2\tau_j$

where

(94) $F_j(\tau) = \sqrt{\frac{1}{k}\sum\{\widehat{\sigma_j} + 2\widehat{\sigma}(\widehat{\sigma_j})\}^2 + (k-2)\{\widehat{\sigma_j} + 2\widehat{\sigma}(\widehat{\sigma})\}^2} : \sqrt{k-1} = \sqrt{\frac{\overline{\tau^2} + (k-2)\tau_j^2}{k-1}}$

analogously to (77), when replacing each σ_j by τ_j

The selection criterium becomes

(95) $$\boxed{|\widehat{\Delta A_j}| > 4\tau_j + 2F_j(\tau)\sqrt{\frac{k-1}{kn}} = 4\tau_j + 2\frac{F_j(\tau)}{F_k}\frac{\bar{\sigma}}{f}}$$ when substituting

n from (82). The largest F (that is F_k with the largest σ_k) can be expressed in σ_1 as a result of the assumption that we know exactly or approximately the interrelation between $\sigma_1, \sigma_2, \ldots \sigma_k$, without knowing each σ separately. That implies that in $\sigma_1 : \sigma_2 : \ldots \sigma_k =$

$1 : a_2 : a_3 \dots a_k$ we know the a's exactly or approximately. In the same way $\bar{\sigma}$ is expressed in σ_1 and hence $F_k : \bar{\sigma}$ does not certain σ_1.
Attention: it is sufficient to know exactly the interrelation between the σ's, in order to find the minimum value of n.

Numerical example 1:

Let be k = 3, the σ's are unknown, but we expect $\sigma_1 : \sigma_2 : \sigma_3 = 1:3:11$.
Put $\sigma_2 = 3\,\sigma_1$ and $\sigma_3 = 11\,\sigma_1$, so that $\bar{\sigma} = 5\,\sigma_1$ and $\overline{\sigma^2} = 43.7\,\sigma_1^2$.
In order to estimate the minimum duration n we compute, (see (77)) $F_3 = 9.1\,\sigma_1$ and $T = 1.82$.
If f = 5 is chosen a priori then (see (82)) $n = \frac{2}{3} \cdot f^2 . T^2 = 55$ days.
After having made 55 simultaneous measurements the estimates $\hat{\sigma}_1$, $\hat{\sigma}_2$ and $\hat{\sigma}_3$ prove to be e.g. 0.01, 0.04 and 0.13, so that, according to (51) $\hat{\sigma}(\hat{\sigma}_1) = 0.0007$, $\hat{\sigma}(\hat{\sigma}_2) = 0.0014$ and $\hat{\sigma}(\hat{\sigma}_3) = 0.0088$, so that $\tau_1 = 0.011$, $\tau_2 = 0.023$, $\tau_3 = 0.148$.
With these τ values we obtain $F_1 = 0.062$, $F_2 = 0.063$, $F_3 = 0.121$ (see (94)). In this way the selection levels (see the righthand sides (95)), can be calculated. The following table contains details.

Instr. No.	1	2	3	remark
Suppose σ_j =	σ_1	$3\sigma_1$	$11\sigma_1$	
Let be measured $\hat{\sigma}$	0.01	0.02	0.13	
Let be measured $\widehat{\Delta A}$	-0.60	-0.08	+0.68	sum 0
Computed $\hat{\tau}$	0.011	0.023	0.148	$\hat{\sigma} + 2\hat{\sigma}(\hat{\sigma})$
Computed $F(\tau)$	0.062	0.063	0.121	see (94)
Selection level	0.058	0.106	0.619	see (95)
Needs correction for s.e.	yes	no	yes	
Because	0.60>0.058	0.08<0.106	0.68>0.619	

In this example the interdependence turned out to be $\hat{\sigma}_1 : \hat{\sigma}_2 : \hat{\sigma}_3 = 1 : 2 : 13$ (which is an inaccurate estimate of the true value). We assumed 1 : 3 : 11 before the experiments were started. With 1 : 2 : 13 the minimum duration would have been [with $\bar{\sigma} = 5\ 1/3\ \sigma_1$; $\overline{\sigma^2} = 58.0\ \sigma_1^2$; $F_3 = 10.6\ \sigma_1$; $T = 2.07$), see (82)] n = 72 days (larger than 55).

Numerical example 2:

Two cases I, II,

I. k = 3; all σ's are known, e.g. $\sigma_1 = 0.05$ $\sigma_2 = 0.=0$ $\sigma_3 = 0.15$

II. k = 3; σ's unknown, but it is assumed that $\sigma_1 : \sigma_2 : \sigma_3 = 1 : (2 \pm \frac{1}{2}) : (3 \pm \frac{1}{2})$. Extreme possibilities $1 : 1\frac{1}{2} : 3\frac{1}{2}$ and $1 : 2\frac{1}{2} : 2\frac{1}{2}$

Case I. $n = \frac{k-1}{k} f^2 T^2$; choose f = 5, hence $n = \frac{2}{3} \times 25\ T^2$

$\bar{\sigma} = 0.10$; $\overline{\sigma^2} = 0.0117 \therefore F_3^2 = \frac{0.0117 + 0.0225}{2} = 0.0171 \therefore F_3 = 0.130$

$T = F_3 / \bar{\sigma} = 1.3$ and n = 28 days

Case II.

a) $\sigma_1 : \sigma_2 : \sigma_3 = 1 : 1\frac{1}{2} : 3\frac{1}{2} \therefore \bar{\sigma} = 2\sigma_1$; $\overline{\sigma^2} = 5.17\sigma_1^2 \therefore F_3^2 = \frac{5.17 + 49/4}{2} \sigma_1^2 =$

$8.17\ \sigma_1^2$; $F_3 = 2.96\sigma_1 \therefore T = F_3 : \bar{\sigma} = 1.48$

Hence $n = \frac{2}{3} \times 25 \times 1.48^2 = 36$ days.

b) $\sigma_1 : \sigma_2 : \sigma_3 = 1 : 2\frac{1}{2} : 2\frac{1}{2} \therefore \bar{\sigma} = 2\sigma_1$; $\overline{\sigma^2} = 4.42\ \sigma_1^2 \therefore F_3^2 =$

$\frac{4.42 + 25/4}{2} \sigma_1^2 = 5.34\ \sigma_1^2 \therefore F_3 = 2.31\ \sigma_1 \therefore T = F_3 : \sigma_1 = 1.15$

Hence $n = \frac{2}{3} \times 25 \times 1.15^2 = 22$ days.

The uncertainty as to the true interrelation $\sigma_1 : \sigma_2 : \sigma_3$ leads to a fairly large uncertainty as to the minimum duration n, namely 22 to 36 days. Preferably one should choose 36 days.

5.4 Appendix

Reference is made to (88). We wish to solve the unknown n. Putting $p = f\sqrt{\frac{k-1}{k}}$, $q = \sqrt{\frac{2}{k-1}}$ and $x = \sqrt{n}$ we get:

$(x + q)\ p = x^2$ or $x^2 - p\,x - p\,q = 0$; The solution becomes

$x_{1,2} = \frac{1}{2} p \pm \sqrt{\frac{1}{4} p^2 + pq}$

Here only the positive sign can be used

$x = \sqrt{n} = \frac{1}{2} p [1 + \sqrt{1 + (4\sqrt{2k}) / f\ (k-1)}]$

$n = \frac{1}{4} p^2\ [1 + \sqrt{\ }]^2$ or

$$n = \frac{k-1}{k} f^2 \left[\frac{1 + \sqrt{1 + (4\sqrt{2k}) / f(k-1)}}{2} \right]^2$$

Now for sufficiently large f and (or) k, we have approximately

$[----] \cong \frac{1}{2} \{1 + \sqrt{1 + 4\sqrt{2} / f\sqrt{k}}\} \cong \frac{1}{2}\{1 + 1 + 2\sqrt{2}/f\sqrt{k} = 1 + \sqrt{2}/f\sqrt{k} > 1$

and $n \cong \frac{k-1}{k} f^2 (1 + \sqrt{2}/f\sqrt{k})^2 \cong \frac{k-1}{k} f^2 (1 + \frac{2\sqrt{2}}{f\sqrt{k}}) > \frac{k-1}{k} f^2$

or written in a different way

$n \cong \frac{k-1}{k} (f^*)^2$ with $f^* = f \cdot g$ and $g = 1 + \frac{1}{f}\sqrt{\frac{2}{k}}$ $f = \bar{\sigma} / d$

In this way (88a) appears.

6. Numerical examples

6.1 Numerical example 17/18 September 1961, De Bilt, K.N.M.I.

From 9 h 17-9-1961 to 9 h 18-9-1961 a volume of air 34.4 m^3 was sucked through the filter at De Bilt. The measurements of the filter-activity and other qunatities are assembled in the following table.

date	b imp/min	N imp	k imp/min	D min	f=b/k	δ%	T day	t = τ + T, τ=5	τ=16
Sept. 17/18									
18	445.8	4000	17.2	8.62	26.0	1.62	0.2	5.2	16.2
19	360.4	4000	20.4	16.46	17.7	1.67	1	6	17
20	305.8	4000	20.2	12.28	15.2	1.69	2	7	18
21	258.6	4000	20.3	14.30	12.8	1.71	3	8	19
22	222.6	4000	20.4	16.46	10.9	1.73	4	9	20
23	189.1	4000	17.0	19.40	11.1	1.72	5	10	21
25	155.5	4000	17.3	23.17	9.0	1.76	7	12	23
26	144.7	4000	17.1	24.74	8.5	1.77	8	13	24
27	133.7	4000	17.6	26.42	7.6	1.79	9	14	25
28	116.2	4000	17.7	29.85	6.6	1.83	10	15	26
29	112.7	2000	17.5	15.37	6.4	2.58	11	16	27
30	111.2	4000	17.8	31.00	6.2	1.83	12	17	28
Oct. 4	87.8	4000	17.5	37.98	5.0	1.89	16	21	32
6	83.1	2000	20.6	19.29	4.0	2.79	18	23	34
7	77.9	4000	20.8	40.56	3.7	2.00	19	24	35
13	59.6	4000	20.6	49.86	2.9	2.13	25	30	41
17	49.5	4000	17.7	59.70	2.8	2.13	29	34	45
25	39.1	4000	18.0	59.32	2.2	2.31	41	46	57
27	38.6	4000	21.3	66.69	1.8	2.46	43	48	59

<u>Explanation</u>: b = mean number of impulses per minute of the filter-activity corrected for the background k; b is based upon a total N number of impulses (pure activity + background) in D minutes.
δ = percentual accuracy of b, computed as $\delta = \frac{1 + 1/f}{\sqrt{N}}$; f = b/k
τ = age of the artificial atmospheric radioactivity = number of days between the day of explosion of the bomb and the day of sampling the air.
T = number of days between the day of sampling air and the day of counting the exposed filter.
$\tau' = \tau + T$ = time in days between the explosion day and the day of counting the filteractivity.

Since the sampling took place from 9 to 9 h and all countings of the filteractivity took place in the morning, say between 9 and 11 h, each T equals the duration between the end of the sampling and the time of counting the filteractivity, expressed in a whole number of days. The very first counting always takes place at about 5 hours after the end of the sampling in order to obtain a rough idea of the activity (5 h = $\frac{5}{24}$ = 0.2 days).

6.1.1 <u>Using only two measurements</u>

The zero day's activity (extrapolation) as shown in the monthly published table is based on the countings of the 4^{th} and 11^{th} day (T_1 = 4; T_2 = 11). The table, given above, says: b_1 = 222.6; k_1=20.4; $f_1 = b_1/k_1$ = 10.9; N_1 = 4000 so that δ_1 = 1.73%; likewise b_2 = 112.7; k_2 = 17.5; $f_2 = b_2/k_2$ = 6.4; N_2 = 2000 and δ_2 = 2.58% (weaker activities are measured less accurate). Hence $\delta_* = \sqrt{\delta_1^2 + \delta_2^2}$ = 2.58% = relative accuracy of $q = b_2/b_1$ = 0.505.

Next nomogram D with q = 0.505 and $m = T_2/T_1$ = 2.75 ($\tilde{m}$ = 1.58), gives c = 2.05 and $\tilde{c}$ = 1.8. Hence the zero day's value b_o becomes $b_o = c\, b_1$ = 457 imp/min., with an accuracy $\delta_o = \tilde{c}.\tilde{m}.{}_*$ = 8.9%. Consequently for the specific artificial atmospheric radioactivity in the sampled volume of 34.4 m^3 we find the point estimate $\hat{I} = \frac{b_o}{34.4}$ 1.5= 19.9 pC/m^3. The 95% confidence interval (broadth $4\delta_o\hat{I}$) becomes <u>16.4<I<23.5</u> pC/m^3.

Next we compute the age and its accuracy. The $q = 0.505$ gives $\tilde{q} = 1.93$ (see nomogram D). Further $q = 0.488$ (that is $1/c$) and $c = \infty$ give $m_\tau = 1.78$, so that
$\hat{\tau} = T_1 : (m_\tau - 1) = \underline{5.1}$ days, the point estimate of the age τ. Next the accuracy $\delta_\tau = \tilde{c}\,\tilde{q}\,\delta_* = 10.7\%$, so that the 0.95 reliability interval for the unknown τ becomes $\underline{4.0 < \tau < 6.2}$ days.

<u>Remark:</u> the nomogram D is based on the law $\beta = \text{const}\,(\tau + T)^{-1.20}$; it is therefore assumed that the value of the exponent is exactly known.

The considerations given above refer to the random counting error, but there exists also a random sampling error. The standarddeviations are σ_B resp. σ_e, whereas the total error possesses a standarddeviation $\sigma_E = \sqrt{\sigma_e^2 + \sigma_B^2}$. Now duplo measurements at De Bilt (July '58 - July '59) of the atmospheric radioactivity around a mean level 2.9 pC/m^3 (range 0.2 to 10.0) have learnt that $\sigma_e \cong 0.18$ and that σ_e does not depend significantly of the level within this range. Assuming that σ_e possesses the same value on higher levels, we obtain $\sigma_E = \sqrt{0.18^2 + 0.089^2} = 0.20$. Hence, taking into account both the spatial sampling and the counting errors, the 0.95 reliability region for I becomes $\underline{11.9 < I < 27.9}$ pC/m^3, with a point estimate 19.9 pC/m^3. The large breadth of this interval (in other words: the large unreliability of the figure 19.9) certainly will surprise the reader. Of course the question as to whether such a result is almost without value, is a matter of the user. Since $\sigma_e \gg \sigma_B$, a more accurate counting of the filteractivities does not decrease σ_E markedly.

6.1.2 <u>Using all measurements, two different graphs</u>

We have at our disposal 18 countings of the filteractivity (see the table shown above for T = 1, 2 etc.). Let us plot two different graphs. In the first one we plot b against $\tau + T$, for $\tau = 5$ resp. 16, in a double logarithmic scale; in the second one we plot $b^{-1/1.20}$ against T in a double linear scale.

a) <u>b against $\tau + T$</u> in a double linear scale.
In case the atmospheric radioactivity belongs to only one specified

bomb (τ given) then, because of the law β = const $(\tau + T)^{-p}$, the points log β and log (τ+ T) lie on one straight line. We have plotted the couples b, T + τ for two τ values : 5 and 16 days (see fig.1). It is obvious that the points in graph A are better linearly arranged than in graph B, although there seem to be two groups of points, that is the points up to T = 12 seem to lie better linearly than the points for higher values of T. The points T = 4 and T = 11 (used in the foregoing section) have been connected by a straight line, which (of course only by chance) seems to be also the best fitting straight line through all points. The slope of this line gives a value $\hat{p}$ = 1.20 .

The line cuts the vertical b axis in the zero day's value 446, which value of course should be the same as the value 457 found by means of nomogram D, see under section 6.1.1. The difference is only a matter of accuracy of reading the nomogram and of plotting graph A. We have indicated in the same graph the 95% confidence interval for β_0.

Although the points in graph B do not lie linearly, one could draw the best fitting straight line. If this is done only visually, this straight line gives $\hat{p}$ = 2.05; the fact that this value differs so strongly from 1.20, illustrates that the activity cannot possess an age of about 16 days. Moreover, attacking the question purely synoptically, it is quite probable that the activity which was sampled 17/18 September originated from the explosion on 12 September.

b) <u>$b^{-1/1.20}$ against T</u> in a double linear scale.

If 1.20 would be the correct value of p we could plot the points $b^{-1/1.20}$ against T. Then the values of b_0 and $\hat{\tau}$ are unknown (see graphs C_1, C_2 in fig. 2). Again there seem to be two groups of points, the left hand group up to T = 12 and the right hand side group for larger values of T. One could draw two straight lines, the first one C_1, fitting the left hand group (nearly the straight line passing through the points T = 4 and T = 11) and the second one C_2 through the points of the other group. As is known from the basic law the straight line cuts the horizontal axis in the value of τ ; C_1 gives $\hat{\tau} \cong 5$; C_2 gives $\hat{\tau} \cong 12$.

6.1.3 <u>The increase of the accuracy of the extrapolated value b_o when more than two countings of the filteractivity are used.</u>

We should distinguish two cases.

a) Given the fact that only countings on the days $T_1 = 4$ and $T_2 = 11$ have been made, one could ask what may be the influence on the accuracy of the extrapolated (zero day's) value b_o if also countings on days between T_1 and T_2 would have been performed and used in the extrapolation. With n countings in total (first one on $T = 4$; last one on day $T = 11$) we find for the accuracy $(\delta_o)_n$ of the extrapolated value b_o the expression

$$(\delta_o)_n = \bar{\sigma}\sqrt{\frac{12\,m^2 + 12\,m + 2(2n-1)/(n-1)}{n(n+1)/(n-1)}} \quad , \quad n = 3, 4, 5 \text{ etc.,}$$

where $m = \frac{4}{11-4}$, $\bar{\delta}$ = average of the accuracies of the n countings and the countings are supposed to be equidistant in time.

When substituting $\bar{\delta} = \frac{1}{7}(1.73+1.72+1.76+1.77+1.79+1.83+2.58)=1.88\%$ we obtain

n =	2	3	4	5	6	7	8	times
$(\delta_o)_n$ =	8.9	2.7	2.5	2.3	2.2	2.1	2.0	%

We observe that: $(\delta_o)_n$ decreases with increasing n.

Consequently, when counting the filteractivity on each of 8 successive days (T = 4, 5, 6, 7, 8, 9, 10,11), all countings being equally accurate (1.88%), and when using all these countings to find b_o, then the accuracy of b_o becomes <u>2.0%</u>; based on only two countings, namely the first one (T = 4) and the last one (T = 11), the accuracy was <u>8.9%</u>.

b) Given the fact that only countings on the days $T_1 = 4$ and $T_2 = 11$ have been made, one could ask what may be the influence on the accuracy of b_o if also countings on the days 18, 25, 32 etc. would have been made and used in the extrapolation; again we suppose that the countings lie equistant in time. Then:

$$(\delta_o)_n = \bar{\sigma}\sqrt{\frac{12\,m^2 + 12\,m(n-1) + 2(2n-1)(n-1)}{n(n^2-1)}} \quad ; \quad n = 3, 4, 5 \text{ etc.}$$

with $m = \frac{4}{11-4}$ and $\bar{\sigma} = \frac{1}{7}(1.73+2.58+ \ldots 2.46) = 2.30\%$

Hence we obtain

n =	2	3	4	5	6	7	times
$(\delta_o)_n$ =	8.9	2.7	2.4	2.1	1.9	1.8	%

We see: $(\delta_o)_n$ decreases with increasing n, but quicker than in case a.

6.1.4 Three measurements to compute c, τ, p in $\beta = c(\tau + T)^{-p}$

We need at least three measurements b_1, b_2, b_3 on the days T_1, T_2, T_3 in order to solve three unknowns c, τ, p. For τ we get

$$\frac{\log q_{12}}{\log q_{13}} = \frac{\log(\tau + T_2) - \log(\tau + T_1)}{\log(\tau + T_3) - \log(\tau + T_1)} \; ; \; q_{12} = b_2/b_1 ; \quad q_{13} = b_3/b_1$$

Suppose we want to use the countings on the days $T_1 = 4$; $T_2 = 11$, $T_3 = 18$ with $b_1 = 222.6$; $b_2 = 112.7$; $b_3 = 83.1$, so that $q_{12} = 0.505$ and $q_{13} = 0.372$; $(\log q_{12}) : (\log q_{13}) = 0.692$.
The τ equation can be solved only approximately. Since the right hand side of the equation gives 0.670 for $\tau = 0$ and decreases for increasing τ, the equation does not give a positive solution for τ. How to explain this fact ?

a) first explanation

The point $T = 18$ probably does not belong to the groups of points for $T \leqslant 12$.

When using the countings for the days $T = 4, 8, 11$, then the τ becomes nearly 6 days.

b) second explanation

Since b_1, b_2, b_3 are inaccurate, also the ratio $R = (\log q_{12}) : (\log q_{13})$ is accurate and even far more inaccurate than each of b_1, b_2, b_3 separately. The question arises how to find the accuracy of R, written as

$$\sigma\left(\frac{\log q_{12}}{\log q_{13}}\right) \quad ?$$

Here we use a statistical theorem, which says:

$$\frac{\sigma^2(\underline{x}/\underline{y})}{\varepsilon^2(\underline{x}/\underline{y})} \cong \frac{\sigma_x^2}{\mu_x^2} + \frac{\sigma_y^2}{\mu_y^2} \text{ , if } \underline{x} \text{ and } \underline{y} \text{ are uncorrelated.}$$

Substitute $\underline{x} = \log \underline{q}_{12}$ and $\underline{y} = \log \underline{q}_{13}$, so that

$$\sigma_x^2 = \frac{\sigma^2(q_{12})}{\mu^2(q_{12})} = \delta_1^2 + \delta_2^2 \quad \text{and} \quad \sigma_y^2 = \frac{\sigma^2(q_{13})}{\mu^2(q_{13})} = \delta_1^2 + \delta_3^2$$

Hence

$$\frac{\sigma^2(R)}{\mu_R^2} \cong \frac{\delta_1^2 + \delta_2^2}{\mathcal{E}^2 \log q_{12}} + \frac{\delta_1^2 + \delta_3^2}{\mathcal{E}^2 \log q_{13}}$$

This is not quite correct because x and $\underline{y}$ are "pseudocorrelated", since q_{12} and q_{13} are ratio's with the same denominator b_1.

Substituting $\mathcal{E}\log q_{12} \cong \log \mathcal{E}\, q_{12}$ and $\mathcal{E}\, q_{12} \cong 0.505$; $\mathcal{E}\, q_{13} \cong 0.372$ we finally obtain

$\sigma(R)/\mu_R \cong 12.4\%$

This means, that the 0.95 reliability region of the true value, from which 0.692 is a point estimate, is situated between 0.692 (1-2x0.124) and 0.692(1+2x0.124) or between 0.52 and 0.37. In other words: the value 0.692 is so inaccurate that the reliability interval of τ , if computed with only three countings of the filteractivity, is extremely broad and even contains the zero value. The lesson could be to use more than the minimum number of three countings. We did not elaborate on this aspect.

6.2 Duplo measurements at the K.N.M.I., De Bilt

To begin with we analysed in the beginning of September 1958 the differences between duplo measurements made in De Bilt (two sampling units were installed close to each other). Two so called identical instruments had been measuring during 48 days simultaneously the atmospheric artificial radioactivity; it appeared sometimes that the components of one and the same duplo differed strongly, even as much as the largest difference between the measuring stations in the country on that day in question. It is exactly for this reason that we started to study the statistical aspects of measurements on the atmospheric radioactivity.

The results referring to this period of 48 days are summarized in table 1 . They suggested to continue these simultaneous measurements in order to obtain a better estimate of the random error.

The second analysis refers to 212 duplo's, see table 2.

The general conclusion is that, at least within the range 0.1 - 13.0 pC/m^3, the standardeviation σ_E of the total random error did not depend statistically significant on the level. Further: systematical errors seem to be absent, so that it is not necessary to reduce the one measured value positively and the other negatively in order to make the measurements with two adjacent sampling units better comparable.

Further: a statistical test learnt that the hypothesis $\sigma_E(1)=\sigma_E(2)$ needed not rejection. In other words: the spatial random errors of both instruments probably are equal (the random counting errors certainly are).

It is regrettable, that the accuracy of the counting of the filteractivities varies much if the level varies much. This is due to the counting rules in the laboratory of the K.N.M.I. Consequently, it is difficult to compute σ_e from σ_E by means of $\sigma_e^2 = \sqrt{\sigma_E^2 - \sigma_B^2}$, unless one would do so on each day separately. Of course an approximation was obtained by substituting the overall mean value $\bar{\bar{s}}_B$ (an average over all daily s_B-values for σ_B).

A second difficulty is the fact that the samples are not constant. See e.g. instrument 1 in table 2, where the daily samples varied between 24 and 52 m^3, around a mean value 35 m^3. Still the assumption of their constancy underlies our statistical considerations.

In the first analysis (table 1) we worked with the zero day's values (each based upon two countings of the filteractivity, usually on the days 4 and 11). As is shown in this report the inaccuracy of the extrapolation value is always larger (sometimes much larger) than the inaccuracies of the basic measurements separately. For this reason in the second analysis (table 2) only the 4th day countings of the filteractivities are used. Since $\sigma_E^2 = \sigma_e^2 + \sigma_B^2$ the consequence will be (σ_B in analysis 1 being larger than in analysis 2) that $\sigma_E(1) > \sigma_E(2)$. Now we see that the s_E values in table 1 surpass the s_E values in table 2, but statistically spoken it is not justified to conclude that the fact $s_E(1) > \sigma_E(2)$ implies $\sigma_E(1) > \sigma_E(2)$ without considering the large reliability intervals for the σ's as a consequence of the fairly small amount of data.

Further, it is questionable whether to reject the "suspected" duplo's (the components differing markedly) or not when computing s and $\widehat{\Delta A}$. If there are no reasonable special physical arguments for such a rejection, also these duplo's should be taken into account in the statistical considerations. Then s increases considerably. See table 2. There were made radiograms of the filters of some of these special duplo's, but unfortunately they did not indicate for certain whether hot spots caused the extremely large differences or not. So the problem what to do with such duplo's remains unsolved.

Table 1: Duplo measurements; July + Augustus 1958 De Bilt, K.N.M.I.

periods of nearly constant level	pC/m^3 $\bar{z}_1$	$\bar{z}_2$	n days	s_E%	0.95 reliability interval for σ_E%	$\lvert\widehat{\Delta A}\rvert$	$\hat{\sigma}(\widehat{\Delta A})$	correction for syst. error?	air sample Vm^3/day
12-15Aug.	0.52	0.36	4	20	6-34				
23J-10Aug.	0.86	0.73	19	21	14-28				
14 -20July	1.29	1.17	7	19	9-29				
all levels whatever	0.95	0.84	48	19	16-22	0.060	0.019	no	38-49 mean 44

z_1 varied between 0.14 and 1.84 pC/m^3
z_2 varied between 0.18 and 1.62 pC/m^3

Duplo measurements: two filters were installed close to each other. The one was counted in De Bilt (z_1), the other in Rijswijk (z_2).

Both z values are based on extrapolation values of the filteractivity.

Table 2: Duplo measurements; November 1958 - July 1959 De Bilt, K.N.M.I.

$\bar{z}=\frac{1}{2}(z_1+z_2)$ interval of $\bar{z}$ pC/m^3	mean $\bar{z}_1$ range	mean $\bar{z}_2$ range	n days	s_E%	0.95 confidence region for σ_E%	$\lvert\widehat{\Delta A}\rvert$	$\hat{\sigma}(\widehat{\Delta A})$	correction for syst. errors?	Vm^3/day	s_B=% 1 2	s_E if $\sigma_1=\sigma_2$ %
0-2	1.37 0.39-2.14	1.42 0.32-2.25	55 (52)	14 (12)	11-17						
2-4	2.95 1.75-4.26	2.98 0.93-4.34	95 (93)	15 (10)	13-18						
>4	5.51 2.29-12.09	5.51 2.40-11.61	62 (59)	19 (9)	16-20						
all	3.3 0.39-12.09	3.3 0.32-11.61	212 (204)	16 (10)	14-18	0.007 (0.013)	0.008 (0.005)	no	1) 24-52 mean 35 2) 19-46 mean 30	4.2 4.4	15 (12)

z_1 varied between 0.39 and 12.09 pC/m^3

z_2 varied between 0.32 and 11.61 pC/m^3

All z values are based upon the 4th day countings of the filteractivity.

The figures between brackets are valid if the suspected duplo's are dropped.

Then s_E is decreased considerably.

TABLE 3

Suspected duplo's; July 1958 - July 1959; De Bilt, K.N.M.I.

date	z_1	z_2 pC/m^3	z_i/z_j
13/14 July 1959	0.21	1.56	7.4
12/13 Nov. 1958	3.63	0.93	3.9
9/10 Nov. 1958	10.11	3.26	3.1
30/1 Dec. 1958	5.47	2.40	2.7
26/27 Dec. 1958	2.27	6.45	2.2
10/11 July 1959	0.79	1.65	2.1
7/8 Dec. 1959	1.38	2.64	1.9
9/10 July 1959	0.60	1.08	1.8
25/26 Dec. 1958	2.27	3.98	1.8

6.3 Analysis of duplo measurements of atmospheric radioactivity (calcination and non calcination method) in Belgium

0. Introduction

Two so called identical instruments for measuring the artificial atmospheric radioactivity were installed close to each other at Ukkel in Belgium. The daily sampled volumes of air were large, say about 1600 m^3. For each of both instruments each day the exposed filter is divided into two parts in a ratio 7:93. The activity of the small part is counted directly with a Geiger Müller counter. The larger part is calcinated and the activity of a part of the ash is counted. In this way each of both apparates gives two values for the specific atmospheric radioactivity, each of which may be considered as an estimate of one and the same unknown true activity. These four figures will be mentioned here, for the sake of simplicity, the "non calcination" (A) and "calcination" (B) figures. The first pair will be denoted with z_1 and z_2

(instruments 1 and 2); the second pair with z_3 and z_4 (also given by instruments 1 and 2).

As said already, these figures z_1, z_2, z_3, z_4[1] will be considered as simultaneous measurements of one and the same unknown value ζ , which generally will vary from day to day. It is as if four instruments (in stead of two) have been installed "at one point".

Several questions arise:

1. Is there a systematical difference between the non calcination data z_1 and z_2 ?
2. Is there a systematical difference between the calcination data z_3 and z_4 ?
3. Is there a systematical difference between the calcination figure and the non calcination figure of one and the same instrument and is this difference different for the instruments ?
4. Do these systematical differences depend on the radioactivity level ?
5. How large are the random errors for each of both methods and for each of both apparates ?
6. Do the random errors depend on the level ?

1. Random errors

The 94 daily simultaneous measurements z_1, z_2, z_3 and z_4, kindly placed at my disposal by Dr. Grandjean, relate to the period 28-7-1958 to 18-11-1958. The z_1 values varied between 0,2 and 8,0 pC/m^3. We have made two groups:

I. all quadruplets with $z_1 < 1.00$ (45 cases);

II. all quadruplets with $z_1 \geqslant 1.00$ pC/m^3 (49 cases). (See table 1).

All calcinations are based on the natural logarithms $y_i = \lg z_i$; i = 1, 2, 3, 4.

If the standarddeviations of the random errors in the non calcination case are equal ($\sigma_1 = \sigma_2$, say σ_A), and do not depend on the level, and if also $\sigma_3 = \sigma_4$, say σ_B, then the best estimates $\widehat{\sigma}_A$ and $\widehat{\sigma}_B$ of σ_A and σ_B are given by $\frac{1}{2} s_\Delta \sqrt{2}$, with $\Delta = y_1 - y_2$ resp. $y_3 - y_4$. Then $\widehat{\sigma}_A = 0.13$; $\widehat{\sigma}_B = 0.06$. The difference between 0.13 and 0.06 proves to be significant. This result could be understood a priori,

Note[1]) All figures are computed with the 4th day countings. We preferred to avoid the use of extrapolation values.

because the ratio of the sampled volumes of air for both methods will be (nearly) equal, day for day, to the ratio 7:93 in which the filter was divided and it is easily understood that σ decreases with increasing air sample (see table 2). Next the level of activity is taken into consideration by grouping all data in only two groups, so that the numbers per group are not too small and not too unequal.

In this way the boundary level was laid at 1.00 pC/m^3. Then the non calcination σ's become 0.15 and 0.12 and the calcination σ's become 0.058 and 0.056. The difference between 0.058 and 0.056 proves to be not significant. The difference between 0.15 and 0.12 seemed to be caused by too rough a calculation (to save time all logarithms were noted in only two decimals) and by several more or less "outlying" pairs (large ratio's of z_1 and z_2 or z_3 and z_4). The conclusion may be:

a. there is a calcination effect on σ;

b. there is not a level influence.

Next we verify the assumption $\sigma_1 = \sigma_2$ and $\sigma_3 = \sigma_4$ for the measurements in total.

We obtained $\widehat{\sigma_1^2} = 0.01695$; $\widehat{\sigma_2^2} = 0.01705$ for method A and

$\widehat{\sigma_3^2} < 0$; $\widehat{\sigma_4^2} = 0.0159$ for method B.

The values 0.01695 and 0.01705 are statistically equal. Of course it will surprise that $\widehat{\sigma_3^2} < 0$ (whereas $\sigma_3^2 > 0$). This strange result is both a sampling effect and the consequence again of too rough a calculation.

The conclusion may be: both with regard to the calcination results (B) and to the non calcination results (A) the instruments seem to be identical in the random errors.

2. <u>Systematical errors.</u>

When comparing the "four" instruments, the systematical errors are calculated in the standard way with the help of the mean y values $\bar{y}$ and the overall mean y value $\bar{\bar{y}}$; $\bar{\bar{y}} = \frac{1}{4} (\bar{y}_1 + \bar{y}_2 + \bar{y}_3 + \bar{y}_4)$;
$\widehat{\Delta A}_1 = \bar{\bar{y}} - \bar{y}_1$ etc. (See table 3).

The $\widehat{\Delta A}$ figures are very accurate. In the method A both figures possess a standarddeviation 0.011; in method B 0.007. There seems to be a slight level effect. Since, however, the difference between the group $\widehat{\Delta A}$ values in case A (see table 4) is not significant and likewise in case B (table 5), the conclusion could be, when using all four figures together, to correct the A figures with -16% and the B figures +16%. However, looking in detail it seems senseless to do even this, because the systematical errors turn out to be small with respect to the random errors.

To explain this we would refer to the formulae developed for the so called selection or rejection level, treated in chapter 5. In this case we are dealing with k = 4; the four σ's are unknown a priori and they are certainly not all equal. The section level is:

$$4\tau_j + 2F_j(\tau)\sqrt{\frac{k-1}{nk}} \quad \text{with} \quad \tau_j = \hat{\sigma}_j + 2\hat{\sigma}(\hat{\sigma}_j);\ \hat{\sigma}(\hat{\sigma}_j) = \hat{\sigma}_j : \sqrt{2n(k-1)};$$

$$F_j^2 = \left\{\overline{\tau^2} + (k-2)\,\tau_j^2\right\} : (k-1);\quad \overline{\tau^2} = \frac{1}{k}\Sigma\,\tau_j^2;\quad \bar{\tau} = \frac{1}{k}\Sigma\,\tau_j$$

Substituting $\sigma_1 = \sigma_2 = 0.13$ and $\sigma_3 = \sigma_4 = 0.06$, we obtain as levels
0.597 0.597 0.282 0.282.
Only in case the measured $\widehat{\Delta A}_j$ surpasses its corresponding rejection level, it deserves application. In any case $\widehat{\Delta A}_j$ should surpass the $4\tau_j$ (nearly $4\sigma_j$) value. In all four cases the $\widehat{\Delta A}$ values is much smaller than the rejection level. Hence: the random errors are so large with respect to the systematical errors that application of these systematical errors seems to be rather senseless.

3. Interpretation

Suppose the instruments are installed in future at two different stations any-where in Belgium. Suppose at some day the one instrument gives $z_1 = 2.0$, the other $z_2 = 2.2$ pC/m^3 (non calcination method). These figures look different, but is the difference real ? As said already the application of a correction for systematical errors is not necessary, hence we only should consider the random error.
$y_1 = \lg z_1 = 0.682$ and $y_2 = \lg z_2 = 0.786$. The 0.95 confidence region for the true value η_1 (measured as y_1) is 0.682 - 2 x 0.13 = 0.422 to 0.682 + 2 x 0.13 = 0.942; hence ζ_1 is situated between 1.53 and 2.58 ($z_1 = 2.00$ is an estimate of ζ_1).

Likewise ζ_2 is situated between 1.69 and 2.85 ($z_2 = 2.20$ is an estimate of ζ_2). The confidence intervals are so broad[1] because of the large σ values. There need not to be any difference between ζ_1 and ζ_2 (and even $\zeta_2 < \zeta_1$ may be true), although $z_2 > z_1$. The values of σ_1 and σ_2(0.13 and 0.06) are so large, that, keeping $z_1 = 2.0$ and increasing z_2 from 2.2 to higher values, only for z_2 larger than 3.2 it is "statistically certain" (with a 0.05 chance of being wrong) that $\zeta_2 > \zeta_1$. This value 3.2 is found by means of $y_2 \geqslant 0.786 + 2 \times 0.13 \cdot \sqrt{2} = 1.154$ and $y_2 = \lg z_2$.

Table 1

pC/m³	n	non calculation A			calculation B			V m³ mean			
		$\bar{z}_1$	$\bar{z}_2$	extreme z_1/z_2	$\bar{z}_3$	$\bar{z}_4$	extreme z_1/z_2	1	2	3	4
I $z_1 < 1.00$	45	0.90	0.66	0.64; 1.47	0.50	0.49	0.88;1.31				
II $z_1 \geqslant 1.00$	49	2.23	2.11	0.81; 1.95	1.60	1.62	0.76;1.12				
all	94	1.60	1.42	0.64; 1.95	1.07	1.08	0.76;1.31	110	126	1385	1576
mean		1.51			1.08						
z_1 values varied between 0.2 and 8.0 pC/m³											

Table 2

σ values in %

under the assumption $\sigma_1 = \sigma_2 = \sigma_A$; $\sigma_3 = \sigma_4 = \sigma_B$

n	A(1 ; 2) non calc.	B(3 ; 4) calc.
I 45	12	5.6
II 49	15	5.8
all 94	13	5.7

Note[1]) And they are still broader when taking into account also the random counting error.

Table 3

$\widehat{\Delta A}$ values in %

	A non calc.		B calc.		
n	1	2	3	4	sum
I 45	-17.3	-14.6	+12.7	+19.1	0
II 49	-18.5	-12.8	+15.8	+15.3	0
all 94	-17.8	-13.5	+14.3	+17.1	0
mean	-16		+16		

Table 4

A

duplo measurements (non calc)

n	$\widehat{\Delta A_1}$%	$\widehat{\Delta A_2}$ %	sum
I 45	-1.4	+1.4	0
II 49	-2.8	+2.8	0
all 94	-1.0	+1.0	0

Table 5

B

duplo measurements (calc.)

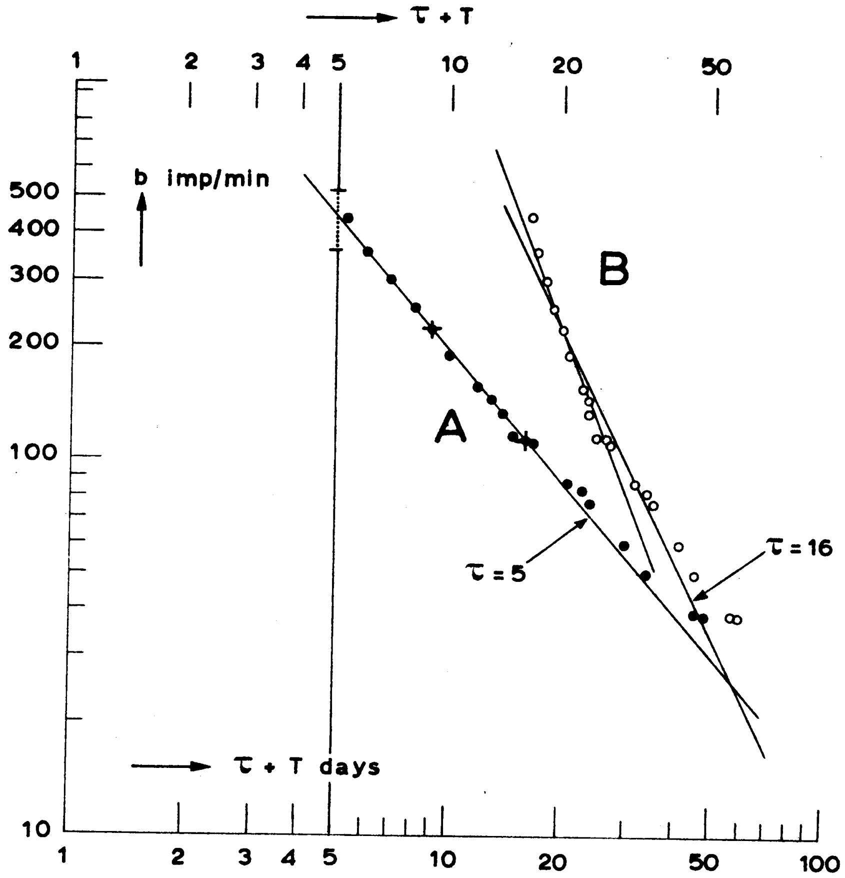
n	$\widehat{\Delta A_3}$%	$\widehat{\Delta A_4}$%	sum
I 45	-3.2	+3.2	0
II 49	-0.2	+0.2	0
all 94	-0.6	+0.6	0

There was no time for reproducing ADDENDUM A and B. As the report can be read without studying the Addenda, it has been reproduced in the present form.

Addenda can be forwarded as soon as possible on request to interested readers.

o-o-o-o-o

Fig. 1



$\beta = \text{const.}\ t^{-p};\ t = \tau + T$

$\log \beta = \log \text{const.} - p \log(\tau + T)$

A $\tau = 5 \longrightarrow p \cong 1.20$

B $\tau = 16 \longrightarrow p \cong 2.0$

$T_1 = 4;\ b_1 = 222.6;\ \delta_1 = 1.73\%$
$T_2 = 11;\ b_2 = 112.7;\ \delta_2 = 2.58\%$ $\longrightarrow b_0 = 446$

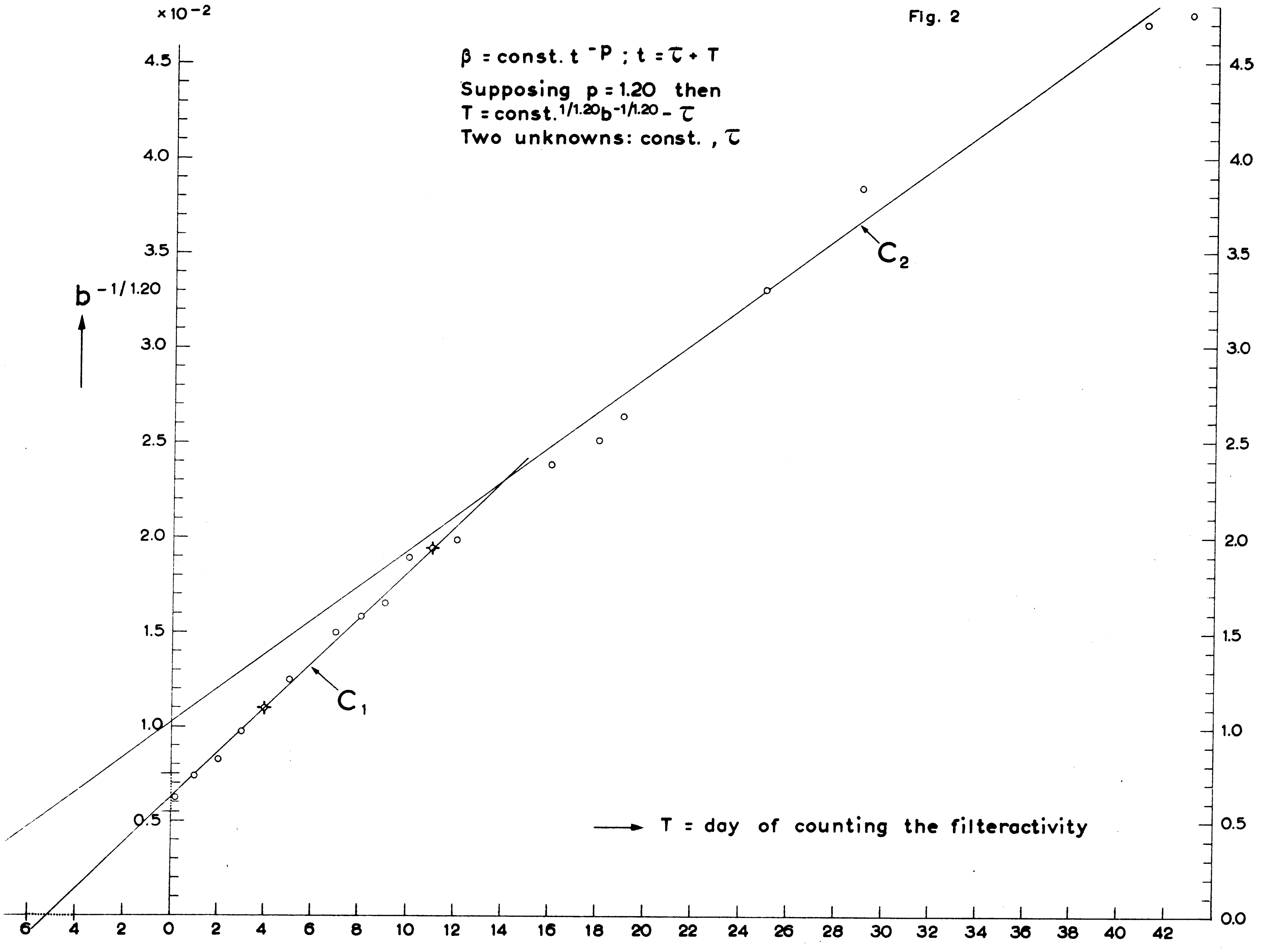
$\delta_0 = 8.9\%$

0.95 reliability interval for the true specific atm. art. radioact. in the sample 34.4 m^3 is $16.4 < I < 23.5$ pC/m^3; $\hat{I} = 19.9$

Taking into account the spatial sampling error:

$11.9 < I < 27.9$ pC/m^3; $\hat{I} = 19.9$

Fig. 2



× 10⁻²
β = const. t^{-p} ; $t = \tau + T$
Supposing p = 1.20 then
$T = \text{const.}^{1/1.20} b^{-1/1.20} - \tau$
Two unknowns: const. , τ
$b^{-1/1.20}$
C1
C2
→ T = day of counting the filteractivity

ADDENDUM A

Statistical aspects when counting the activity of the exposed filter.

CONTENT

.-.-.-.-.

0. Introduction

Let be:

β = true, unknown, mean number per minute of impulses (counts) produced by the pure (naked) filteractivity;
$\hat{\beta}$ = estimate of β .

κ = true value of the unknown, mean number of impulses (counts) per minute produced by a filter which is not infected by radioactive atmospheric dust (background, zero effect).
In De Bilt K.N.M.I. $\kappa \cong 17$ imp./min.

$f = \beta/\kappa$

$\hat{f} = \hat{\beta}/\kappa$ = estimate of f

$\sigma(\hat{\beta})$ = standarddeviation of $\hat{\underline{\beta}}$

$\hat{\sigma}(\hat{\underline{\beta}})$ = estimate of $\sigma(\hat{\beta})$

$\delta \equiv \sigma(\hat{\underline{\beta}})/\beta$ = percentual standarddeviation of $\hat{\underline{\beta}}$ or relative accuracy

$\hat{\delta} = \hat{\sigma}(\hat{\underline{\beta}})/\hat{\beta}$ = estimate of δ

Now the specific artificial atmospheric radioactivity (e.g. in pC/m^3) may be written as $R = const.\hat{\beta}/V$ 1), if V m^3 air have been sampled. The purpose is to measure the unknown β . The measurements give an estimate $\hat{\beta}$ of β. There are two possibilities.

i) The measuring-time T is pregiven; then the total number of counts $\underline{N}$ will be stochastic (meaning: for an immediately following measurement of T minutes the value of $\underline{N}$ may be quite different).

ii) The measuring total number N of counts is pregiven; then the measuring time $\underline{T}$ will be stochastic (meaning: for an immediately following measurement the period for obtaining N impulses may be quite different).

In both cases $\hat{\underline{\beta}}$ will be stochastic, i.e. $\hat{\underline{\beta}} - \beta$ will be distributed in some (symmetrical) way around zero and the standarddeviation $\sigma(\hat{\underline{\beta}})$

---

1) Strictly speaking this is not correct. In the formule for R the $\hat{\beta}$ should refer to the zero day's value β_0 of the pure (naked) filter-activity (of only artificial nature). It is, however, impossible to find β_0 by counting the filteractivity only once. At least two measurements (days T_1, T_2, see Addendum B) are needed. These measurements $\hat{\beta}_1$ and $\hat{\beta}_2$ give $\hat{\beta}_0$ by extrapolation.

in this distribution will be a measure for the accuracy with which the computed $\hat{\underline{\beta}}$ represents the unknown value β .

Some practical questions which gave rise to the following statistical considerations are:

A. One wants to measure a given <u>unknown</u> filteractivity β with a <u>given</u> accuracy δ . What must be the total number N of impulses to be "almost sure" that the requirement is satisfied ?

B. When measuring an <u>unknown</u> filteractivity β so long that the number of impulses N attains a <u>pregiven</u> value, what is then the accuracy δ of the measured filteractivity ?

C. Prescribing the total number N of impulses, one wants to know the range of filteractivities β , the measurements of which are more accurate than a pregiven value δ .

These questions can be answered if it is assumed that the background activity is exactly known.

1. Derivation of the formulas

1.1 Prefixed measuring time T; the so called "preset time method"

Suppose κ is known (by means of measuring the unexposed filter during e.g. 24 hours, so that the computed value is highly accurate). After having counted during T minutes, the total number of counts $\underline{N}$ is read off. $\underline{N}$ may be written as:

$$\underline{N} = \underline{N}_{\beta} + \underline{N}_{\kappa}$$

with $\underline{N}_{\beta}$ from β-effect; $\underline{N}_{\kappa}$ from κ-effect. Both $\underline{N}_{\beta}$ and $\underline{N}_{\kappa}$ follow a poissonian distribution.

Consequently, $\mathcal{E}\underline{N}_{\beta} = \sigma^2(\underline{N}_{\beta}) = T.\beta$; $\mathcal{E}\underline{N}_{\kappa} = \sigma^2(\underline{N}_{\kappa}) = T.\kappa$ and $\mathcal{E}\underline{N} = T(\beta + \kappa)$.

N.B.: $\underline{N}_{\beta}$ and $\underline{N}_{\kappa}$ are distributed independently.

Consequently,

$$\sigma^2(\underline{N}) = \sigma^2(\underline{N}_{\beta}) + \sigma^2(\underline{N}_{\kappa}) = T(\beta + \kappa) \tag{1}$$

We only count $\underline{N}$, not $\underline{N}_{\beta}$ and $\underline{N}_{\kappa}$ separately. Nevertheless we need $\underline{N}_{\beta} = \underline{N} - \underline{N}_{\kappa}$. After estimating $\underline{N}_{\kappa}$ by $T.\kappa$ (in individual cases $T.\kappa$ will be either larger or smaller than $\underline{N}_{\kappa}$, but in the long run $T\kappa$

will be a correct estimate of $\underline{N}_\kappa$), $\underline{N}_\beta$ is estimated as $\widehat{\underline{N}}_\beta = \underline{N} - \widehat{\underline{N}}_\kappa = \underline{N} - T.\kappa$.

Next an estimate $\widehat{\underline{\beta}}$ of β is made by:

(2) $$\widehat{\underline{\beta}} = \widehat{\underline{N}} / T = (\underline{N}/T) - \kappa$$

N.B.: Since $\underline{N}$ is stochastic, also $\widehat{\underline{\beta}}$ is stochastic.

Now $\mathcal{E}\widehat{\underline{\beta}} = (\mathcal{E}\underline{N} - T.\kappa) /T = T\beta / T = \beta$; this estimate $\widehat{\underline{\beta}}$ is unbiassed.

The accuracy $\sigma(\widehat{\underline{\beta}})$ of $\widehat{\underline{\beta}}$ is given by:

$$\sigma(\widehat{\underline{\beta}}) = \sigma(\frac{\underline{N}}{T} - \kappa) = \frac{1}{T}\sigma(\underline{N}) = \sqrt{T(\beta + \kappa)} : T,$$

and the percentual accuracy $\delta \equiv \sigma(\widehat{\underline{\beta}}) / \beta$ thus becomes:

(3) $$\boxed{\delta = \frac{\sqrt{1 + 1/f}}{\sqrt{\beta T}}}$$

Here δ is expressed in the pregiven T and the unknown β , which appears as well in the denominator as in the numerator (in f). It is clear that $\widehat{\beta}$ should be substituted for β in order to obtain an estimate $\widehat{\delta}$ of δ .

The formula shows that <u>only for sufficiently large activities (i.e. $f \gg 1$) the percentual accuracy δ is inversely proportional to the square root both of the pure activity itself and of the measuring time.</u>
For radioactivities small with respect to the background, this law does not hold.

In (3) δ is expressed in terms of β and T. Of course it is also possible to express β in δ and T, or T in β and δ :

(4) $$\boxed{\beta = \frac{1 + \sqrt{1 + 4\kappa T\delta^2}}{2T\delta^2}}$$ if $\kappa T\delta^2 \gg 1$, that is $\frac{1}{f}(1 + \frac{1}{f}) \gg 1$ or $f \ll 1$

(5) $$\boxed{T = \frac{1 + 1/f}{\beta\delta^2}}$$

Three points may illustrate the use of these relations:

a) For given β and T the δ follows from (3);

b) If one wants to measure a given β with a given accuracy δ , one should measure during T minutes, as given by (5);

c) Only activities higher than β , given by (4), can be measured with a given accuracy δ , if one wants to measure not longer than a given period of T minutes.

1.2 Prefixed number N of counts; the so called "preset count method"

Again κ is supposed to be known exactly. Now $\underline{T}$ is stochastic. Since $N = \underline{N}_\beta + \underline{N}_\kappa$, and since $\underline{N}_\kappa$ is unknown, an estimate $\widehat{\underline{N}_\kappa} = \underline{T}.\kappa$ is substituted, so that $\underline{N}_\beta$ is estimated as $\widehat{\underline{N}_\beta} = N - \underline{T}.\kappa$ and β is estimated as:

(6) $$\widehat{\underline{\beta}} = \widehat{\underline{N}_\beta} : \underline{T} = (N/\underline{T}) - \kappa$$

Note the difference with (2). In (2) T is pregiven and $\underline{N}$ is stochastic, in (6) N is given and $\underline{T}$ is stochastic.
This estimate $\widehat{\underline{\beta}}$ proves to be biassed, that is $\mathcal{E}\widehat{\underline{\beta}} \neq \beta$.
Before giving the proof it is necessary to derive the probability distribution of $\underline{T}$ and of $1/\underline{T}$, because $\sigma(\widehat{\underline{\beta}}) = N\sigma(\frac{1}{\underline{T}})$.
It is possible to show [1] that $\underline{T}$ is distributed as:

(7) $$\varphi(\underline{T})d\underline{T} = \frac{\mu^{N-1}}{(N-2)!} e^{-\mu\underline{T}} \underline{T}^{N-2} d\underline{T} \text{ , with } \varphi(o) = \varphi(\infty) = 0$$

and $\mu = \beta + \kappa$.
This distribution is not symmetrical. Some of its characteristics are:

(8) $$\mathcal{E}\underline{T} = \frac{N-1}{\mu} ; \quad \sigma^2(\underline{T}) = \frac{N-1}{\mu^2} ; \quad \mu_3 = \frac{2(N-1)}{\mu^3} \text{ etc. Mode } = \frac{N-2}{\mu}$$

The larger the N, the less asymmetric the $\underline{T}$ distribution.
Let be $\underline{x} = 1/\underline{T}$, then $\underline{x}$ is distributed as:

(9) $$\psi(\underline{x})d\underline{x} = \frac{\mu^{N-1}}{(N-2)!} e^{-\mu/\underline{x}} \underline{x}^{-(N-1)} d\underline{x} \text{ , with } \mathcal{E}\underline{x} = \frac{\mu}{N-2} ;$$

$$\sigma^2(\underline{x}) = \frac{\mu^2}{(N-3)(N-2)^2}$$

---

1) The only assumption is that the impulses are produced in time in such a way that the probability distribution of their situations in the time axis is a poissonian one.

Using these formulas, we obtain:

(9a) $$\delta = \frac{1 + 1/f}{\frac{N-2}{N}\sqrt{N-3}}$$

This is the exact expression for δ as a function of f and N.

(9b) Usually N is so large, that $$\delta \cong \frac{1 + 1/f}{\sqrt{N}}$$

Again δ is estimated as $\hat{\delta}$ by substituting $\hat{\beta}$ for β (in f) in (9b).

Special remarks:

a) As said already $\hat{\underline{\beta}}$ is not unbiassed. We have

$$\mathcal{E}\hat{\underline{\beta}} = \mathcal{E}\left(\frac{N}{\underline{T}} - \kappa\right) = N\,\mathcal{E}\left(\frac{1}{\underline{T}}\right) - \kappa = \frac{N}{N-2}\beta + \frac{2\kappa}{N-2} > \beta$$

The bias, however, is very small for large N.

b) Comparing (9b) with (3) we state:

In the preset count method δ diminishes approximately proportional to $\sqrt{N}$, but, strictly speaking, also the value of β itself affects δ, although the influence is less important for larger β values.

Formula (9b) may be rewritten as:

(10) $$\beta = \frac{\kappa}{\delta\sqrt{N-1}}$$

(11) $$N = \left(\frac{1 + 1/f}{\delta}\right)^2$$

Formula (9b) is well known in literature. The usual reasoning is then "Suppose N counts in total have been recorded, then the standarddeviation in this N is $\sqrt{N}$ and the relative standarddeviation is $t = \sqrt{N} : N = 1 : \sqrt{N}$. Now when measuring the unknown pure activity β, the standard-deviation consequently will be $t\,(\beta + \kappa)$, wherein κ represents the background activity, which is exactly known (e.g. 17 impulses/minute); the relative standarddeviation (relative accuracy) will be $t\,(\beta + \kappa) : \beta$. If for instance N = 2000 (hence t = 0.022) and if for instance the relative accuracy should be smaller than 10%, then β should be at least 4.8 impulses/minute, corresponding (if the efficiency is 0.30 and the air sample is 40 m^3) with 0.18 pC/m^3".

So far this reasoning, which in our opinion is not quite correct, since the fact that N is measured with a standarddeviation implies that N satisfies a probability distribution, in any case that N is not prefixed (as in the preset count method). However, the reasoning may be continued approximately as follows. Starting from (3) (in the preset time method) and remembering $\bar{N} = T(\beta + \kappa) = T\beta(1+\frac{1}{f})$, we obtain:

$$\delta = \frac{1 + 1/f}{\sqrt{\bar{N}}} \qquad (9b^{*})$$

Comparing ($9b^{*}$) with (9b) we like to draw attention to the difference between $\bar{N}$ and N. One should not "forget" the bar. Thus (9b) is "derived" in a way, which mathematically speaking is not correct, but which is permissible from a practical point of view. The error is small if the following condition is fulfilled. The preset number N_o should be so large that a relatively small difference between the actual measuring time T_o and the value $\bar{T}$, which is needed on the average to count N_o, is highly probable.

A second formulation of the same condition is the following. The number N_o should be prefixed in such a way, that a relatively small difference between N_o and the value $\bar{N}$, which is counted on the average in the actual measuring time T_o, will be highly probable.

With the help of the relations 9b, 10 and 11 it is possible to answer the questions A, B and C mentioned in the introduction.

Some numerical illustrations:

A. Let be N = 2000. Let be required $\delta \leqslant 10\%$. This is only possible (see (11)) for f - values equalling or exceeding 1 : $(0.10\sqrt{2000} - 1) = 0.299$. In case κ would be 17, this would indicate a range of β values $\geqslant 4.9$ impulses/minute.

B. At the time of writing this report, the Netherlands' rules for measuring filteractivities were as follows. Each measurement should not last less than 15 minutes. In case the pure activity is so small that within 15 minutes the total number of counts (pure activity plus background) would be less than 400, then this number 400 is preset and the measuring time is read off (larger than 15 minutes). If the pure activity is so large, that within 15 minutes the total number of counts would surpass 400, then one of a set of possible N values is chosen (400, 800, 1000, 2000, 3000 etc. according to circumstances) in such a way that the total measurement should not last much longer than 15 minutes.

Three practical examples may illustrate these rules.

a. Let β be very small, e.g. 3.4, so that (κ= 17) f = 0.2. With a reduction constant 1.5 and for a sample of 40 m^3 this β value would imply a specific artificial atmospheric radioactivity R=1.5x3.4/40 = 0.13 pC/m^3. Since 15 (β+ κ) = 15 (3.4 + 17.0) = 306< 400, one should prefix the total number of counts N on 400 and measure so long (of course longer than 15 minutes) that this number 400 is obtained. Then $\delta = (1 + \frac{1}{f}) : \sqrt{N} = (1 + \frac{1}{0.2}) : \sqrt{400}$ = <u>30%</u> ; $\bar{T}$ = 19.6 min.

b. Let β be very large, e.g. 3400, so that f = 200. With a reduction constant 1.5 and for a sample of 40 m^3 we obtain R = 1.5 x 3400/40 = 128 pC/m^3, being very large. Since 15 (β + κ) = 51255≫400, one should prefix the total number of counts N on (in accordance with the possibilities of the apparatus) for instance 55000 and read the measuring time (somewhat larger than 15 minutes). In this case $\delta = (1 + \frac{1}{f}) : \sqrt{N} \cong 1 : \sqrt{N} = 1 : \sqrt{55000}$ = 0.4% and $\bar{T}$ = 16.1 min.

c. For an intermediate value β = 415 (f = 23.4; R = 15 pC/m^3) we would obtain, if N = 8000 (hence $\bar{T}$ = 19 min.) a value δ = 1.2%, situated between 0.4 and 30.

These examples illustrate that <u>extremely different</u> activities are <u>not measured equally accurately</u>, see the difference between δ = 30, δ = 1.2 and δ = 0.4%, for 0.13, 15 and 128 pC/m^3.

This is, of course, a consequence of the instruction. If also in case a) the δ value should be about 0.4%, than N should have been (instead of 400): $[(1 + \frac{1}{f}) : \delta]^2 = (6 : 0.004)^2$ = 2250000 !! (measuring time nearly 110294 min = 1838 hours !)

Concl.: If one wants that the measuring periods are nearly constant and short (e.g. 15 minutes), even in case that the levels of radioactivity vary markedly, then the necessary consequence will be that the relative accuracies of the measurements of the filteractivities will vary strongly. If it is required that the percentual accuracies of the countings of the filteractivities will be next to constant, then the measuring durations will vary strongly and may be too long in practical routine.

The rules imply : since from 15 (β+ κ) = 400 and κ = 17 follows f = 0.57 we should distinguish between two classes : f <0.57 and> 0.57.

$f = \beta/\kappa$

for all cases $f < 0.57$ prefixed $N = 400$ $T > 15$ min. The larger T, the smaller $\delta = \frac{1 + 1/f}{\sqrt{400}}$	for all cases $f > 0.57$ prefixed N according to circumstances $T \cong 15$ min. The larger N, the larger f and $\delta \cong \frac{1}{\sqrt{N}}$ for large values of f
for $z = 1.5 \frac{\beta}{V}$ pC/m^3; $z < 0.36$ $V = 40$ m^3	$z > 0.36$ pC/m^3

Conclusion: if it desirable or advisable [1] to measure all artificial atmospheric radioactivities with (nearly) equal relative accuracies, then this wish should be taken into account when establishing the instructions for measuring. Application of the formulas and nomograms developed in this paper could be useful.

In Belgium N is prefixed at 3000, whatever may be the activity. Then δ has a value of $(1 + \frac{1}{f}) : \sqrt{3000}$.
For a comparison between the effects of the difference between the Netherlands and the Belgian counting rules, the following table is instructive.

			Netherlands			Belgian		
pure effect	$f=\beta/\kappa$	κ imp/min	N	$\bar{T}$ min	δ%	N	$\bar{T}$ min	δ%
0.13 pC/m^3	0.2	17	400	19.6	30	3000	147	11.0
15	23.4	17	8000	19.3	1.2	3000	7.2	1.9

N.B. A specific artificial atmospheric radioactivity corresponding with a pure β value, which would equal numerically the background effect (say 17 imp/min) amounts to ($V = 40$ m^3; reduction constant $1\frac{1}{2}$) 0.64 pC/m^3.

---

1) This is mainly a question for the user of the data. Because of the great variety of the users, the wishes with respect to the numerical value of δ differ markedly.

1.3 <u>The methods 1.1 and 1.2 do not differ statistically.</u>

There seems to be no statistical difference between the preset time and preset count method. This may be shown with three examples.

i) We want to measure certain β with a prescribed accuracy δ . What method (preset count or preset time) should be chosen ? In case of the preset time method, T should be fixed on $T_T \equiv (1+\frac{1}{f}) : \beta\delta^2$, see form (5).

In case of the preset count method, N should be prefixed on $N_c \equiv (1+\frac{1}{f})^2 / \delta^2$. The measuring time to obtain this number is on an average $\bar{T}_c \equiv N_c : (\beta+\kappa) = [(1+\frac{1}{f})^2 : \delta^2] / (\beta+\kappa) = (1+\frac{1}{f}) : \beta\delta^2$.

Consequently $T_T = \bar{T}_c$

ii) A given β is measured during a given time T in the preset time method. Consequently $\delta_T = \sqrt{1+\frac{1}{f}} : \sqrt{\beta T}$. In this T minutes on an average $\bar{N}_T = T.(\beta+\kappa)$ counts will be observed.

Next we measure the same activity in the preset count method by fixing N on $\bar{N}_T = T(\beta+\kappa)$.

What will δ_c be ? $\delta_c = (1+\frac{1}{f}) : \sqrt{\bar{N}_T} = \sqrt{(1+\frac{1}{f}) : \beta T}$.

Consequently $\delta_T = \delta_c$.

iii) Suppose β (and hence f) and T are pregiven in the preset time method. Then the relative accuracy becomes $\delta = \sqrt{1+\frac{1}{f}} : \sqrt{\beta T}$ and the number of counts, $\bar{N}_T$, in this time T is on an average $(\beta+\kappa) T = T\beta(1+\frac{1}{f})$.

Next suppose we measure the same β in the preset time method, and we require the same δ . Then the preset number of counts should be $N_c = (1+\frac{1}{f})^2 / \delta^2 = (1+\frac{1}{f})^2 / \{(1+\frac{1}{f}) : \beta T\} = T\beta(1+\frac{1}{f})$ [1].

Consequently $\bar{N}_T = N_c$.

---

1) These computations are based on the approximation 9b, which holds the better, the larger N. Starting from 9a, instead of 9b, N must be solved from a third degree equation. It can be shown that there is only one real root which is somewhat smaller than the right hand side of 9b, but the difference vanishes for large β values. Consequently only for very small values of β the preset count method may be considered better than the preset time pethod, i.e., on an average, the measurement will take a shorter period of time. Generally this aspect is completely unimportant.

Conclusion: Statistically speaking, the methods are identical.
Both methods require an exact knowledge of κ . Usually κ is measured during a large number of hours before and after the measurement of the exposed filter, so that the estimate $\hat{\kappa}$ is highly accurate.
The question ariseswhether it is preferable to count only the impulses of the naked activity or to measure the ($\beta + \kappa$)-effect and the naked κ effect simultaneously and separately (with two separate apparates). Let us first consider the first idea.

1.4 The naked β impulses (the so called anticoïncidence method).

This method usually is called the "anticoïncidence method". The name does not seem to be correct. If a coïncidence is defined as a β and κ - impuls occurring simultaneously, then the alternative (anti)situation would be either a naked β or a naked κ impuls. In fact in the anticoïncidence method only the naked β impulses are counted; the β impulses coinciding with κ impulses are lost, but this loss is unimportant.
In any case the marked advantage is: it is not necessary to know κ .
Again we may distinguish a preset time and a preset count method.

1.4.1 Preset time; given T

Substituting of $\kappa = 0$ ($f = \infty$) in (3), (4) and (5) gives:

(12)	$\hat{\underline{\beta}} = \underline{N} / T$	
(13)	$\delta = 1 / \sqrt{T\beta}$	relations between T, β and δ
(14)	$\beta = 1 / T\delta^2$	
(15)	$T = 1 / \beta\delta^2$	

1.4.2 Preset count; given N

We obtain:

(16)	$\hat{\underline{\beta}} = N / \underline{T}$	
(17)	$\delta = 1 / \sqrt{N}$	relations between δ and N;
(18)	$N = 1 / \delta^2$	β does not appear !

1.5 **Measuring total effect and background effect separately and simultaneously; two instruments-method.**

In the first instrument the exposed filter and in the second an unexposed filter is measured. So we measure $\beta + \kappa'$ and κ'' simultaneously; the smaller the distance between the instruments, the smaller the difference $\tau = \kappa'' - \kappa'$, where κ' and κ'' represent the backgrounds for the instruments. We want to derive the formulas as generally as possible and again for the use of either the preset time or the preset count method.

1.5.1 **Preset time method; T is fixed**

After T minutes in the first instrument $\underline{N}' = \underline{N}_\beta + \underline{N}'_\kappa$ counts are measured and $\underline{N}''_\kappa$ in the second. If $\underline{N}'_\kappa$ is replaced by $\underline{N}''_\kappa$, then $\underline{N}_\beta$ may be estimated as:

$$\widehat{\underline{N}}_\beta = \underline{N}' - \underline{N}''_\kappa = \underline{N}_\beta + (\underline{N}'_\kappa - \underline{N}''_\kappa)$$

(19) so that $$\boxed{\widehat{\underline{\beta}} = (\underline{N}' - \underline{N}''_\kappa) : T}$$

is an estimate of the unknown β.

Evidently $\widehat{\underline{N}}_\beta$ is an unbiassed estimate of $\underline{N}_\beta$ only if the two background effects are equal. However, we do not want to make this assumption, but we do suppose that $\underline{N}'_\kappa$ and $\underline{N}''_\kappa$ are correlated with a correlation coefficient ρ

Then:

$$\sigma^2(\widehat{\underline{N}}_\beta) = \sigma^2(\underline{N}_\beta)+\sigma^2(\underline{N}'_\kappa)+\sigma^2(\underline{N}''_\kappa) - 2\rho\sigma(\underline{N}'_\kappa)\sigma(\underline{N}''_\kappa) = T\beta + T\kappa' + T\kappa'' - 2\rho\sqrt{T\kappa' . T\kappa''}$$

Introducing $\tau = \kappa'' - \kappa'$ and $|q| = |\tau| : \kappa' \ll 1$, then

$$\sigma(\widehat{\underline{N}}_\beta) = \sqrt{T[\beta + (1-\rho)(2\kappa' + \tau)]}$$

Since $\widehat{\underline{\beta}} = \widehat{\underline{N}}_\beta / T$ and $\delta = \sigma(\widehat{\underline{\beta}})/\beta$ we obtain

(20) $$\boxed{\delta \cong \frac{1}{\sqrt{T\beta}}\sqrt{1 + \frac{(1-\rho)(2+q)}{f}}}$$ with $f = \beta/\kappa'$

This is the general expression; δ depends not only on T, β, f, but also on q and ρ.

Formula (20) may be rewritten as:

(21) $$\boxed{\beta \cong \frac{1}{2T\delta^2}\left[1 + \sqrt{1 + 4(1-\rho)(2+q)\,\kappa T\delta^2}\right]}$$

and

$$T \cong \frac{1 + \frac{(1-\rho)\ (2+q)}{f}}{\beta\delta^2} \tag{22}$$

The approximative nature is a consequence of the assumption $|q| \ll 1$.

Of course we wish to compare (20) with (3) and examine whether the two instruments method has any advantage above the one instrument method. Therefore it is necessary that $(1-\rho)(2+q) < 1$ or $(1-\rho) < 1 : (2+q)$. Neglecting q against 2 (since it is very small, e.g. 0.02) the inequality would be fulfilled if $\rho > 0.5$. What can be said about the ρ value in practice ? We must bear in mind that ρ represents the correlation between the paired T minute numbers $\underline{N}'_{\kappa}$ and $\underline{N}''_{\kappa}$ and probably depends on T, i.e. grows with increasing T.
Now if ρ would surpass 0.5 and if $|q| \ll 1$, then the surprising result is that the preset time method using a second apparatus to measure the naked background effect simultaneously with the measurement of the total effect leads to <u>smaller</u> δ values (for given T and β) than the one instrument method (where κ should be known exactly). It is true that knowledge of both κ' (in order to compute f) and κ'' (in order to find τ and q) would be necessary, but, although κ' and κ'' may show small day to day variations, the difference τ may be (nearly) constant, so that only one very accurate measurement of τ would be needed, whereas the κ' could be measured less accurate. Unfortunately we do not know whether ρ depends on T and how this dependence may be.

We were able to estimate the above mentioned ρ in the following way. In two neighbouring instruments during 93 nights the background effects were measured. The measuring times were not equal night for night. They varied between 15 and 16 h. The total numbers varied around 16000. For each of the 93 nights the one minute mean value was calculated for each of the apparates; the overall averages proved to be 17.1 for the first instrument, 16.7 for the second. These values 17.1 and 16.7 differ statistically significant. The correlation coefficient between the 93 paired one minute mean values turned out to be r=0.73, which differs statistically from zero (the 0.95 reliability interval for the correlation coefficient is 0.59-0.82)[1].

---

1) See following page.

Let us return to (20) again and consider four special situations A, B, C, and D.

A) Suppose $\tau = 0$ ($\kappa' = \kappa''$) N.B. This does not include that $\rho = 1$

Then:

(23)
$$\boxed{\delta = \frac{1}{\sqrt{T\beta}}\sqrt{\frac{1+2(1-\rho)}{f}}}$$

B) Suppose $\rho = 1$. This needs not to include that $\tau = 0$. If for instance in each individual minute the background effect in the second instrument would be twice the background effect in the other one then $\rho = 1$ and $\tau \neq 0$. It looks curious that even if τ would be not zero, still (20) changes into (13) of the anticoïncidence method, but we should bear in mind that using (20) β must be substituted by its estimate $\hat{\beta}$ in (19) and just if $\tau \neq 0$ this estimate would be biassed. Generally in practice, however, the situation $\rho \rightarrow 1$ will include $\tau \rightarrow 0$.
Now, substituting $\rho = 1$ and $\tau = 0$ in (20), (21), (22), the relations (13), (14) and (15) appear. So this hypothetical situation is analoguous to the anticoïncidence method.

---

1) At the time of finishing this report some results as to the value of ρ with regard to its dependence on T became available. For two counting apparates, installed near to each other, at every hour the one hour sums of counts were read photographically simultaneously. These background measurements could be made only during the night. In this way n=364 pairs of one hour totals were obtained. The correlation coefficient r was $r_1 = -0.006$, which value does not differ statistically significant from zero. The 0.95 reliability interval is -0.11 to +0.11. Next adjacent and not overlapping groups of k=12 were made and the group sums were calculated. The correlation coefficient r_{12} between the n=30 pairs became 0.37 (reliability interval 0.00 to 0.65). Next k was made 18. Now n=20 and r_{18}=0.58 (interval 0.19 - 0.79). Next k was made 24. Now n=15; r_{24}=0.592 (0.11-0.84). The 364 one hour sums for the one instrument varied around 1063, with a standarddeviation 31; for the second instrument these figures were 1083 resp. 34. The one minute means were 17.7 and 18.1 imp/min.
Concl.: it is highly probable that ρ increases with T and is 0.5 at about T = 12 hours.

(24) C) Suppose $\rho = 0$, then $$\delta = \frac{1}{\sqrt{T\beta}}\sqrt{1 + (2+q)\frac{1}{f}}$$

Since $2 + q > 1$ for every κ' and κ'', this δ value always exceeds the value in (3). This result could be understood a priori: if the instruments (the one measuring the true effect plus background; the other only the background) are installed so far from each other that the background effects are uncorrelated, then it is always certain, whatever may be the background values, that, for given β and T, the two instruments method leads to larger δ values than the one instrument method (this means that the two instrument method is less accurate).

D) Suppose $\rho = 0$ and $\tau = 0$ ($\kappa' = \kappa''$)

This means that the backgroundeffects are exactly equal, whereas the correlation is zero. These facts need not to be contradictory.

Then we obtain:

(25) $$\delta = \frac{\sqrt{1 + 2/f}}{\beta T}$$

(26) $$\beta = \frac{1 + \sqrt{1 + 8\kappa T\delta^2}}{2T\delta^2} \cong \frac{\sqrt{2}}{\delta}\sqrt{\frac{\kappa}{T}} \quad \text{if } \kappa T\delta^2 \gg 1$$

(27) $$T = \frac{1 + 2/f}{\beta\delta^2}$$

$$\text{i.e. } \frac{1}{f}\left(1+\frac{1}{f}\right) \gg 1 \quad \text{or } f \gg 1$$

We wish to compare (25), (26), (27) with the expressions (3), (4) and (5).

The approximations for β hold good if $f \gg 1$, that is for sufficiently small β values and, consequently, for large T values.

Then from (3) to (25) the δ is increased with a factor of almost $\sqrt{2}$;

from (4) to (26) the β is multiplied with nearly $\sqrt{2}$;

and from (5) to (27) the T is nearly twice as large (whereas T itself is already large).

The larger the β (that is the smaller the f) the smaller are these enlargements.

1.5.2 Preset count method; N is fixed

This somewhat unrealistic case implies the following: in the first instrument the measurement of the exposed filter would last so long that a pregiven number N of counts is reached, whereas in the second one an unexposed filter is counted also so long that the number of counts attains the same value N. Then $N = N_\beta + N'_\kappa = N''_\kappa$, but N is measured in T'minutes in the first instrument and in T'' minutes in the second one, $T'' \neq T'$ generally.

Since N'_κ is unknown, we substitute $\frac{T'}{T''} N''_\kappa$ for N'_κ, so that the estimate of N_β becomes:

$$\hat{N}_\beta = N - \hat{N}'_\kappa = N - N''_\kappa \frac{T'}{T''} = N\left(1 - \frac{T'}{T''}\right) \quad \text{and}$$

$$\boxed{\hat{\beta} = \frac{\hat{N}_\beta}{T'} = N\left(\frac{1}{T'} - \frac{1}{T''}\right)}$$

As to the bias of $\hat{\beta}$ we find

$$\mathcal{E}\hat{\beta} = N\left[\mathcal{E}\frac{1}{T'} - \mathcal{E}\frac{1}{T''}\right] = N\left[\frac{\beta+\kappa'}{N-2} - \frac{\kappa''}{N-2}\right] = \frac{N}{N-2}(\beta - \tau) \neq \beta$$

so that the bias is smaller, the larger N and the smaller the difference between the two background effects.

The computation of $\sigma^2(\hat{\beta})$ becomes very difficult.
Although it is possible to calculate both $\sigma(\frac{1}{T'})$ and $\sigma(\frac{1}{T''})$ we also need the correlation coefficient ρ^* between $1/T'$ and $1/T''$. We have not made the calculation, since this situation seems to be too unrealistic. When measuring with two instruments simultaneously (the one for the total effect, the other for the background effect) it seems more reasonable that the measurements should last the same time T and that T should be pregiven.

2. A comparison between the one and two instruments method.

Nomograms.

Nomograms enable the user to avoid calculations.
In the first place a nomogram A based on formula (3) has been drawn. We take f as parameter and plot δ and N on axes with logarithmic scales. Then each f curve is a straight line. There exists no straight line at the upper side, but we propose to take $f = 0.1$ ($\beta = 0.1\,\kappa$). However, there is a boundary line at the lower side ($f = \infty$), representing the anticoincidence method. The density of lines increases with increasing values of f.

Numerical examples

1) Suppose $\beta \cong \frac{1}{2}\kappa$. Require a measurement with $\delta = 0.05$. Then the nomogram shows that N should be fixed on 3600 or so. If however δ needs not to be smaller than 0.10, then N = 900.
For $\beta \cong 10\,\kappa$ the analoguous N numbers would be 500 and 120.
2) If N would be fixed on 400, then the nomogram shows that there are no β values which can be measured with accuracies $\delta < 0.05$, since the straight line $f = \infty$ passes through the point with coordinated $\delta = 0.05$ and N = 400.
3) If a very accurate measurement is required ($\delta = 0.01$) then the low activities (say $f < 1$) require preset numbers of counts larger than 40000.

Next a second, supplementary graph B is constructed by means of which a new auxiliary parameter f^* can be read so that even in the two instruments method the formula (3) referring to the one instrument method can be used only by substituting f^* for f.
Now compare (20) and suppose $\kappa' = \kappa''$ (i.e. q = 0), so that (23) results. Introducing $\varepsilon = 1 - \rho$ we obtain:

(28) $$\delta = \frac{\sqrt{1 + 2\frac{\varepsilon}{f}}}{\sqrt{\beta T}} = \frac{\sqrt{1 + 2\frac{\varepsilon}{f}}}{\sqrt{\bar{N} : (1+\frac{1}{f})}} = \frac{\sqrt{(1 + \frac{1}{f})(1 + 2\frac{\varepsilon}{f})}}{\sqrt{\bar{N}}} = \frac{1 + \frac{1}{f^*}}{\sqrt{\bar{N}}}$$

In this way we have defined f^*. Write $f^* = f.c$, so that

$$c = \frac{1}{-f + \sqrt{f^2 + (1 + 2\varepsilon)f + 2\varepsilon}}$$

We have drawn curves of c against f for the values ρ = 0, 0.1, 0.2, 0.4, 0.5, 0.6, 0.8, 0.9 and 1.0 . Then if we want to measure according to the two instruments method during a pregiven time T, we can compute, as soon as ρ is given, δ as follows.
The measurements give an estimate $\hat{\beta}$ of β , see (19).
With this estimate $N = T(\hat{\beta} + \kappa)$ can be computed. Nomogram B gives (with $f = \hat{\beta} : \kappa$) the value of c. Hence $f^* = f.c$ is known. These values $T(\hat{\beta} + \kappa)$ and f^* are substituted for $\bar{N}$ and f in ($9b^*$). Hence δ is known.

When drawing the nomogram $c=c\ (f, \rho)$ we should bear in mind that:

a) If $\rho = \frac{1}{2}$, then $c=1$ irrespective of f (horizontal curve).

b) If $\rho < \frac{1}{2}$, then $c<1$ for all f values; c decreases very slowly for increasing f, the slower, the nearer ρ to 0.5.

c) If $\rho > \frac{1}{2}$. then $c>1$ for all f values; c decreases with increasing ρ the quicker (particular for $f<1$) the nearer ρ to 1.

d) If $\rho = 0$, then $\delta = (1 + \frac{1}{f^*}) : \sqrt{\bar{N}}$, with $f^* = cf$ and $c = 1 : (-f + \sqrt{f^2 + 2f + 2})$

For the range $f = 0 \rightarrow \infty$ then $c = 0.71 \rightarrow 0.67$

e) If $\rho = 1$, then $\delta = \sqrt{(1 + \frac{1}{f}) : \bar{N}} = 1 : \sqrt{\beta T}$;
also $\delta = (1 + \frac{1}{f^*}) : \sqrt{\bar{N}}$
with $f^* = fc$ and $c = 1 : (-f + \sqrt{f^2 + f})$.
This represents the anticoincidence method.

f) If $f \rightarrow \infty$ then $c \rightarrow 1 : (0.5 + \varepsilon)$, with $\varepsilon = 1 - \rho$

Some numerical applications.

1) Suppose f is about 2 and suppose two apparates are used of which the average one minute values of the backgrounds are equal and of which the background values are correlated strongly e.g. with $\rho = 0.8$. Suppose one wants to measure the unknown β with a relative accuracy $\delta = 0.05$. Then the question arises: should we measure with only one instrument (preset count) the ($\beta + \kappa$)-effect and next with the same apparatus the κ effect separately, highly accurately ? Or, should we apply to the two instruments method ($\beta + \kappa$ and κ simultaneously)? Answer: graph A gives for $\delta = 0.05$ and $f = 2$ a value $N = 900$ in the one instrument preset count method. With graph B we obtain for $f = 2$ and $\rho = 0.8$ a value $c = 1.3$ and consequently $f^* = cf = 3$. Next we enter graph A with $f = 3$ and $\delta = 0.05$. Consequently $N = 760$. Hence the method of simultaneous measurements with two separate instruments requires a 16 percent smaller number of counts and consequently a 16 percent smaller measuring-time. The disadvantage, however, is that two instruments are occupied simultaneously, whereas in the one instrument method only one apparatus is needed. During nights before and after the measurement of the total effect this apparatus may be used for measuring the naked background effect. It is to be decided by the user which of the two methods is preferable.

2) For very large activities, say $f > 10$, c is nearly constant with regard to f but depends strongly on ρ . But since the density of straight lines in graph A grows fast with increasing f, the two instruments method, whatever may be the value of ρ , does not change N seriously. Now the difference between the two methods is relatively small.

3) For very low activities, say $f < 0.5$, c is not constant with regard to f, especially for $\rho > \frac{1}{2}$, and moreover depends strongly on ρ ; say, c lies between 1 and 4 (provided that $\rho > 0.5$). Consequently the preset number of counts is diminished markedly when using the two instruments method and the more markedly, the larger ρ .

.-.-.-.-.

ADDENDUM B

Statistical aspects concerning the estimates of
the "age" and "the zero day's " value
of the pure filteractivity
by means of measurements of the filteractivity
on two, three, or more days after
the sampling day.

.-.-.-.-.-.-.-.-.

.-.-.-.-.

Statistical aspects concerning the estimates of the "age" and the "zero day's value" of the pure filteractivity by means of measurements of the filteractivity on two, three, five or more days.

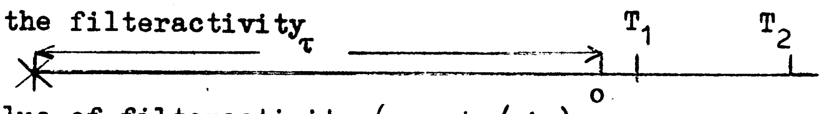
0. Introduction

Definitions

τ = number of days between the day of explosion and the day of sampling the air; "age"

T = number of days between the sampling day and the day of counting the filteractivity

$t = \tau + T$



β = true mean value of filteractivity (counts/min), corrected for background

(1) Basic principle: $\boxed{\beta = \text{const. } t^{-p}}$

p = constant, dependent on the composition of the radioactivity (e.g. $p = 1.20$). (The value of p is constant during the period between the day of explosion and the day of last measurement of the filteractivity).

Given β_1 (on day T_1) is measured as b_1, with a relative accuracy δ_1; β_2 (on day T_2) is measured as b_2, with a relative accuracy δ_2.

Problem 1) Find the age $\hat{\tau}$ and its accuracy δ_τ.
2) Find the zero day's activity b_0, and its accuracy δ_0, using the two inaccurate measurements b_1 and b_2.

1. Age, estimate and accuracy

1.1 Estimate Substituting $\beta_1, \tau + T_1$, resp β_2, $\tau + T_2$ in (1) for β and t, we find two equations with two unknowns, to wit the constant and τ.

(2) For τ: $\boxed{\tau = \dfrac{T_2\varphi^a - T_1}{1 - \varphi^a}}$ with $a = 1/p$; $\varphi = \beta_2/\beta_1$

In practice $q = b_2/b_1$ is substituted for $\varphi = \beta_2/\beta_1$ and consequently an estimate $\hat{\tau}$ of τ is found.

(3) $$\boxed{\widehat{\tau} = \frac{T_2 q^a - T_1}{1 - q^a}} = f\,(b_1 ; b_2) \quad \text{with} \quad q = b_2/b_1$$

1.2 Accuracy

A sharper formulation of the problem goes as follows. Suppose one would be able te remeasure the same true, but unknown, values β_1 and β_2 again and again, then generally each time a new pair of values b_1 and b_2 would result, each pair being an estimate of β_1 , β_2 . Each pair b_1, b_2 furnishes a value $\widehat{\tau}$ (see (3)). Since b_1 satisfies a probability distribution around value β_1, with a standarddeviation σ_1 (with $\delta_1 = \sigma_1/\beta_1$) and since the same applies to b_2, also $\widehat{\tau}$ follows a distribution. We like to know the mean value $\mathcal{E}\widehat{\tau}$ and the standard-deviation $\sigma(\widehat{\tau})$. Even if b_1 and b_2 would follow gaussian distributions, the distribution of $\widehat{\tau}$ will not be symmetrical and hence certainly not normal because of the fact that $\widehat{\tau}$ is not a linear function of b_1 and b_2.

Three attacks of the problem will be treated here:

a) Strictly spoken, as soon as the probability functions $g_1(x_1)dx_1$ of x_1, with $m_1 \leqslant x_1 \leqslant n_1$, and $g_2(x_2)dx_2$ of x_2, with $m_2 \leqslant x_2 \leqslant n_2$, are given, with $y = f(x_1, x_2)$, then the probability function $g(y)dy$ can be derived as:

$$g(y)dy = \int_{m_1}^{y} g_1(x_1) g_2(x_2) dx_1 \quad \text{with } x_2 = h(x,y) \text{ from } y = f(x_1,x_2)$$

with $\min.y \leqslant y \leqslant \max.y$, where min.y and max.y depend on m_1, n_1, m_2, n_2. Next all moments in this y-distribution could be calculated. This is very difficult for a function as given in (3). We have given up this attack.

b) It seems reasonable to reason as follows. Given the probability functions of x_1 and x_2 (cf. previous point), there are 4 combinations of pairs: $\mathcal{E}x_1 \pm 2\sigma_1$, $\mathcal{E}x_2 \pm 2\sigma_2$, namely ++, +−, −+ and −−. For each of these pairs the value of y can be computed; take the smallest and the largest one. Call the difference between these extremes $4\sigma_y$ and the arithmetic mean of the four y values $\mathcal{E}y$. This attack may be very incorrect and even senseless for "complicated" functions $y = f(x_1 ; x_2)$, as will shown further on.[1)]

---

1) See footnote on page 4B.

c) The third attack of the problem is followed usually. Before describing this attack, a statistical theorem should be mentioned. Let be given n interdependent stochastrical variables x_1, x_2 . . . x_n, with correlation coefficients ρ_{ij}, $i \neq j$; i, j=1, 2 . . . n. Let be μ_i the mean value of x_i; σ_i^2 the variance; μ_{ki} the k^{th} moment of x_i; i=1, 2 . . . n.
Let be given a function $y = f(x_1, x_2, \ldots x_n)$.
Then $\mathcal{E}_y$ and $\sigma(y)$ can be developed in a series:

(4) $$\mathcal{E}_y = f(\mu_1 . . \mu_2) + \tfrac{1}{2}[f_{11}\sigma_1^2 + \ldots f_{nn}\sigma_n^2] + \frac{1}{3!}[f_{11}\mu_{31} + \ldots f_{nn}\mu_{3n}] + \text{etc.}$$

Note 1) Of course, this attack is a better approximation of the exact method treated under c, the smaller the values of δ_1 and δ_2.
When substituting $b_2(1+2\delta_2)$, $b_1(1-2\delta_1)$ in q in (3) we obtain:

$$\tau^1 = \frac{T_2 q^a - T_1 + 2a(T_2 q^a \delta_2 + T_1\delta_1)}{1 - q^a - 2a(\delta_1 + q^a\delta_2)}$$

Likewise, substituting $b_2(1-2\delta_2)$, $b_1(1+2\delta_1)$ in q in (3) we obtain:

$$\tau^{11} = \frac{T_2 q^a - T_1 - 2a\,(\text{----})}{1 - q^a + 2a(\text{----})}$$ Hence $\tau^{11} < \tau^1$. Define $4\sigma_\tau = \tau^1 - \tau^{11}$.

In these expressions δ_1 and δ_2 are considered so small, that $(1\pm 2\delta_1)^a \cong 1 \pm 2a\delta_1$. The case $\delta_1 \neq \delta_2$ is somewhat complicated. Take for the sake of simplicity $\delta_1 = \delta_2$, say δ. Then:

$$\tau^1 - \tau^{11} = 4\sigma_\tau \cong 8a\delta\frac{q^a}{(1-q^a)^2}(T_2 - T_1).$$ substitute $q = \left(\frac{T_1+\tau}{T_2+\tau}\right)^{+p}$

(see (11)), and it follows that:

$$\delta_\tau = \sigma_\tau/\tau \cong 2\frac{(T_1+\tau)(T_2+\tau)}{(T_2-T_1)\tau}\,a\delta \quad \text{and } 2\delta \cong \delta$$ *

Hence the expression (12) reappears ! In this way the fact that attack b) holds the better, the smaller both δ_1 and δ_2, has been illustrated.

(5) $\sigma^2(y)=(f_1^2\sigma_1^2 + ..f_n^2\sigma_n^2) + 2[f_1 f_2 \rho_{1,2}\sigma_1\sigma_2 + ..f_{n-1}f_n\rho_{n-1,n}\sigma_{n-1}\sigma_n] +$ etc.

Here $f_i = \partial f/\partial x_i$ and $f_{ii}=\partial^2 f/\partial x_i^2$ for the arguments μ_1 etc. We wish to use this theorem. In our case $n = 2$; $x_1 \equiv b_1$; $x_2 \equiv b_2$; $y \equiv \hat{\tau}$; $f =$ function in (3); $\rho_{12}=0$.

Here $\rho_{12}=0$ since b_1 and b_2 are completely independent, for instance the chance that $b_2 > \beta_2$ does not depend on whether $b_1 > \beta_1$ or $b_1 < \beta_1$. Of course we use the exact (5) and truncate (4). Then:

(6) $\mathcal{E}\hat{\tau} \equiv \tau = f(\beta_1,\beta_2) = (T_2\varphi^a - T_1) : (1-\varphi^a)$ with $\varphi = \beta_2/\beta_1$ and

(7) $\sigma^2(\hat{\tau}) = f_1^2\,\sigma_1^2 + f_2^2\,\sigma_2^2$

The truncation is made here after the first term in (4); of course the consequence of this truncation should be verified. Therefore we also investigated the importance of the term in brackets with the σ_1^2 and σ_2^2; it turned out that (6) is a sufficiently close approximation of (4).

Next f_1 and f_2 are computed. We obtain:

(8) $f_1 = \partial f/\partial b_1 = -a\,\varphi^a \dfrac{T_2 - T_1}{\beta_1(1-\varphi^a)^2}$; $f_2 = a\,\varphi^a \dfrac{T_2 - T_1}{\beta_2(1-\varphi^a)^2}$

Substituting (8) in (7) it follows:

(9) $\sigma(\hat{\tau}) = \dfrac{\varphi^a(T_2-T_1)}{(1-\varphi^a)^2}\, a\,\delta_*^2$

where per definition $\delta_1 = \sigma_1/\beta_1$; $\delta_2 = \sigma_2/\beta_2$; $\delta_*^2 = \delta_1^2 + \delta_2^2$

Next follows $\delta_\tau \equiv \sigma(\hat{\tau}) : \tau$ For τ see (2). We then obtain:

(10) $\delta_\tau = \dfrac{\varphi^a(T_2-T_1)}{(1-\varphi^a)(T_2\varphi^a-T_1)}\, a\delta_*$

(11) In view of (1) we have $\varphi \equiv \dfrac{\beta_2}{\beta_1} = \left(\dfrac{T_2+\tau}{T_1+\tau}\right)^{-p}$

Substituting (11) in (10), with $a = 1:p$, we obtain:

(12) $$\boxed{\delta_\tau = \frac{(\tau + T_1)(\tau + T_2)}{(T_2 - T_1)\tau} a\delta_*} = v.a\delta_* \quad ; \quad v > 1$$

This is the most general expression for δ_τ.
Next the assumption is made, that the value of $\hat{\tau}$ is distributed around τ with a relative standarddeviation δ_τ in a gaussian way[1), so that there is a 0,95 probability to find $\hat{\tau}$ between $\tau(1-2\delta_\tau)$ and $\tau(1+2\delta_\tau)$

Expression (12) shows, that δ_τ is smaller,

a) the smaller τ, provided that $\tau > \sqrt{T_1 T_2}$;
b) the smaller T_1, irrespective of τ, T_2, T_1;
c) the larger T_2, for all T_1, T_2, τ;
d) the smaller δ_*, that is the smaller δ_1 and δ_2;
e) the larger $T_2 - T_1$, for all T_1, T_2.

From (12) follows that δ_τ certainly exceeds a δ_*
The factor $v \approx 1$ only if $T_1 T_2 + 3T_1\tau + \tau^2 \ll T_2\tau$
When $T_2 = mT_1$, with $m > 1$, this would require $\tau^2 + (3-m)\tau T_1 + mT_1^2 \ll 0$
Since this inequality cannot be fulfilled for reasonable values of T_1 and m, it is almost certain that δ_τ surpassed $a\delta_*$ and usually in a marked way.

There are two extreme cases:

A) $\tau \gg T_1$; the explosion took place "very long" ago. Then (12) becomes:

(12A) $$\boxed{\delta_\tau \cong \frac{\tau}{T_2 - T_1} a\delta_*}$$

If moreover $T_2 \gg T_1$, then $\delta_\tau = (1 + \frac{\tau}{T_2})\, a\delta_*$

B) $\tau \approx T_1$; $\tau \ll T_2$; explosion took place recently. Then (12) becomes:

(12B) $$\boxed{\delta_\tau \cong \frac{T_1 + \tau}{\tau} a\delta_*}, \text{ with } \frac{T_1+\tau}{\tau} > 1, \text{ but not } \frac{T_1+\tau}{\tau} \gg 1 .$$

1.3 How to choose T_2 ?

Generally $T_1 = 3, 4, 5$; how to choose the best possible value of T_2 ?
Suppose we want to estimate τ by only two measurements b_1 (day T_1)

1) It would be a very difficult problem to investigate what would be the degree of skewness of the true $\hat{\tau}$ distribution.

and b_2 (day T_2), then the question arises how to choose T_2 for given T_1, δ_1, δ_2 in order to find a sufficiently accurate $\hat{\tau}$? Now solving (12) with respect to T_2, we obtain:

(13) $$T_2 = \frac{\delta_\tau T_1 + (T_1 + \tau)\, a\delta_*}{\delta_\tau \tau - (T_1 + \tau)\, a\delta_*}\,\tau$$

1.3 Numerical examples

1) Let be $\delta_1 = \delta_2 = 5\%$, so that $\delta_* = 7.0\%$; let be $T_1 = 4$; $p = 1.20$ $a = 0.83$. Let on certain grounds be supposed that τ is about 100 days. We will require $\delta_\tau = 10\%$. Now, how to choose T_2 ? Equation (13) gives 163 days. Of course this T_2 is very large. It is "impossible" to wait so long before the second measurement b_2 can be made. What to do ?

a) One could choose T_2 smaller and be content with a far much larger δ_τ ; for instance $T_2 = 20$ and $\delta_\tau = 45\%$.

b) One could fix T_2 on 20 and prefer to make a third measurement, on day T_3. How to choose T_3 so that $\delta_\tau = 10$ to 20% ? This question leads to the formulation of a new problem: given n measurements $b_1, b_2 \ldots b_n$ with relative accuracies $\delta_1, \delta_2, \ldots \delta_n$, what is the best estimate of τ and what is the relative accuracy $\delta(\hat{\tau})$ and how should $T_1, T_2 \ldots T_n$ be chosen to fulfil special conditions ? This problem has only partly been solved in the present Addendum (see chapter 5).

c) One could measure β_1 and β_2 far more accurate. In case a): $\delta_\tau = 45\%$. If we want to decrease 0,45 to 0,10, then δ_1 and δ_2 should be, say, 4 times as small and the measuring times 16 times as large. If this would give practical difficulties, one could prefer the "two instruments method" or the anticoïncidence method for counting the filteractivity in order to save time.

If one wants not to wait so long before making the second measurement so that in any case $T_2 < \tau$ or even $T_2 \ll \tau$, then (13) gives $\delta_\tau T_1 + (T_1 + \tau)a\delta_* < (\ll)\ \delta_\tau \tau - (T_1 + \tau)\, a\,\delta_*$ and hence

(14) $$\delta_\tau > (\gg)\ 2\,\frac{\tau + T_1}{\tau - T_1}\, a\,\delta_* > 2a\delta_*$$

This implies that one should be content with an accuracy of τ , which is in any case larger than $2\, a\, \delta_*$ namely $1.7\, \delta_* = 2.4\delta$ for $p = 1.20$ and $\delta_1 = \delta_2$, day δ.

2) The second example refers to an observation at 6/7 - 1 - '59, de Bilt. On day $T_1=5$ the value $b_1=52.4$ imp/min and on day $T_2=12$ the value $b_2=46.7$ imp/min was measured (both corrected for background). The relative accuracies were $\delta_1=3.0$ and $\delta_2=3.1\%$. Substituting $p=1.20$ and $a=0.83$ (3), we obtain $\hat{\tau} = 65$ days.
Here $\delta_* = \sqrt{0.030^2+0.031^2} = 0.043$ and $a.\delta_* = 0.0356$.
Certainly $\delta_\tau > 0.0356$. The exact expression (12) gives $11.8\, a\, \delta_* = 0.42$, whereas the approximation ($\tau \gg T_1, T_2$, see 12A) gives $9.3 \times a\delta_* = 0.33$. The difference between 0.42 and 0.33 is not large. Note the large value of δ_τ which can be interpreted as: there is a probability of 0.95 that τ is situated between $(1-2 \times 0.42)65$ and $(1+2 \times 0.42)65$, that is between 12 and 118 days. (Observe the marked breadth of this interval!). Let us now see how the incorrect attack b), section 1.2, works. Here consider the pairs $b_1(1+2\delta_1) = 55.5$; $b_2(1-2\delta_2) = 43.8$; and $b_1(1-2\delta_1) = 49.3$; $b_1(1+2\delta_1) = 49.6$. The first pair gives, see (12), $\hat{\tau} = 26$ days; for the second the computation comes to a deadlock, since $49.6 > 49.3$[1]. Nevertheless, there is a 0.95 reliability interval for τ , namely 12 - 118 day's.

3) Let be $T_1 = 3$; $b_1 = 50$; $\delta_1 = 0.01$ and $T_2 = 10$; $b_2 = 45$; $\delta_2 = 0.01$. Then (3) gives $\hat{\tau} = 74$ days; (12) gives $\delta_\tau = 0.146$, whereas (12A) furnishes 0.124. Here $a\, \delta_* = 0.0118$. The values 0.146 and 0.124 do not differ much. The 0.95 confidence region for τ is 52 to 96 days. The attack b), section 1.2, runs as follows: $b_1(1+2\delta_1) = 51.0$; $b_2(1-2\delta_2) = 44.1$ give $\hat{\tau} = 52$ days, whereas $b_1(1-2\delta_1) = 49.0$; $b_2(1+2\delta_2) = 45.9$ give $\hat{\tau} = 121$ days. This example illustrates that attack b now gives an incorrect answer, although not too incorrect.

Note 1) Notwithstanding the rather small values of δ_1 and δ_2 this approximative attack gives a senseless answer here. (See also the note on page 4B). Obviously, there are still more requirements to be fulfilled (such as: q sufficiently different from 1) before this attack can reasonably be applied.

Nevertheless, generally, we prefer attack c, described extensively in this Addendum.

2. Zero day's activity; estimate and accuracy

2.1 Estimate

Substituting β_o on $T_o = 0$, β_1 on $T = T_1$, β_2 on $T = T_2$ in the relation

$$\beta = \text{const.}\ t^{-p}\ , \text{ with } t = \tau + T$$

and eleminating the constant and τ, we obtain as true value of the filteractivity on day zero:

(15) $$\beta_o = \beta_2 \left(\frac{T_2 \varphi^a - T_1}{T_2 - T_1} \right)^{-p} \qquad \text{with } \varphi = \beta_2/\beta_1; \quad a = 1/p$$

Next we substitute b_o for β_o, b_1 for β_1 and b_2 for β_2, then

(15a) $$\boxed{b_o = b_2 \left(\frac{T_2 q^a - T_1}{T_2 - T_1} \right)^{-p}} = f(b_1, b_2) \qquad \text{with } q = b_2/b_1$$

2.2 Accuracy

As explained in the foregoing chapter on the $\hat{\tau}$ accuracy, we again follow the attack expressed in the formulas (4) and (5), here with $y = b_o$; $x_1 = b_1$; $x_2 = b_2$; f is the function (15a) and $\rho_{12} = 0$. Then we obtain:

(16) $$\mathcal{E} b_o = \beta_o = (15) \qquad \text{and}$$

(17) $$\sigma^2(b_o) = f_1^2 \sigma_1^2 + f_2^2 \sigma_2^2 \qquad \text{with } f_1 = \frac{\partial b_o}{\partial b_1} \quad \text{and} \quad f_2 = \frac{\partial b_o}{\partial b_2}$$

for the arguments β_1 and β_2.
We obtain:

(18) $$f_1 = \left(\frac{T_2 \varphi^a - T_1}{T_2 - T_1} \right)^{-\frac{a+1}{a}} \frac{\varphi^{a+1}}{T_2 - T_1} T_2 \quad \text{and} \quad f_2 = \left(\frac{T_2 \varphi^a - T_1}{T_2 - T_1} \right)^{-\frac{a+1}{a}} \frac{\varphi^a}{T_2 - T_1} T_2$$

Substituting (18) in (17) we obtain:

(19) $$\sigma^2(b_o) = \left[\frac{\varphi^a T_2}{T_2 - T_1} \left(\frac{T_2\varphi^a - T_1}{T_2 - T_1} \right)^{-\frac{a+1}{a}} \right]^2 (\varphi^2\sigma_1^2 + \sigma_2^2)$$

We wish to find the relative accuracy of β_o, denoted by $\delta_o \equiv \sigma(b_o):\beta_o$. With (15) and (19) we obtain:

$$\delta_o = \frac{T_2\varphi^a}{T_2\varphi^a - T_1}\delta_* \quad \text{with} \quad \delta_* = \sqrt{\delta_1^2 + \delta_2^2}$$

Substitution of (11) in (20) results in:

(21) $$\boxed{\delta_o = \frac{T_2(\tau + T_1)}{(T_2 - T_1)} \delta_*}$$

When substituting $\hat{\tau}$ (see (3)) for τ one obtains:

$$\hat{\delta}_o = \frac{T_2^*}{T_2^* - T_1}\delta_* = \frac{1}{1 - \frac{T_1}{T_2^*}}\delta_*, \text{ if } T_2^* = T_1 q^a$$

With regard to (12) we obtain:

(22) $$\boxed{\delta_\tau = \left(1 + \frac{\tau}{T_2}\right) a\delta_o}$$ Note that the expression does not contain T_1.

Obviously the estimate of τ is usually (very much) more inaccurate than the estimate of β_o, and this is the more so, the longer ago the explosion.

We see, that δ_o is smaller:

a) the larger τ
b) the smaller T_1
c) the larger T_2
d) the smaller δ_*
e) the larger $T_2 - T_1$

There are two extreme situations.

(23a) $\tau \gg T_1$ so that $\delta_o \cong \delta_*$ [1]. Then $\delta_o \cong \frac{T_2}{T_2 - T_1} \delta_*$. When moreover $T_2 \gg T_1$, then $\delta_o \cong \delta_*$ [1]. Explosion long ago.

(24b) $\tau \approx T_1 ; \tau \ll T_2$; explosion short ago. The $\delta_o \cong (1 + \frac{T_1}{\tau}) \delta_*$

whereas $\delta_\tau \cong a\delta_o$

2.3 How to choose T_2 ?

Generally $T_1 = 3, 4, 5$. How to choose T_2 ? Suppose we wish to estimate β_o by only two measurements b_1 (on day T_1), b_2 (day T_2) and to choose T_2, for given T_1, δ_1, δ_2, in order to find a sufficiently accurate b_o.

See (21) and solve T_2. We obtain:

(25)
$$T_2 = \frac{\delta_o T_1}{\delta_o \tau - (\tau + T_1)\delta_*} \tau$$

2.4 Numerical examples

1) Let be $\delta_1 = \delta_2 = 0.05$, so that $\delta_* = 0.07$; $T_1 = 4$; $p = 1.20$; $a = 0.83$. Suppose on certain grounds we know that $\tau \approx 100$ days. Suppose we require $\delta_o = 0.10$. Then (25) gives $T_2 = 15$ days. (Remember that $\delta_\tau = 0.10$ would require $T_2 = 163$ days).

Since usually $\tau \gg T_1$, (25) approximates to:

(25a)
$$T_2 \cong \frac{\delta_o T_1}{\delta_o - \delta_*}; \quad \text{consequently} \quad \frac{\delta_o}{\delta_o - \delta_*} > 1, \text{ since } T_2 > T_1.$$

If a δ_o value is wanted k times as large as δ_*, then T_2 should be chosen $\frac{k}{k-1}$ as large as T_1.

Note 1) In this way, provided that $\tau \gg T_1$ and $T_2 \gg T_1$ (e.g. $T_1 = 4$; τ , $T_2 > 15$), we see that the relative accuracy δ_o of the extrapolation value b_o only depends on the relative accuracies δ_1 and δ_2 of the basic measurements, because $\delta_o^2 \cong \delta_1^2 + \delta_2^2$, so that δ_o depends neither on p nor on τ.

$\delta_o = k\delta_*$, then $\frac{T_2}{T_1} = \frac{k}{k-1}$; $(k = 1; \frac{k}{k-1} = \infty)$ (1.5; 3) (2; 2) (3; 1.5) (4; $1\frac{1}{3}$) (5; $1\frac{1}{4}$) (10; 1.1).

2) See the second example in 1.4.
$b_1 = 52.4$; $T_1 = 5$; $b_2 = 46.7$; $T_2 = 12$; $\delta_1 = 0.030$; $\delta_2 = 0.031$; $p = 1.20$, so that (3) and (12) give $\hat{\tau} = 65$ days and $\delta_\tau = 0.42$.
Further (15a) gives $b_o = 57.4$ and (21) gives $\delta_o = 0.079$;
β_o lies between 48.3 and 67.5 imp/min.

The β_o region is rather broad in spite of the small δ values.

3) $b_1 = 50$; $T_1 = 3$; $b_2 = 45$; $T_2 = 10$; $\delta_1 = \delta_2 = 0.01$; $p = 1.20$, so that (3) and (12) give $\hat{\tau} = 74$ days and $\delta_\tau = 0.15$.
Further (15a) gives $b_o = 56.0$ and (21) gives $\delta_o = 0.020$;
β_o lies between 48.0 and 52.0 imp/min.

3. Nomogram.

3.1 Modification of existing nomogram

In the routine work at the Royal Meteorological Institute of the Netherlands a nomogram is used with the help of which the measurements b_1 and b_2 on the days T_1 resp. T_2 enable one to find quickly both the extrapolated (zero day's) value b_o (estimate of β_o) and the age $\hat{\tau}$ (estimate of τ).
We shall show how the nomogram has been modified so that it may be used also for reading the relative accuracies δ_o and δ_τ of b_o resp. $\hat{\tau}$.

The existing nomogram (see fig.D) contains a left hand vertical axis with the values of $m = T_2/T_1$ and a right hand vertical axis with the values of $q = b_2/b_1$. The oblique axis contains the values of:

$$c = q\left(\frac{mq^{0.834} - 1}{m - 1}\right)^{-1.20} \qquad (26)$$

The nomogram holds only for $p = 1.20$ ($a = p^{-1} = 0.834$).
The procedure is as follows:
Let be given measurement b_1 on day T_1 and measurement $b_2 (< b_1)$ on day T_2.

Calculate $m = T_2 : T_1$ and $q = b_2 : b_1$. The straight line through the points m and q cuts the c-axis at a value $c = c'$. Then $\underline{b_o = c'b_1}$. The straight line through the points $q = 1:c'$ and $c = \infty$ cuts the m-axis in m_τ ; then $\underline{\hat{\tau} = T_1 : (m_\tau - 1)}$.

(27) Now, see (10), δ_τ can be written: $\delta_\tau = c^{0.834} \dfrac{0.834\delta_*}{1-q^{0.834}}$

(28) and, see (21), δ_o can be written: $\delta_o = c^{0.834} \dfrac{m}{m-1} \delta_*$

Define $\tilde{c} = c^{0.834}$; $\tilde{q} = 0.834 : (1-q^{0.834})$ and $\tilde{m} = m : (m-1)$.

(29) Then: $\delta_\tau = \tilde{c}.\tilde{q}\,\delta_*$ and $\delta_o = \tilde{c}\tilde{m}\delta_*$.

This suggests to write the values of $\tilde{c}$ along the c-axis, the values of $\tilde{q}$ along the q-axis and the values of $\tilde{m}$ along the m-axis.

The following numerical example illustrates the use of the modified nomogram to find b_o, $\hat{\tau}$, δ_o, δ_τ by means of b_1(day T_1) and b_2(day T_2).

3.2 Numerical application

De Bilt 6/7-1-1959;

$b_1 = 52.4$; $\delta_1 = 0.30$; $T_1 = 5$; $b_2 = 46.7$; $\delta_2 = 0.031$; $T_2 = 12$.

Then $\delta_* = 0.043$; $m = 2.40$; $q = 0.891$; $\tilde{m} = 1.72$; $\tilde{q} = 9.3$.

Step a) gives $c' = 1.09$; $1/c' = 0.916$; $\tilde{c'} = 1.076$. Hence $\underline{b_o = 57.0}$ (57.4) imp/min.

Step b) gives $m_\tau = 1.08$; hence $\underline{\hat{\tau} = 62.5}$ (65) days.

Step c) gives $\delta_\tau = 0.43$ (0.42); confidence interval for τ is 9 to 116 days

$\delta_o = 0.079$ (0.079); confidence interval for β_o is 48 to 66 imp/min.

The results of calculation with the formulas have been placed between brackets. Hence the nomographic method is in general sufficiently accurate.

4. The exponent p in the basic law

4.1 The exponent is unknown and should be estimated

Let us again start from the relation $\beta_i = c(\tau + T_i)^{-p}$; $i = 1,2,3$. Now there are three unknowns c, p and τ and it is the purpose to compute on the basis of three measurement (the least number with which it will be possible) the values of $\hat{\tau}$ and β_o and their accuracies. In this Addendum only τ will be considered (the β_o case is very complicated).

4.1.1 Estimate of τ

From $\lg \beta_1/\beta_2 = -p \lg t_1/t_2$, with $t_i = \tau + T_i$, and $\lg \beta_2/\beta_3 = -p \lg t_2/t_3$, follows:

(30) $$\frac{\lg t_1/t_2}{\lg t_2/t_3} = \frac{\lg \beta_1/\beta_2}{\lg \beta_2/\beta_3}, \text{ say } k. \quad \text{Let be } k_o = \frac{\Delta 23}{\Delta 12} k$$

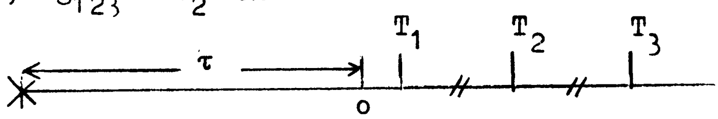
Now suppose $\tau \gg T_1, T_2, T_3$, so that, introducing $\Delta_{12} \equiv T_2 - T_1$; $\Delta_{23} \equiv T_3 - T_2$; $\Delta_{13} \equiv T_3 - T_1$; $g_{12} \equiv \frac{1}{2}(T_1 + T_3)$; $g_{123} \equiv \frac{1}{2}(g_{12} + g_{23})$; $g_{23} \equiv \frac{1}{2}(T_2 + T_3)$, we obtain:

(31) $$k \cong \frac{\Delta_{12}(\tau - g_{12})}{\Delta_{23}(\tau - g_{23})}$$

(32) Whence $$\boxed{\tau \cong \frac{1}{4} \frac{k_o + 1}{k_o - 1} \Delta_{13} + g_{123}}$$

Next let be $T_2 - T_1 = T_3 - T_2$, equidistant measuring days, so that $\Delta_{23} = \Delta_{12}$, say Δ; $\Delta_{13} = 2\Delta$; $g_{123} \equiv T_2$ and

(33) $$\boxed{\tau \cong \frac{1}{2} \frac{k+1}{k-1} \Delta + T_2}$$



Substituting the measurements b_1, b_2, b_3 for β_1, β_2, β_3 one obtains an estimate $\hat{\tau}$ of τ.

4.1.2 Accuracy of $\hat{\tau}$

Next we want to compute the relative accuray δ_τ of $\hat{\tau}$.

Using the expression $\sigma^2(\hat{\tau}) = f_1^2\sigma_1^2 + f_2^2\sigma_2^2 + f_3^2\sigma_3^2$, where $f_i = \partial f/\partial b_i$ with i = 1, 2, 3; $f \equiv \hat{\tau}$ is a function of b_1, b_2, b_3. We finally obtain, under the assumption $T_1, T_2, T_3 \ll \tau$,

(34) $$\boxed{\delta_\tau \cong \left(\frac{\tau}{\Delta}\right)^2 a\sqrt{\delta_1^2 + \delta_2^2 + \delta_3^2}}, \text{ with } \delta_i = \frac{\sigma_i}{\beta_1} \text{ and } i = 1, 2, 3;$$

a = 1/p, where p and τ are the true unknown values.

In the case of two measurements b_1 and b_2(days T_1, T_2) and under the assumption that p would be known exactly, we would have found:

(35) $$\boxed{\delta_\tau \cong \frac{\tau}{2\Delta} a\sqrt{\delta_1^2 + \delta_2^2}}$$

In the 2 measurements method $(\delta_\tau)_2$
In the 3 measurements method $(\delta_\tau)_3$

(36) Note the relation: $(\delta_\tau)_3 \cong \frac{2\tau}{\Delta}(\delta_\tau)_2$ with $\tau/\Delta \gg 1$.

We seem to win nothing. Of course this surprises, but let us state explicitly, that these two situations cannot be compared easily:

a) two measurements: only c and τ are unknown; p is known. At least two measurements are needed. They give two equations with two unknowns, of which only τ interests us; via $\hat{\tau}$, the estimate b_o of the zero day's value can be found.

b) three measurements c, p and τ are unknown; three equations with three unknowns, of which only p and τ interest us. By way of p and $\hat{\tau}$, the estimate b_o of the zero day's value can be found.

A comparison which would make more sense could be:

case a) take three measurements, although there are two unknowns c and τ. Then use a fitting formula to find τ and compute $(\delta_\tau)_{2;3}$; 2 unknowns, 3 equations;

case b) take four measurements, although there are three unknowns c, τ and p. Then use a fitting formula and find τ and compute $(\delta_\tau)_{3;4}$; 3 unknowns, 4 equations.

Next compare $(\delta_\tau)_{3,4}$ with $(\delta_\tau)_{2,3}$. We have worked out this statistical problem, which is rather difficult, only partly, because of the fact that the basic relation between β and t is exponential.

Next compare

$(\delta_\tau)_{3,5}$ with $(\delta_\tau)_{2,4}$; in general $(\delta_\tau)_{3.j}$ with $(\delta_\tau)_{2.j}$;
j = 3, 4, 5 etc.

4.2 The exponent p is inaccurately known

4.2.1 Expressions

In the foregoing chapters the basic law is $\beta = c.t^{-p}$ with a known value of p. The nomogram was constructed for p = 1.20. Each particular value of p would require a new nomogram. The question arises what may be the consequence of a small inaccuracy in the "known" value of p on the zero day value, its accuracy and on the age and its accuracy. We will derive expressions for $\partial b_o/b_o$, $\partial\delta_o/\delta_o$, $\partial\hat{\tau}/\hat{\tau}$, $\partial\delta_\tau/\delta_\tau$, for an infinitesimal increase ∂p of p. On the basis of the formulas for b_o, δ_o, $\hat{\tau}$ and δ_τ and since $\varphi^a = \frac{\tau + T_1}{\tau + T_2}$; $a = 1/p$, we come to:

(37)

$\partial\tau/\tau = -\lg q \frac{\delta_\tau}{\delta_*} \frac{\partial p}{p}$	$\frac{\partial\delta_\tau}{\delta_\tau} = -\frac{\tau^2 - T_1T_2}{\tau\Delta} \lg q \frac{\partial p}{p} \frac{1}{p}$
$\partial b_o/b_o = \lg q \frac{\delta_o}{\delta_*} \frac{\partial p}{p}$	$\frac{\partial\delta_o}{\delta_o} = \frac{T_1(\tau + T_2)}{\tau\Delta} \lg q \frac{\partial p}{p} \frac{1}{p}$

where $\varphi = \beta_2/\beta_1$; $q = b_2/b_1$; $\Delta = T_2 - T_1$; $\delta_* = \sqrt{\delta_1^2 + \delta_2^2}$; $\delta_\tau = \frac{\sigma(\hat{\tau})}{\tau}$; $\delta_1 = \frac{\sigma(b_1)}{b_1}$; $\delta_2 = \frac{\sigma(b_2)}{b_2}$.

These expressions hold only if $\partial p/p$ is small.
In case $\tau >> T_1$, T_2 one obtains:

(38)

$$\partial\tau/\tau \cong -\lg q \frac{\delta_\tau}{\delta_*} \frac{\partial p}{p} \qquad \partial\delta_\tau/\delta_\tau = -\frac{\tau}{\Delta} \lg q \, \partial p/p^2$$

$$\partial b_o/b_o \cong \lg q \frac{\delta_o}{\delta_*} \frac{\partial p}{p} \qquad \partial\delta_o/\delta_o = \frac{T_1}{\Delta} \lg q \, \partial p/p^2$$

4.2.2 Numerical example

De Bilt 6/7-1-1959; $T_1 = 4$; $b_1 = 52.4$ imp/min.; $\delta_1 = 0.030$;
$T_2 = 12$; $b_2 = 46.7$; $\delta_2 = 0.031$.
Consequently $b_o = 57.1$, with $\delta_o = 0.064$ and
$\hat{\tau} = 65$ days, with $\delta_\tau = 0.42$
(small δ_1 and δ_2 and still large δ_τ, and rather small δ_o).
The expressions derived above give:

(39) $\partial b_o/b_o = -0.21\ \partial p/p \qquad \partial\delta_o/\delta_o = -\frac{0.098}{p}\frac{\partial p}{p}$; if $p = 1.20$ then $-0.08\ \frac{\delta p}{p}$

$\partial\hat{\tau}/\hat{\tau} = 1.36\ \partial p/p \qquad \partial\delta_\tau/\delta_\tau = \frac{1.06}{p}\frac{\partial p}{p}$; if $p = 1.20$ then $0.08\ \frac{\delta p}{p}$

Interpretations:

per 1% increase of p, then b_o decreases with 0.2%; $\hat{\tau}$ increases with 1.4%; δ_o decreases with 0.1%; δ_τ increases with 0.9%.

For instance: if p would lie anywhere between 1.10 and 1.30 around 1.20, then:

the b_o range would be	56.1	to 58.1	imp/min	(range of	3.5%)
" $\hat{\tau}$ " " "	58	" 72	days	(" "	23 %)
" δ_o " " "	0.063	" 0.065		(" "	1.3%)
" δ_τ " " "	0.39	" 0.45		(" "	15 %)

In this example the relative total ranges of b_o and $\hat{\tau}$, caused by the inaccuracy of the exponent p, are almost the half of the corresponding relative inaccuracies δ_o and δ_τ of b_o and $\hat{\tau}$, caused by the inaccuracies in the measurements b_1 and b_2. In other words: it becomes fairly senseless to measure b_1 and b_2 extremely accurate if p itself is rather inaccurate. We will not elaborate here on this statistical aspect.

Substituting φ from (11) for q into (38) one obtains:

(a) $$\varepsilon\left(\frac{\partial b_o}{b_o}\right) = \frac{T_2(T_1+\tau)}{(T_2-T_1)\tau}\left(\lg\frac{\tau+T_1}{\tau+T_2}\right)\partial p < 0,$$ since $T_1 < T_2$ and $\partial p > 0$;

and

(b) $$\varepsilon\left(\frac{\partial\hat{\tau}}{\hat{\tau}}\right) = \frac{(T_1+\tau)(T_2+\tau)}{(T_2-T_1)}\left(\lg\frac{\tau+T_2}{\tau+T_1}\right)\partial p > 0,$$ since $\partial p > 0$.

Differentiating $\varepsilon\frac{\partial b_o}{b_o}$ resp. $\varepsilon\frac{\partial\hat{\tau}}{\hat{\tau}}$ as to T_2 or τ, we obtain:

$\partial(a)/\partial T_2 > 0$; $\partial(a)/\partial\tau < 0$; $\partial(b)/\partial T_2 < 0$ and $\partial(b)/\partial\tau \gtrless 0$, for sufficiently large τ, then > 0.

In other words:

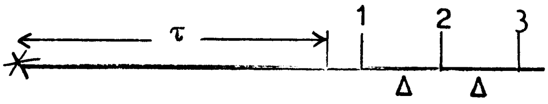
the percentual change of b_o caused by a small positive change ∂p of p is the smaller, the larger the age and the smaller the T_2 value (that is the nearer T_2 to T_1) and the percentual change of $\hat{\tau}$ caused by a small positive change ∂p of p is the larger, the larger the age (provided it being sufficiently large) and the smaller, the larger the T_2 value (that is the larger the distance between T_1 and T_2).

5. More measurements than are strictly necessary

If p in $\beta = c(\tau + T)^{-p}$ would be exactly known a priori, then two measurements b_1 on day T_1 and b_2 on day T_2 (after the sampling day) are sufficient to compute c and τ (the two unknowns) and hence to compute the extrapolation value b_o on day zero ($T = 0$). As soon as there have been made more than two measurements, there are more equations than unknowns and we deal with a fitting problem. Plot the points b_i^{-a} against T_i ($i = 1, 2 \ldots n$) in a Cartesian system, the axes of which contain linear scales, the vertical axis for the value b^{-a} and the horizontal one for the value T.
Would all measurements have been exact, so that the β values would have been found, then the n points would have been situated on one and the same straight line, provided that the basic law $\beta = ct^{-p}$ would hold good. Since the measurements are inaccurate, the points do not lie on the same straight line and the best fitting straight line should be found. The position of this line depends on the degree in which the measurements b_i differ from the corresponding values β_i ($i = 1, 2 \ldots n$) and hence the segment cut from the ordinate axis and the segment cut from the time axis (b_o^{-a} and $\hat{\tau}$) will be inaccurate estimates of the true unknown values β_o^{-a} and τ (zero day's value; age). We wish to compute these inaccuracies among other things as a function of the relative accuracies δ_i of the separate measurements b_i.

5.1 Measurements on three days

The first case treated here is the situation of three measurements made on equidistant days: T_1; $T_2 = T_1 + \Delta$; $T_3 = T_1 + 2\Delta$. Denoting the ordinate with $y = b^{-a}$, the best (least squares) straight line through the three points y_i, T_i, becomes:

(41) 

$$y = \frac{\Sigma_1^3 (T_i-\bar{T})(y_i-\bar{y})}{\Sigma_1^3 (T_i-\bar{T})^2} (T-\bar{T}) + \bar{y}$$

Since $T_2 - T_1 = T_3 - T_2 = \Delta$, we have $\bar{T} \equiv T_2$ and

(42) $$y = \frac{y_3 - y_1}{2} (T - T_2) + \bar{y} .$$

5.1.1 <u>The zero day's value. Estimate.</u>

First of all we want to know the value $y_o (= b_o^{-a})$ for day $T = 0$.

$$y_o = - T_2 \frac{y_3 - y_1}{2\Delta} + \bar{y} = y_1 \left(\frac{3T_1 + 5}{6\Delta}\right) + \frac{1}{3} y_2 - y_3 \left(\frac{3T_1 + \Delta}{6\Delta}\right)$$

(43) $$\boxed{b_o = \left[\frac{3T_1 + 5\Delta}{6\Delta} b_1^{-a} + \frac{1}{3} b_2^{-a} - \frac{3T_1 + \Delta}{6\Delta} b_3^{-a}\right]^{-p}}$$

Hence after measuring b_1 on day T_1, b_2 on day $T_2 = T_1 + \Delta$ and b_3 on day $T_3 = T_2 + \Delta$, the foregoing formula gives the extrapolated zero day's value, provided p being known. A nomogram could avoid calculations.

5.1.2 <u>Accuracy</u>

Next we are interested in the relative accuracy $\delta_o = \sigma(b_o) : \beta_o$ of b_o on the ground of the relative accuracies δ_1, δ_2, δ_3 of the measurements b_1, b_2 and b_3. This b_o is a function of b_1, b_2 and b_3 and hence

(44) $$\sigma^2(b_o) = \left(\frac{\partial b_o}{\partial b_1}\right)^2 \sigma_1^2 + \left(\frac{\partial b_o}{\partial b_2}\right)^2 \sigma_2^2 + \left(\frac{\partial b_o}{\partial b_3}\right)^2 \sigma_3^2$$

for the arguments β_1, β_2, β_3, provided that the separate errors of measurements are uncorrelated. Finally we obtain:

(45) $$\delta_o = \frac{\sqrt{A^2\beta_1^{-2a} + B^2\beta_2^{-a} + C^2\beta_3^{-a}}}{A\beta_1^{-a} + B\beta_2^{-a} - C\beta_3^{-a}}$$

with

$$A = \frac{3T_1 + 5\Delta}{\Delta} \; ; \; B = \frac{1}{3} \; ; \; C = \frac{3T_1 + \Delta}{\Delta}$$

Since the β's are unknown, the b's must be substituted.
Now the basic law gives $\beta_i = c^{-a}(\tau + T_i)$, $i = 1, 2, 3$ and therefore

(46) $$\delta_o = \frac{1}{\tau}\sqrt{A^2(\tau + T_1)^2\delta_1^2 + B^2(\tau + T_2)^2\delta_2^2 + C^2(\tau + T_3)\,\delta_3^2}\;.$$

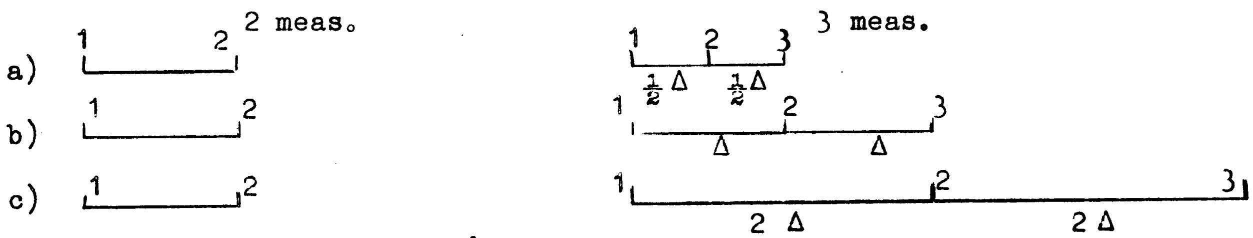
Let us suppose for the sake of simplicity $\tau \gg T_1, T_2, T_3$, then

(47) $$\delta_o \cong \sqrt{A^2\delta_1^2 + B^2\delta_2^2 + C^2\delta_3^2}$$

If moreover $\delta_1 = \delta_2 = \delta_3$, say δ (all measurements relatively equally accurate) then δ_o approximates to

(48) $$\boxed{\delta_o \cong \delta\,\frac{\sqrt{3T_1^2 + 6T_1\Delta + 5\Delta^2}}{\Delta\sqrt{6}}}$$

This is the moment for a comparison with δ_o in the case of two measurements and $\delta_1 = \delta_2$, say δ, and $\tau \gg T_1, T_2$.
How should this comparison be made ?
There are many possibilities.



and so on.

(49) a) In case a) $(\delta_o)_2 = \dfrac{T_1 + \Delta}{\Delta}\,\delta\sqrt{2}$ and $(\delta_o)_3 = \dfrac{2\sqrt{3T_1^2 + 3T_1\Delta + \frac{3}{4}\Delta^2}}{\Delta\sqrt{6}}\,\delta$

and consequently $(\delta_o)_3 < (\delta_o)_2$

(50) b) In case b) $(\delta_o)_2$ is as above and $(\delta_o)_3 = \dfrac{\sqrt{3T_1^2 + 6T_1\Delta + 5\Delta^2}}{\Delta\sqrt{6}}\,\delta$

and again $(\delta_o)_3 < (\delta_o)_2$, now in a more marked way.

(51) c) In case c) $(\delta_o)_2$ is as above and $(\delta_o)_3 = \dfrac{\sqrt{3T_1^2 + 12T_1\Delta + 20\Delta^2}}{2\,\Delta\sqrt{6}}\,\delta$

and $(\delta_o)_3 < (\delta_o)_2$ in a still more markedly degree.

Case a) may be the most illustrative: in what degree becomes δ_o smaller if halfway between the measuring days of the two measurements method a third measurement would have been made ?
Compute the ratio $K = (\delta_o)_3 : (\delta_o)_2$

(52) $$K^2 = \frac{12m^2 + 12m + 5}{12m^2 + 24m + 12} \qquad m = T_1/\Delta$$

For decreasing m (increasing Δ; fixed T_1) K decreases asymptotically to $0.65 = \sqrt{5/12}$. For $T_1 = 4$; $\Delta = 8$ ($T_2 = 12$); $K = 0.72$

Hence in case a) δ_o be decreased to 65% of the value in the 2-measurements method, for sufficiently large Δ (provided $\tau \gg T_1, T_2$), by making a measurement halfway.
Next consider case b).

Now $$K^2 = \frac{3m^2 + 6m + 5}{12(m+1)^2} \rightarrow \frac{5}{12} = 0.42$$ and $K \rightarrow 0.64$ if $m \rightarrow 0$

Again K decreases asymptotically to $\sqrt{5/12}$ for increasing Δ. For $T_1 = 4$, $T_2 = 12$, then $K = 0.57$.
Next consider case c).

Now $$K^2 = \frac{m^2 + 4m + 6}{20(m+1)^2} \rightarrow \frac{3}{10}$$ if $m \rightarrow 0$.

Again K decreases to $\sqrt{5/12}$ for increasing Δ. For $T_1 = 4$, $T_2 = 12$ we have $K = 0.50$.

5.1.3 Estimate $\hat{\tau}$

Substituting T = 0 in the equation for y, (see (42)), we obtain:

(53) $$\boxed{\hat{\tau} = 2\frac{\bar{y}}{y_3 - y_1}\Delta - T_2} \qquad \text{with } y_i = b_i^{-a}; \quad a = 1/p$$

Hence obtaining the values b_1 on day T_1, b_2 on day $T_2 = T_1 + \Delta$ and b_3 on day $T_3 = T_1 + 2\Delta$, the above mentioned formula gives the estimate of the age of the activity. A nomogram could avoid calculations.

5.1.4 Accuracy $\sigma(\hat{\tau})$

We are also interested in the relative accuracy $\delta_\tau = \sigma(\hat{\tau}) : \tau$ of $\hat{\tau}$. Using the approximation $\hat{\tau} \cong \frac{2}{3}\Delta \frac{y_1 + y_2 + y_3}{y_3 - y_1}$ (suppose $\tau \gg T_1, T_2, T_3$) we finally come to

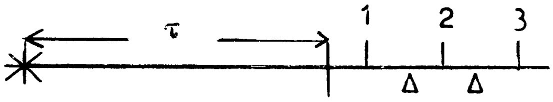
(54) $$\delta_\tau^2 = \frac{a^2 c^{4a}}{36(\tau + T_1 + \Delta)^2 \Delta^2} [\beta_1^{-2a}(3\tau + 3T_1 + 5\Delta)^2 + \beta_2^{-2a}\Delta^2\delta_2^2 + \beta_3^{-2a}(3\tau + 3T_1 + 5\Delta)^2\delta_3^2]$$

Next substituting $\beta_i^{-a} = c^{-a}(\tau + T_i)$, then δ_τ approximates to:

(55) $$\delta_\tau \cong \frac{a\tau\sqrt{\delta_1^2 + \frac{4}{9}\Delta^2/\tau^2\delta_2^2 + \delta_3^2}}{2\Delta}$$

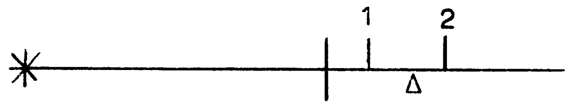
Suppose $\delta_1 = \delta_2 = \delta_3$, say δ, then

(56) $$\delta_\tau \cong \frac{a\tau}{2\Delta}\delta\sqrt{2 + \frac{4}{9}\Delta^2/\tau^2} \cong \frac{a\tau\sqrt{2}}{2\Delta}\delta$$



Next compare this result with the two measurements method, in which case we obtained

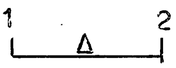
(57) $$\delta_\tau = \frac{a\tau\sqrt{2}}{\Delta}\delta$$



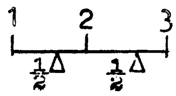
Note that δ_τ does not contain T_1.

Next we make comparisons as we did in the section on δ_0.

Case a):



$(\delta_\tau)_2 \cong \frac{a\tau\sqrt{2}}{\Delta}$

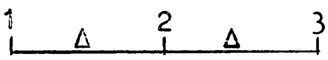


$(\delta_\tau)_3 \cong \frac{a\tau\sqrt{2}}{\Delta}$

Hence there is no difference.

Case b):

$(\delta_\tau)_2 \cong \frac{a\tau\sqrt{2}}{\Delta}$



$(\delta_\tau)_3 = \frac{a\tau}{\Delta\sqrt{2}}\delta$

Here $(\delta_\tau)_3 : (\delta_\tau)_2 = \frac{1}{2}$.

Case c): 1 ⊢ Δ ⊣ 2   1 ⊢ 2Δ ⊣ 2 ⊢ 2Δ ⊣ 3

$$(\delta_\tau)_2 \simeq \frac{a\tau\sqrt{2}}{\Delta}\,\delta \qquad\qquad (\delta_\tau)_3 = \frac{a\tau}{2\,\Delta\sqrt{2}}\,\delta$$

Consequently $(\delta_\tau)_3 : (\delta_\tau)_2 = \frac{1}{4}$.

Conclusion: if one wants to obtain a more accurate estimate $\hat{\tau}$ of the age τ by using three instead of two measurements (b_1, b_2, b_3 instead of b_1, b_2) of the pure filteractivity, then the third measurement b_3 should not be made on a day T_3, situated half way between T_1 and T_2 (that is $T_3-T_1 = T_2-T_3$), but on a day T_3, situated so that $T_3-T_2 = T_2-T_1$ (not 1 ⊢ 3 ⊣ 2 but 1 ⊢ 2 ⊣ 3)

Although the relative accuracies, denoted by $(\delta_\tau)_2$ and $(\delta_\tau)_3$, of the age estimates $\hat{\tau}_2$ and $\hat{\tau}_3$, found in both methods (b_1, b_2 resp. b_1, b_2, b_3), is constant (nearly 0.5), the accuracy $(\delta_\tau)_3$ is inversely proportional with the equidistance defined by $T_2-T_1 = T_3-T_2$.

In this conclusion all measurement are supposed to be equally accurate ($\delta_1 = \delta_2 = \delta_3$).

5.1.5 <u>Numerical example</u>

De Bilt 6/7-1-1959; $T_1 = 4$; $b_1 = 52.4$; $\delta_1 = 0.030$; $T_2 = 12$; $b_2 = 46.7$; $\delta_2 = 0.031$ (say $\delta_1 = \delta_2 = 0.030$).

Consequently $b_o = 57.1$, with $\delta_o = 0.064$ and $\hat{\tau} = 65$ days, with $\delta_\tau = 0.42$.

Suppose we would make a third measurement, on day $T_3 = 20$ (so that the equidistance $T_2 - T_1 = T_3 - T_2 = 8$ days). What are the new values of δ_o and δ_τ ?

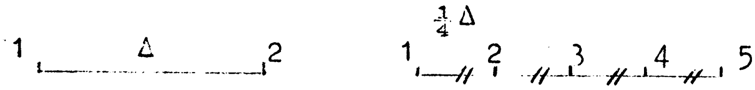
We obtain $(\delta_o)_3 : (\delta_o)_2 = 0.57$ (see above), so that $(\delta_o)_3 = 0.57 \times 0.064 = 0.036$ and $(\delta_\tau)_3 : (\delta_\tau)_2 = 0.50$ and $(\delta_\tau)_3 = 0.50 \times 0.42 = 0.21$.

5.2 <u>Measurements on five days</u>

A next step could be to start from 5 equidistant measurements, so that $T_2 = T_1 + \Delta$; $T_3 = T_1 + 2\Delta$; $T_4 = T_1 + 3\Delta$; $T_5 = T_1 + 4\Delta$ and $\bar{T} = T_3$.

Let be again $\tau \gg T_1, \ldots . T_5$ and $\delta_1 = \ldots\ldots \delta_5$, say δ , then we obtain:

(58) $$\delta_o = \frac{\sqrt{T_1^2 + 4\Delta T_1 + 6\Delta^2}}{\Delta\sqrt{10}}\,\delta$$

Then in the case 

We would have

(59) $$(\delta_o)_2 = \frac{T_1+\Delta}{\Delta}\,\delta\sqrt{2};\qquad (\delta_o)_5 = \frac{\sqrt{T_1^2 + \frac{4}{5}\Delta T + \frac{6}{25}\Delta^2}}{\Delta\sqrt{10}}\,\delta$$

Substituting $\frac{1}{4}\Delta$ for Δ in (58) we obtain

(60) $$K^2 \equiv \left[\frac{(\delta_o)_5}{(\delta_o)_2}\right]^2 = \frac{12m^2 + 12m + 4\frac{1}{2}}{15(m+1)^2} \rightarrow \frac{3}{10} \quad \text{and } K \rightarrow 0.55, \text{ if } m \rightarrow 0$$

For increasing Δ the K decreases to $\sqrt{6/20} = 0.55$

<u>Conclusion: when making 5 equidistant measutements during the total period fixed for the two measurements method, the δ_o value would decrease with almost 45% (3 eq. meas.: 35%).</u>

5.3 Measurements on more than five days

It is interesting to see what the influence on both the accuracy of the zero day's value and the age would be, if still more measurements on equidistant days would have been made within the period between day T_1 and day T_2 fixed for the two measurements method. The problem can be attacked rather easy for δ_o, but is difficult with regard to τ. For this reason only δ_o is considered here.

First of all we refer to a statistical theorem. Suppose to each of a set of n values of x_i ($i = 1, 2.... n$) belongs a distribution of y values, each distribution possessing the same known standarddeviation σ.

Suppose that the mean value of y, denoted with η, is lineair related to x according to $\eta = A + B(x-\bar{x})$, with A and B unknown constants. Now let be observed only one y_1 value corresponding with x_1, only one y_2 value corresponding with x_2 etc.

Usually these points x_i, y_i, do not lie on one and the same straight line.

It is possible to prove that the statistical best estimate of A is given by $\hat{A} = \bar{y} = \frac{1}{n} \sum_i^n y_i$; the statistical best estimate of B is given by

(61) $$\hat{B} = \sum_1^n (x_i-\bar{x})(y_i-\bar{y}) : \sum_1^n (x_i-\bar{x})^2;\quad \bar{x} = \frac{1}{n}\sum_1^n x_i$$

With these values $\hat{A}$ and $\hat{B}$ the straight line can be drawn and for each given value g of x the corresponding mean value y_g of y can be read from the straight line. Since this straight line lies statistically (because $\hat{A}$ and $\hat{B}$ are stochastical quantities as being based on the statistical variables $\underline{y}_1 \ldots \underline{y}_n$), also $\underline{y}_g$ is stochastical. The variance $\sigma^2(\underline{y}_g)$ is given by

$$\sigma^2(\underline{y}_g) = \sigma^2 \left[\frac{1}{n} + \frac{(x_g - \bar{x})^2}{\Sigma (x - \bar{x})^2} \right] \tag{62}$$

We want to use this formula in our problem.
Identify $\underline{y}$ with $\underline{b}^{-a}$ (a being known) and x with T and think of an <u>odd</u> total number of measuring points $n = 2m + 1$; $m = 1, 2, \ldots$ Now σ^2 is the variance of $\underline{y}$ and hence not of $\underline{b}$.
Since $\sigma^2(\underline{y}_i) = \sigma^2(b_i^{-a}) = a^2\beta_i^{-2a}(\sigma_i^2/\beta_i^2) = a^2\beta_i^{-2a}\delta_i^2$, the values $\sigma^2(y_i)$ with $i = 1, 2 \ldots n$, can not be equal if the values of δ_i are. Nevertheless we will approximate all values $\sigma^2(\underline{y}_i)$ by $a^2\beta_o^{-2a}\delta^2$, which approximation is the better the smaller the slope of the line y against T_i, that is the larger τ (here appears the assumption $\tau >> T_1, T_2 \ldots T_n$). Next we consider the ordinate of the straight line for T = 0. This point, mentioned $\underline{y}_o = b_o^{-a}$, possesses a variance given by (see formula (62); substitute x = q).

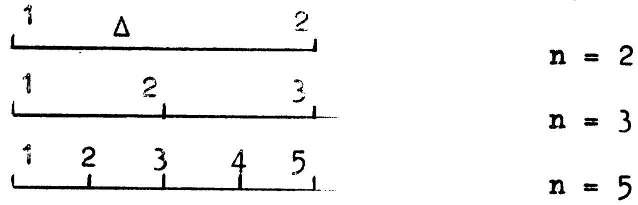
$$\sigma^2(y_o) = a^2\beta_o^{-2a}\delta^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\Sigma(x - \bar{x})^2} \right] \tag{63}$$

whence

$$\boxed{\delta_o^2 = \delta^2 \left[\frac{1}{n} + \frac{\bar{T}^2}{\Sigma_1^n (T_i - \bar{T})^2} \right]} \tag{64}$$

Now we will distinguish two cases A and B.

A.



etc.

Here

$$(\delta_o)^2_n = \frac{12m^2 + 12m + 2(2n-1)/(n-1)}{n(n+1)/(n-1)}\ \delta^2 \quad , \text{ with } m = \frac{T}{\Delta}$$

$n = 3, 4, 5, 6, 7...$

$$(\delta_o)^2_2 = 2(m+1)^2\delta^2 \quad \text{if} \quad \tau \gg T_1$$

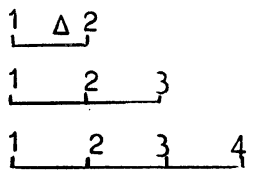
So that

$$K^2_A = (\delta_o)^2_n : (\delta_o)^2_2 = \frac{12m^2 + 12m + \frac{2(2n-1)}{n-1}}{\frac{2n(n+1)}{n-1}(m+1)^2}$$

For $m \to o$ (that is $\Delta \to \infty$), then $K^2 \to \frac{2n-1}{n(n-1)} \underset{n \to \infty}{\to} 2/n$

Increasing Δ, decreasing K_A .

n	K^2_A	K^2_A for $m \to 0$	K_A
3	$(12m^2+12m+5) : 12(m+1)^2$	5/12	0.64
5	$(12m^2+12m+4\frac{1}{2}) : 15(m+1)^2$	$4\frac{1}{2}/15$	0.55
7	$(12m^2+12m+4\frac{1}{3}) : 18.7(m+1)^2$	4.33/18.7	0.48
21	$(12m^2+12m+4.1) : 44.2(m+1)^2$	4.1 /44.2	0.30
∞	$(12m^2+12m+4.0) : \infty$	$4.0/\infty = 0$	0

B. 

n = 2

3

4

etc.

$$(\delta_o)^2_n = \frac{12m^2 + 12(n-1)m + 2(2n-1)(n-1)^2}{n(n^2-1)}\ \delta^2 \quad , \text{ with } m = T/\Delta$$

$n = 3, 4, 5...$

$$(\delta_o)^2_2 = 2(m+1)^2\delta^2 \quad \text{if} \quad \tau \gg T_1$$

Hence

$$K^2_B = (\delta_o)^2_n : (\delta_o)^2_2 = \frac{12m^2 + 12(n-1)m + 2(2n-1)(n-1)}{2n(n^2-1)(m+1)^2}$$

For $m \rightarrow 0$ (that is $\Delta \rightarrow \infty$), $K_B^2 \rightarrow \frac{2n-1}{n(n+1)} \underset{n \rightarrow \infty}{\rightarrow} 2/n$

Increasing Δ, decreasing K_B

n	K_B^2	K_B^2 for $m \rightarrow 0$	K_B
3	$(12m^2+24m+20) : 48(m+1)^2$	5/12	0.64
5	$(12m^2+48m+72) : 240(m+1)^2$	3/10	0.55
7	$(12m^2+72m+156) : 672(m+1)^2$	156/672	0.48
21	$(12m^2+240m+1640) : 70560(m+1)^2$	1640/70560	0.30
∞	$\infty : \infty$	0	0

For each m and each n one obtains:

$$(K_A)_{n,m} > (K)_{n,o} > (K_B)_{n,m}$$

Consequently: the larger the number of equidistant measurements and the larger the equidistance, the more accurate the zero day's value, provided that all measurements are equally accurate. For sufficiently large equidistance and sufficiently large number of measurements the decrease of the accuracy of the extrapolated value is inversely proportional with the square root of the number of measurements.

.-.-.-.-.

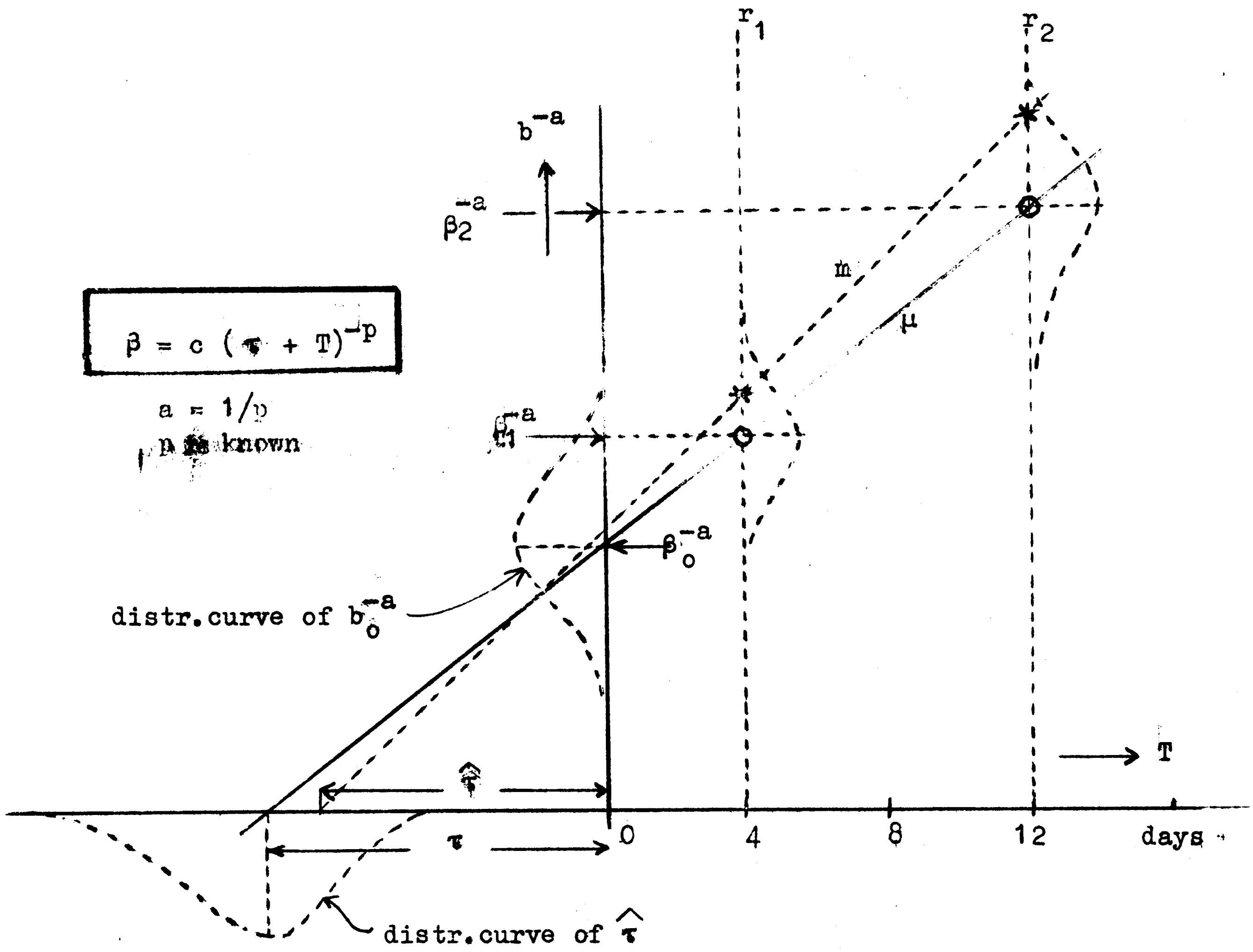
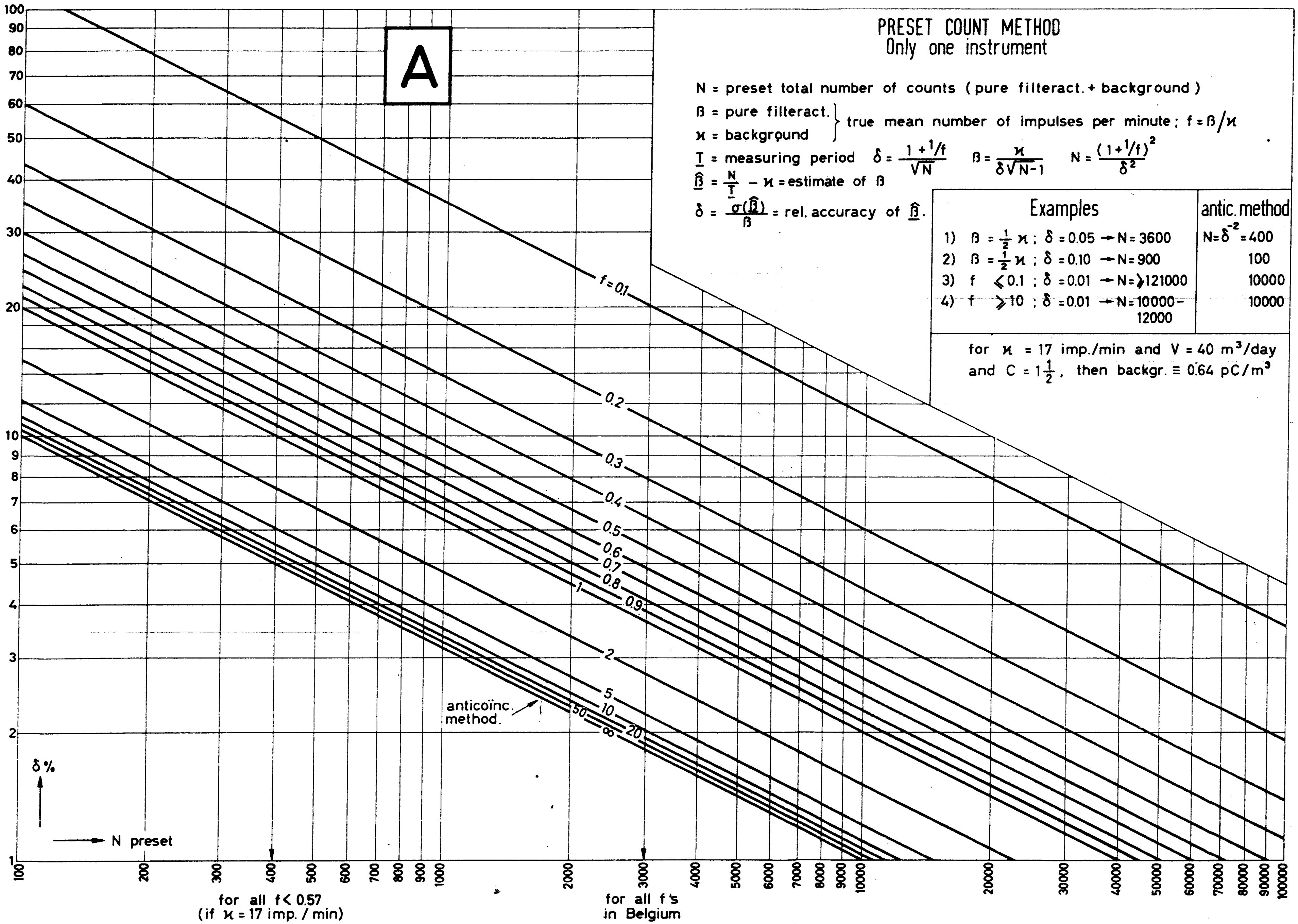


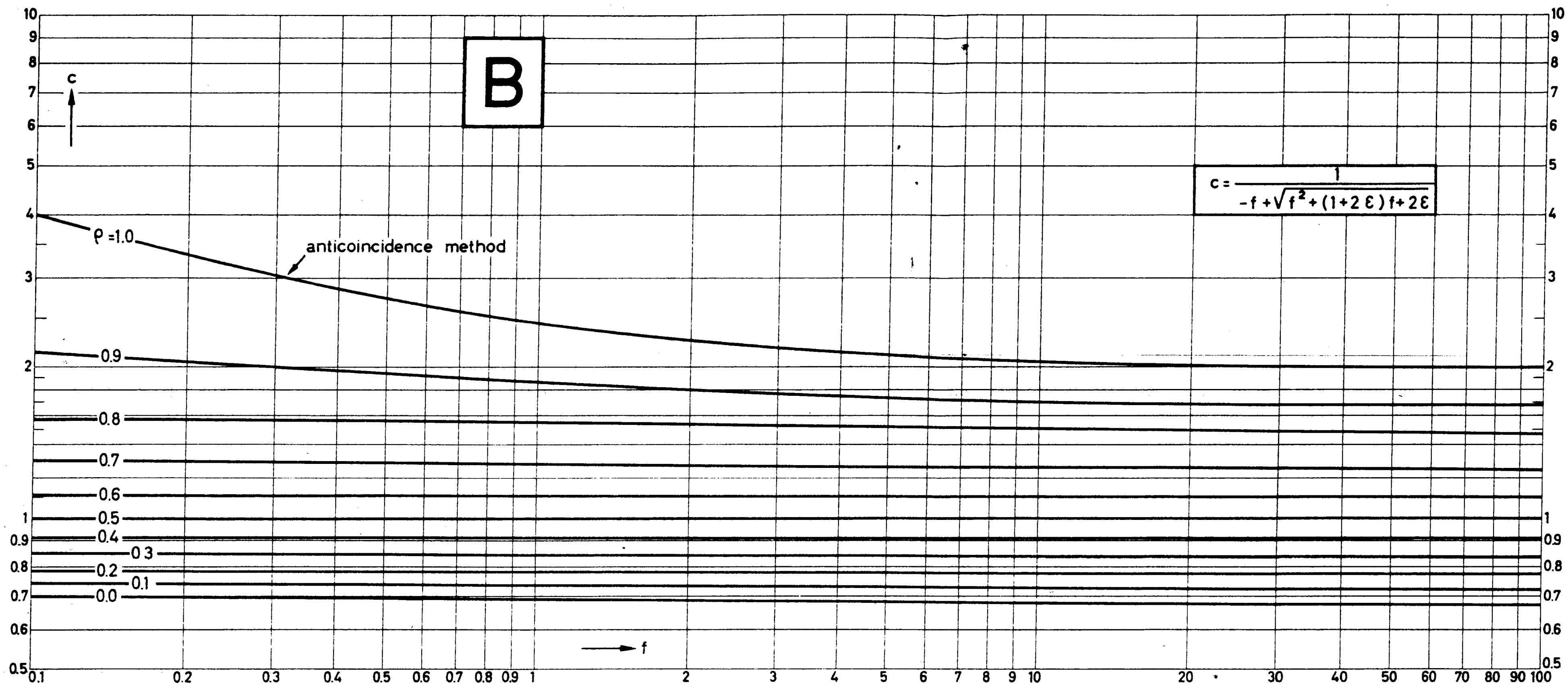
Fig. 1

Explanation:

Linear scales for b^{-a} and T; p is supposed to be known. The true points o T, β^{-a}, e.g. 4, β_1^{-a} and 12, β_2^{-a}, determine the true straight line μ, which cuts the vertical axis in β_0^{-a} and the horizontal axis in τ; here β_0 and τ are resp. the zero day's true filteractivity and the "age". Since the measured values b_1^{-a} and b_2^{-a}, furnishing the crosses in the figure, are stochastically distributed around the true values β_1^{-a} and β_2^{-a} (furnishing the small circles) on the verticals r_1 and r_2), the straight line m through these two points lies stochastically around μ, so that the points of intersection with the vertical and horizontal axes lie stochastically around the mean positions β_0^{-a} and τ respectively.

.-.-.-.-.


A
PRESET COUNT METHOD
Only one instrument
N = preset total number of counts (pure filteract. + background)
ß = pure filteract.
ϰ = background
true mean number of impulses per minute; f = ß/ϰ
T = measuring period δ = (1 + 1/f)/√N ß = ϰ/(δ√(N-1)) N = (1 + 1/f)²/δ²
ß̂ = N/T − ϰ = estimate of ß
δ = σ(ß̂)/ß = rel. accuracy of ß̂.
Examples
1) ß = ½ ϰ ; δ = 0.05 → N = 3600
2) ß = ½ ϰ ; δ = 0.10 → N = 900
3) f ≤ 0.1 ; δ = 0.01 → N = ≥121000
4) f ≥ 10 ; δ = 0.01 → N = 10000–12000
antic. method
N = δ⁻² = 400
100
10000
10000
for ϰ = 17 imp./min and V = 40 m³/day and C = 1½, then backgr. ≡ 0.64 pC/m³
f = 0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
2
5
10
20
50
∞
anticoïnc. method
δ %
N preset
for all f < 0.57 (if ϰ = 17 imp. / min)
for all f's in Belgium



PRESET TIME METHOD

$\beta + \varkappa$ and $\varkappa'$ are measured simultaneously and separately during T minutes.
Let be ρ = correlation coefficient between the total numbers $\underline{N}$ and $N'_{\varkappa}$ in the period of T minutes as unit period and let be $\varepsilon = 1 - \rho$

Example: suppose $\rho = 0.8$; $\varkappa = \varkappa' = 17$ imp./min.

1) Given: $f = 10$ (high activity); $\delta = 0.10$. What is the value of T ?
   Answer: $f = 10$, $\rho = 0.8$, see graph B. Then $c = 1.5$, $f^* = f.c = 15$
   Graph A, with $f = 15$ and $\delta = 0.10$, gives $N = 117$ (the one instrument method gives $N = 120$). T will be $117 : 11\varkappa = 0.6$ min.
2) Given: $f = 0.5$; $\delta = 0.05$. What is the value of T ?
   Answer: $f = 0.5$, $\rho = 0.8$, see graph B. Then $c = 1.55$, $f^* = f.c = 0.78$. Graph A, with $f = 0.78$ and $\delta = 0.05$, gives $N = 2100$ (one instr. method: 3600)
   T will be $2100 : 1\frac{1}{2}\varkappa = 82$ min.
3) Given: $f = 0.1$ (low activity); $\delta = 0.05$. What is the value of T ?
   Answer: $f = 0.1$, $\rho = 0.8$, see graph B. Then $c \cong 1.6$, $f^* = f.c = 0.16$. Graph A, with $f = 0.16$ and $\delta = 0.05$, gives $N = 25000$ (one instr. method : 50000)
   T will be $25000 : 1.1\varkappa = 22$ h

Two instruments

$$\delta = \frac{\sqrt{1 + 2\varepsilon/f}}{\sqrt{\beta T}} = \frac{1 + 1/f^*}{\sqrt{N}} \; ; \; f^* = f.c \; ; \; f = \frac{\beta}{\varkappa}$$

$$\beta = \frac{1 + \sqrt{1 + 8\varkappa\varepsilon T\delta^2}}{2T\delta^2} \; ; \; \hat{\beta} = \frac{\dot{N} - N'_{\varkappa}}{T}$$

$$T = \frac{1 + 2\varepsilon/f}{\beta\delta^2} \; ; \; \varepsilon = 1 - \rho$$

NOMOGRAM

for estimating the "zero day's" atmospheric radioactivity, the "age" of the radioactivity and their accuracies, if measurements of the filteractivity on only two separate days have been made.

D

Basic principle

$\beta = \text{const.}\ t^{-1.20}$

$t = \tau + T$

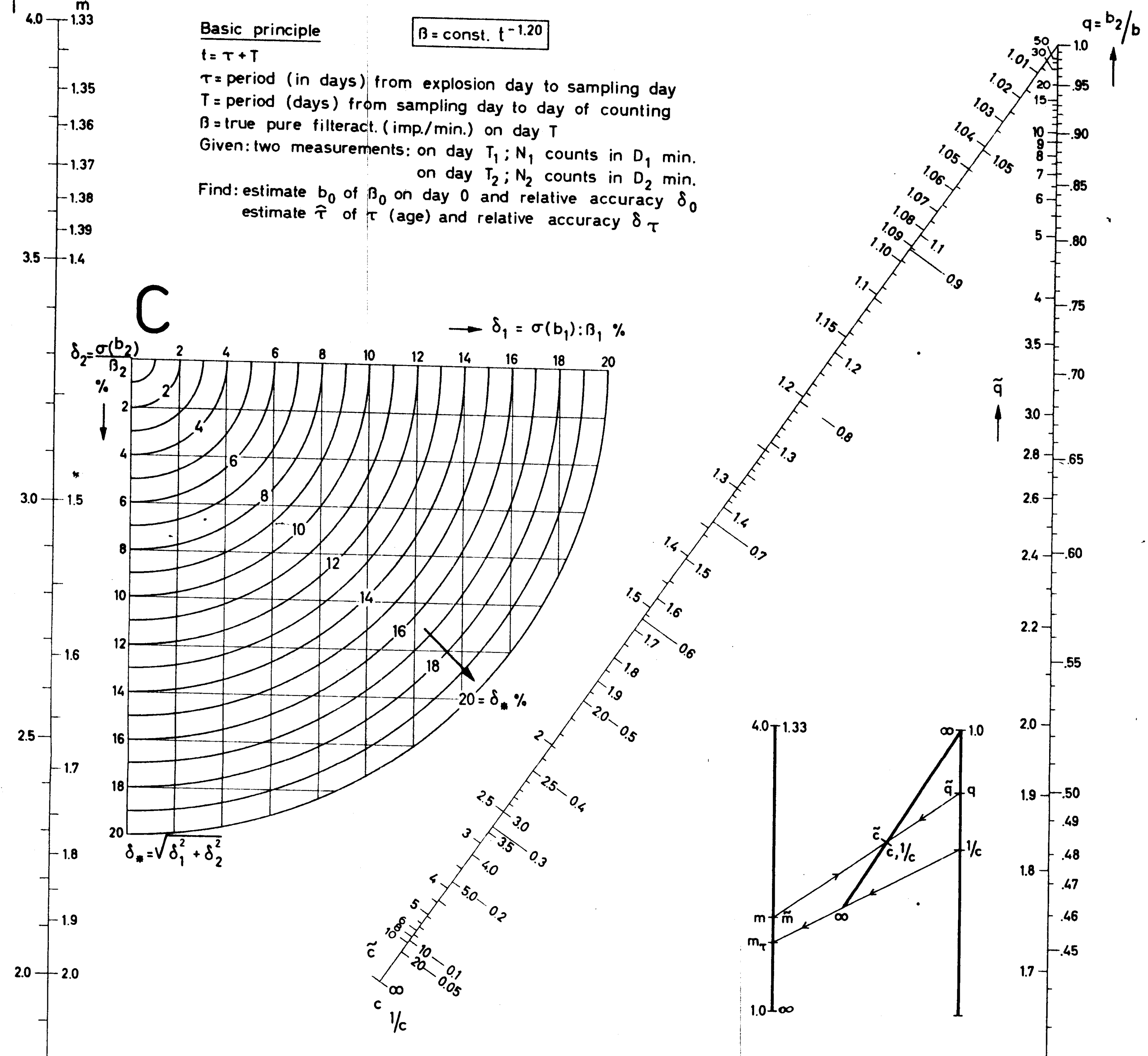
τ = period (in days) from explosion day to sampling day

T = period (days) from sampling day to day of counting

β = true pure filteract. (imp./min.) on day T

Given: two measurements: on day T_1; N_1 counts in D_1 min.
on day T_2; N_2 counts in D_2 min.

Find: estimate b_0 of β_0 on day 0 and relative accuracy δ_0
estimate $\hat{\tau}$ of τ (age) and relative accuracy δ_τ



Numerical example.

Given: $\varkappa = 18.0$ i/min (backgr.); $T_1 = 4$; $N_1 = 9000$; $D_1 = 80$ min.
$T_2 = 12$; $N_2 = 9000$; $D_2 = 100$ min.

Find: b_0, δ_0; $\hat{\tau}$, δ_τ

Answer: $b_1 = \frac{N_1}{D_1} - \varkappa = 94.4$; $f_1 = b_1/\varkappa = 5.25$; $b_2 = \frac{N_2}{D_2} - \varkappa = 72.0$; $f_2 = b_2/\varkappa = 4.00$

Graph A: $(N_1, f_1) \rightarrow \delta_1 = 0.013$; $(N_2, f_2) \rightarrow \delta_2 = 0.013$

Graph C: $(\delta_1, \delta_2) \rightarrow \delta_* = 0.018$

Graph D: $q = b_2/b_1 = 0.764$; $\tilde{q} = 4.2$; $m = T_2/T_1 = 3.0$; $\tilde{m} = 1.5$

$(q, m) \rightarrow c = 1.176$; $\tilde{c} = 1.14$; $1/c = 0.850$; $(q = 0.850, c = \infty) \rightarrow m_\tau = 1.15$

Hence

$b_0 = cb_1 = 110.8$ imp./min $\delta_0 = \tilde{c}.\tilde{m}.\delta_* = 0.031$	$\hat{\tau} = T_1 : (m_\tau - 1) = 26.7$ days $\delta_\tau = \tilde{c}.\tilde{q}.\delta_* = 0.086$
$104 < \beta_0 < 118$ imp./min.	$22 < \tau < 31$ days
$104 ; 118 = 110.8\ (1 \mp 2 \times 0.031)$	$22 ; 31 = 26.7\ (1 \mp 2 \times 0.086)$

95% reliability regions

ADDENDUM C

The reasoning on page 36 is based on the requirement that for each instrument the accuracy $\sigma(\widehat{\Delta A}_j)$ of the systematic error $\widehat{\Delta A}_j$ should not exceed a pregiven value d (e.g. 0.05), this d being the same for all instruments. Then the expression (82) results.

It is, however, also possible to require that the accuracy $\sigma(\widehat{\Delta A}_j)$ of the systematic error $\widehat{\Delta A}_j$ is at the most a pregiven fraction 1/f of the standarddeviation of the total random error for each apparatus separately, this fraction being the same for all instruments. Then there results

$\sigma_j : \sigma(\widehat{\Delta A}_j) \geqslant f$, in stead of $\sigma(\widehat{\Delta A}_j) \leqslant d$ for each j.

Substituting (74) and taking into account $\sigma_1 < \sigma_2 \ldots\ldots < \sigma_k$, we come to

$$(82a) \quad n \geqslant \frac{k-1}{k} f^2 \left[\frac{\bar{\sigma}^2 + (k-2)\sigma_1^2}{(k-1)\sigma_1^2} \right] = n^* \text{ ; e.g. } f = 3 \text{ or } 5.$$

As soon as $\sigma_1 = \sigma_2 = \ldots \sigma_k$, say σ, we obtain $n^* = \frac{k-1}{k} f^2$, as in (68). This reasoning includes a stronger requirement than on the pages 42, 43. Hence the values, estimated for the minimum durations, become larger. Let us give 5 examples, in which 10 instruments have been installed close to each other. Let the σ's be unknown beforehand, although there is some a priori knowledge on the interrelation. Take f = 5. For the interrelation we substitute respectively:

I. $\sigma_1 : \ldots\ldots \sigma_{10} \cong 1 : 1 : \ldots 1$. Equal σ's. Then $n^* = 22$ days.

II. $\sigma_1 : \ldots\ldots \sigma_{10} \cong 1 : 1 : \ldots 1 : 10$. Then $n^* = 47$.

III. $\sigma_1 : \ldots\ldots \sigma_{10} \cong 1 : 1\frac{4}{9} : \ldots 5$ (arithmetic series). Then $n^* = 76$.

IV. $\sigma_1 : \ldots\ldots \sigma_{10} \cong 1 : 2 : 3 \ldots 10$ (arithmetic series). Then $n^* = 116$.

V. $\sigma_1 : \ldots\ldots \sigma_{10} \cong 1 : 10 : 10 \ldots 10$ Then $n^* = 245$ days.

Conclusion: the "larger the inequality" between the σ's, the longer the minimum period of comparison.

Returning to (82a) we state that there are k values of n^*: $n_1^*, n_2^* \ldots\ldots n_k^*$, if the requirement is k fold.

$$(82b) \quad n \geqslant \frac{k-1}{k} f^2 \left[\frac{\bar{\sigma}^2 + (k-2)\sigma_j^2}{(k-1)\sigma_j^2} \right] : j = 1, 2 \ldots\ldots k$$

Since $\sigma_1 < \sigma_2 \ldots < \sigma_k$, we have $n_1^* > n_2^* > \ldots > n_k^*$. We could agree to choose the mean value $n^* = \frac{1}{k} \Sigma n_j^*$, so that n^* will depend less strongly on one or more extremely small or large σ's than n_1^*.

ADDENDUM D

In case the proportionality constant $\underline{C}_j$ (see pg.1, 3, 5) is not constant in time (e.g. because of the fact that the daily filters are not strictly identical), but varies day by day, in such a way that it is composed of a constant part, say $\mathring{C}_j$, and a random part $\underline{C}_{ij}$, then the reasoning is as follows. Let be $\underline{d}_{ij} = \mathring{C}_j - \underline{C}_{ij}$ for apparatus j on day i. Then we have

$$\underline{z}_{ij} = \widehat{C}_j \, \widehat{\underline{\alpha}}_{ij} = (\underline{C}_{ij}\alpha_{ij}) \frac{\widehat{C}_j \, \widehat{\underline{\alpha}}_{ij}}{\underline{C}_{ij}\alpha_{ij}} = (u_i \frac{\underline{u}_{ij}}{u_i}) \frac{\widehat{C}_j \, \widehat{\underline{\alpha}}_{ij}}{\underline{C}_{ij}\alpha_{ij}}$$

Now

$$\lg \frac{\widehat{C}_j}{\underline{C}_{ij}} \cong \frac{\widehat{C}_j - \underline{C}_{ij}}{\underline{C}_{ij}} = \frac{\widehat{C}_j - (\mathring{C}_j - \underline{d}_{ij})}{\mathring{C}_j - \underline{d}_{ij}} \cong \frac{\widehat{C}_j - \mathring{C}_j}{\mathring{C}_j} + \frac{\underline{d}_{ij}}{\mathring{C}_j}$$

Define $\underline{t}_{ij} = \underline{d}_{ij}/\mathring{C}_j$, with $\mathcal{E}\,\underline{t}_{ij} = 0$ and $\mathring{C}_j^2 \, \sigma^2(\underline{t}_{ij}) = \sigma^2(\underline{d}_{ij})$

Hence

$$\underline{z}_{ij} = u_i \exp [A_j + \underline{E}_{ij}] \text{, with } \underline{E}_{ij} = \underline{e}_{ij} + \underline{B}_{ij} + \underline{t}_{ij} \text{ in stead of}$$

$$\underline{E}_{ij} = \underline{e}_{ij} + \underline{B}_{ij}.$$

An additive random variable $\underline{t}_{ij}$ has appeared, which refers to the random part of $\underline{C}_j$ (at least under the assumption that the not constant part of $\underline{C}$ behaves as a random variable). Then the variance of the total error $\underline{E}_{ij}$ is

$$\sigma^2(\underline{E}_{ij}) = \sigma^2(\underline{e}_{ij}) + \sigma^2(\underline{B}_{ij}) + \sigma^2(\underline{t}_{ij})$$

In general: if there are m independent sources of random errors in total (sampling error, counting error, filter error etc.) then

$$\sigma^2(\text{total}) = \sum_m \sigma_m^2.$$

<u>Remark</u> with respect to 5.2.1.2

The author knows that the attack in 5.2.1.2 as to the question of the definition of "sufficiently high" is not correct in a strictly statistical sense, but it is more illustrative in a visual way (see the sketch).

When analyzing the results of the simultaneous measurements proposed by Dr. Grandjean, we could do it better, if desirable, e.g. in the following way: the variance of the corrected value y^*, defined by $y_j + \widehat{\Delta A}_j$, is $\sigma^2 + \frac{k-1}{kn}\sigma^2$ and the difference between the uncorrected y_j and the corrected y_j^* may be called significant if

$$\left|\Delta A_j\right| > 2\sigma\sqrt{1 + \frac{k-1}{kn}} \cong 2\sigma\left(1 + \frac{k-1}{2kn}\right)$$

pg.	line from top (t)/ from bottom (b)		delete	insert
6	19	t	$C_i a_{ij}$	$C_j a_{ij}$
	25	t	$\frac{\widehat{C_{ij}}}{C_j}$	$\frac{\widehat{C_j}}{C_j}$
13	8	b	$(1 - \frac{1}{k})\sigma_{13}^2$	$(1 - \frac{1}{k})^2\sigma_{13}^2$
16	3	b	only	
10	7	b	$\frac{1}{k}\sum_{m\neq j}^{k}\sigma^2(E_m)$	$\frac{1}{k^2}\sum_{\substack{m=1\\m\neq j}}^{k}\sigma^2(E_m)$
19	1,2,3	t	$\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3$	$\hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\sigma}_3^2$
24	2	b	28	24
25	7	t	28	24

pg.	line from top (t) / from bottom (b)		delete	insert
30	9	b	accuracy	systematic error
33	19	t	accuracy	systematic error
34	sketch, in the lower 4σ interval		b	b^1
36	6	t	σ_1^2	σ_j^2
37	12	t	$\frac{F^2 k}{d^2}$	$\frac{F_k^2}{d^2}$
45	3	t	$\sigma_2 = 0. = 0$	$\sigma_2 = 0.10$
46	9	t	qunatities	quantities
	10	t	k	κ
47	5	b	$\delta_o = \tilde{c}.\tilde{m}.$	$\delta_o = \tilde{c}.\tilde{m}.\delta.$
	9	b	2.58	3.1
	2,4,	t	k	κ
	10, 11, 12,	b	k_1 k_2	κ_1 κ_2
53	18	b	s_B-values for σ_B)	s_B-values) for σ_B
56	2	t	s_E	s_e
43	10	t	estimated.	estimated. Since $f= \bar{\sigma}/d$, see (82), the requirement f=5 implies that the period of comparison

pg.	line from top (t) / from bottom (b)		delete	insert
43	10	t		is chosen so long, that all accuracies of the systematic errors should not exceed certain value, which is at least 5 times as small as the mean value of the standard deviations of total random errors of the participating instruments.
3A	4	t	$\hat{\underline{N}}$	$\hat{\underline{N}}_\beta$
6A	14 18	t t	on the on the	on an on an
13A	17	b	became	came
16A	13	t	B is	B (note: κ' and κ'' are replaced by κ and κ')is
4B	3	t	stochastrical	stochastical
8B	3 18	t t	day $b_1(1+2\delta_1)$	say $b_2(1+2\delta_2)$
11B	3	t	The	Then
15B	1	t	Using	Use
21B	8	t	be	can be
23B	10	t	Although	Although the ratio of
26B	3	t	3,4,5,6,7...	3,5,7....