GEODÆTISK INSTITUT

MEDDELELSE NO. 48

THE PROBABILISTIC BACKGROUND OF SOME STATISTICAL METHODS IN PHYSICAL GEODESY

BY

STEFFEN LILHOLT LAURITZEN

KØBENHAVN · DANMARK

1973

Contents

1. Introduction and Summary

The anomalous potential of the earth is described as a gaussian stochastic process $\xi(t, \omega)$ where $t \in M_R$, the space of points outside a sphere concentric with the earth and with a radius R, so that it is contained in the interior of the earth. The mean value function is set to be identically zero and the covariance function has to be estimated.

It is shown that, in order to ensure that the sample functions of the process are elements of some Hilbert space of harmonic functions, it is necessary and sufficient that the covariance function r has the expression

$$r(s, t) = \sum_{n=1}^{\infty} \lambda_n e_n(s) e_n(t) \tag{1}$$

with $\lambda_n \geqq 0$, $\sum_{n=1}^{\infty} \lambda_n < +\infty$ and $(e_n, n \in \mathbf{N})$ a complete orthonormal system in the actual Hilbert space.

If (1) is satisfied, the process defines a gaussian measure on this Hilbert space. From the gaussian measure the covariance properties of different linear functionals are derived in a very convenient way.

It is shown that the covariance function depends only on the spherical distance between s and t if, and only if, sets of potentials that can be carried into each other by means of a rotation of space around the origin have equal probability.

Furthermore, the process is shown to be non-ergodic, i.e.

$$r(\psi) = \int_{\substack{\psi_{s_0, t_0} = \psi \\ \omega \in \Omega}} \xi(s_0, \omega) \xi(t_0, \omega) dP(\omega) \neq \int_{\substack{\psi_{s, t} = \psi \\ s \in K_R}} \xi(s, \omega) \xi(t, \omega) ds dt \tag{2}$$

and that the variance of the random variable on the right side of (2) is fairly large.

It is therefore needless to try to determine the covariance function very exactly, as this is impossible.

As an alternative, it is suggested to use a simple expression as e.g.

$$r(s,t) = \sum_{n=3}^{\infty} \frac{A}{(n-1)(n-2)} \left(\frac{R^2}{r_s r_t} \right)^{n+1} P_n(\cos\theta) \tag{3}$$

which is easy to handle and which fits the empirical covariance function well indeed.

Predictions of a few linear functionals from others are carried through.

To derive the above results it has proved extremely useful to deal with the following mathematical concepts:

1) gaussian measures on Hilbert spaces,
2) reproducing kernel Hilbert spaces,
3) unitary representations of groups,

and of course standard results from functional analysis and probability theory.

I have tried to formulate the theory in a general language, mainly because it has been easier for myself not to be disturbed by irrelevant facts as "the functions are twice differentiable" and similar things.

I am now aware that it would have been more adequate to work only in complex Hilbert spaces and derive the results for the real Hilbert spaces from those. I never made up my mind during the work, and as a result you will find the word "unitary" used in situations, where "orthogonal" would be more appropriate.

The sections 2 and 4 contain only well known results and are included to introduce various notations.

Section 3 is mostly an "explanation" of the importance of a probabilistic study of the problem.

Sections 5, 6.1 and 7 contain a summary of the most important results on the subjects treated, which may be unknown to the reader.

Sections 6.2 and 8 contain results that I have not seen in the literature before, and the concept of a homogeneous random functional is probably new, although it is nothing but an obvious generalization of stationarity.

In section 9, the theoretical results are applied to build the model for the potential.

Section 10 is devoted to the problem of estimating the covariance function, i.e. it is here shown that the estimator in no sense converges towards the true value. Instead a closed expression of a covariance function which seems to fit data quite well is found.

Section 11 contains a non-serious attempt to predict different linear functionals from others, just to show that the method is not quite insane.

2. Some Fundamental Concepts of Potential Theory

This section is written for the benefit of readers, who know nothing or almost nothing about the potential of the earth, and may be omitted by others. A heuristic exposition of the theory is contained in *Heiskanen & Moritz* (1967), and *Sternberg & Smith* (1946) gives a more detailed treatment of the mathematics behind.

2.1. *The Potential of the Earth and Harmonic Functions*

The potential of the earth can be considered as a real-valued function on the space outside the earth. It can be partitioned into *the potential of rotation* and *the potential of mass*, the first of which is due to the earth's rotation about its own axis, and the second to the shape of the earth and the distribution of mass in its interior. The potential of rotation is very well known, and we shall here simply denote the potential of mass as *the potential*, if the opposite is not especially remarked.
This potential u is a *harmonic function* outside the earth, which means that it satisfies *Laplace's equation:*

$$\Delta u = 0 \tag{1}$$

where $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ is the *Laplacian operator*. Furthermore, the potential is *regular at infinity*, which means that $\lim_{r \to \infty} ru$ exists and $r^2 D_1 u$ is bounded; here $r = (x^2 + y^2 + z^2)^{1/2}$ in usual rectangular coordinates, and D_1 is any of the three partial derivatives of first order. The quantity $m = \lim_{r \to \infty} ru$ is denoted as *the mass* of the potential and coincides with the earth's mass in the usual sense.

Example 1.1.
The function u: $\mathbf{R}^3 \setminus \{\mathbf{0}\} \to \mathbf{R}$ defined as

$$u(x, y, z) = \frac{1}{r} = (x^2 + y^2 + z^2)^{-1/2}$$

is a regular potential with mass 1; in fact:

$$r^2 \frac{\partial u}{\partial x} = -x(x^2 + y^2 + z^2)^{-1/2},$$

which is certainly bounded by -1 and 1;

$$\lim_{r \to \infty} ru = 1;$$

$$\frac{\partial^2 u}{\partial x^2} = \frac{2x^2 - y^2 - z^2}{(x^2+y^2+z^2)^{5/2}},$$

and by symmetry you will easily see that u satisfies Laplace's equation. This potential is the potential of the mass 1, situated at the origin of the coordinate system.

As the shape of the earth is in no sense ideal, neither flat, nor spherical or ellipsoidal, and as we know neither the exact shape, nor the distribution of mass in its interior, we are unable to determine the potential by mathematical methods, but will have to search for it among a large class of harmonic functions. We do not even know the domain in which the potential has to be defined.

We shall not consider these difficulties in detail, but a way to overcome them (Krarup, 1969) is to search for the potential among functions, regular and harmonic, outside a sphere which is contained in the earth's interior, the so-called *Bjerhammar sphere*. Of course, this does not solve the problem, but because of the very important theorem of *Runge* (see e.g. Krarup, 1969) we can in this way find an approximation to the potential, which in some sense is arbitrarily good.

(We can make the usual compactification of $\mathbf{R}^3$, adding the point ∞. If we then remove the earth from this compact space, we get a locally compact metric space M, so that the potentials are continuous functions on M. The set of functions harmonic outside the Bjerhammer sphere is then *dense* in the set of harmonic functions on M *in the topology of uniform convergence on compact subsets of* M).

We shall now examine some of the properties of the class of regular harmonic functions outside a sphere with radius R. Let now in the following

$$M_R = \{(x,y,z) \in \mathbf{R}^3 | (x^2+y^2+z^2)^{1/2} > R\}. \tag{2}$$

It can be shown (Sternberg & Smith, 1946) that every regular harmonic function on an open subset M of three-dimensional euclidean space $\mathbf{R}^3$

is *analytic in M*, i.e. can be represented by an *absolute convergent power series:*

$$u(x, y, z) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} a_{ijk} (x - x_0)^i (y - y_0)^j (z - z_0)^k \qquad (3)$$

where $(x_0, y_0, z_0) \in M$ and

$$a_{ijk} = \frac{1}{i!\, j!\, k!} \frac{\partial^{i+j+k} u}{\partial x^i \partial y^j \partial z^k} \Bigg|_{(x, y, z) = (x_0, y_0, z_0)} . \qquad (4)$$

An important class of harmonic functions consists of the spherical harmonics. By a *spherical harmonic of degree n* we mean a homogeneous polynomial of degree n in x, y and z, which satisfies Laplace's equation:

$$K_n(x, y, z) = \sum_{p+q+r=n} a_{pqr} x^p y^q z^r \qquad (5)$$

$$\Delta K_n = 0. \qquad (6)$$

It is not difficult to show (Sternberg & Smith, 1946) that there exist exactly $2n+1$ linearly independent spherical harmonics of degree n. This means that there are functions K_{nm}, $m = -n, \ldots, -1, 0, 1, \ldots, n$, so that any spherical harmonic of degree n has a unique representation as

$$K_n = \sum_{m=-n}^{n} a_{nm} K_{nm}, \qquad (7)$$

where a_{nm} are real numbers.

We shall now introduce spherical coordinates (r, θ, λ), determined by

$$\left.\begin{aligned} x &= r \sin\theta \cos\lambda \\ y &= r \sin\theta \sin\lambda \\ z &= r \cos\theta \end{aligned}\right\} \quad , \qquad (8)$$

or equivalently

$$\left.\begin{aligned} r &= (x^2 + y^2 + z^2)^{1/2} \\ \theta &= \operatorname{Arctan} \frac{(x^2 + y^2)^{1/2}}{z} \\ \lambda &= \operatorname{Arctan} \frac{y}{x} \end{aligned}\right\} \qquad (9)$$

(see fig. 1).

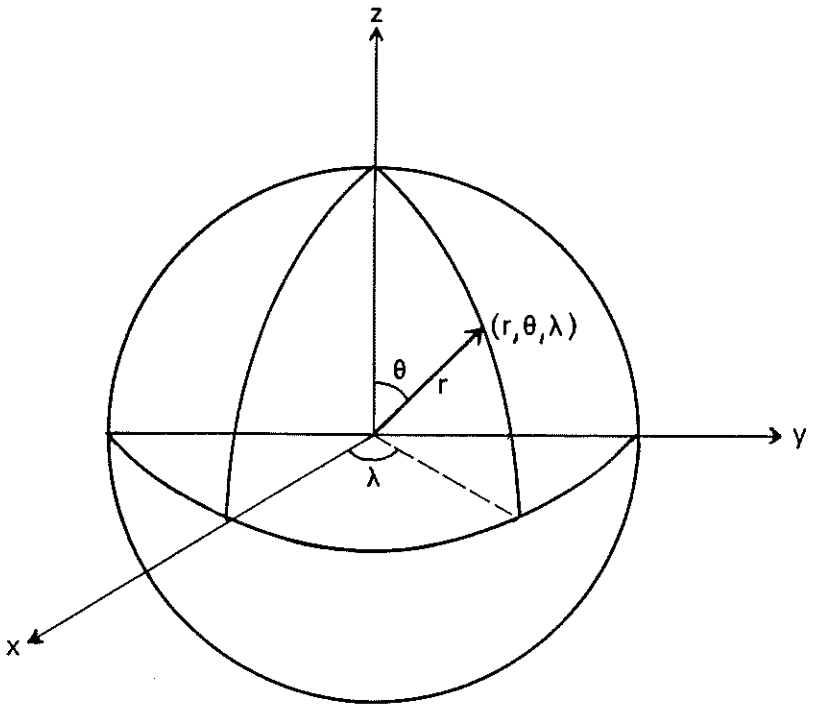

Fig. 1.
θ is *polar distance* – λ is *geocentric longitude* – r is *distance to origin*

It can now be shown that the functions

$$\left.\begin{aligned} K_{nm}(r,\theta,\lambda) = r^n R_{nm}(\theta,\lambda) = r^n P_{nm}(\cos\theta)\cos m\lambda \\ m = 0, 1, \ldots, n \end{aligned}\right\} \quad (10)$$

and

$$\left.\begin{aligned} L_{nm}(r,\theta,\lambda) = r^n S_{nm}(\theta,\lambda) = r^n P_{nm}(\cos\theta)\sin m\lambda \\ m = 1, \ldots, n \end{aligned}\right\} \quad (11)$$

together constitute $2n+1$ linearly independent spherical harmonics of degree n. P_{nm} are the *associated Legendre functions*, given by

$$P_{nm}(t) = (1-t^2)^{\frac{m}{2}} \frac{d^m P_n(t)}{dt^m}, \quad (12)$$

where P_n are the well known *Legendre polynomials*. The functions R_{nm} and S_{nm} in (10) and (11) are called *surface spherical harmonics*.

It can be shown that a function is harmonic inside the unit sphere if, and only if, it can be represented by a series in spherical harmonics:

$$u(r,\theta,\lambda) = \sum_{n=0}^{\infty} r^n \Big(\sum_{m=0}^{n} a_{nm} R_{nm}(\theta,\lambda) + \sum_{m=1}^{n} b_{nm} S_{nm}(\theta,\lambda) \Big). \quad (13)$$

The series converges absolutely and uniformly on any closed subset contained in the interior of the unit sphere.

Similarly, any function is harmonic and regular in the region M_R, if, and only if, it can be represented by a series

$$u(r,\theta,\lambda) = \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} \left(\sum_{m=0}^{n} a_{nm} R_{nm}(\theta,\lambda) + \sum_{m=1}^{n} b_{nm} S_{nm}(\theta,\lambda) \right), \tag{14}$$

which converges absolutely and uniformly on any set of the form

$$\{(r,\theta,\lambda) \mid r \geqq r_0\} \quad \text{with} \quad r_0 > R.$$

The functions:

$$\left.\begin{aligned} \bar{R}_{n0}(\theta,\lambda) &= \sqrt{2n+1}\, P_{n0}(\cos\theta) = \sqrt{2n+1}\, P_n(\cos\theta) = \bar{R}_{n0}(\theta) \\ \left.\begin{matrix} \bar{R}_{nm}(\theta,\lambda) \\ \\ \bar{S}_{nm}(\theta,\lambda) \end{matrix}\right\} &= \sqrt{2(2n+1)\frac{(n-m)!}{(n+m)!}} \left\{\begin{matrix} R_{nm}(\theta,\lambda) \\ \\ S_{nm}(\theta,\lambda) \end{matrix}\right. , \; m = 1, \ldots, n \end{aligned}\right\} \tag{15}$$

are often called *fully normalized surface spherical harmonics.*

The following addition formula is very useful:

$$\left.\begin{aligned} P_n(\cos\psi) &= P_n(\cos\theta) P_n(\cos\theta') \\ &+ 2 \sum_{m=1}^{n} \frac{(n-m)!}{(n+m)!} P_{nm}(\cos\theta) P_{nm}(\cos\theta') \cos m(\lambda - \lambda'), \end{aligned}\right\} \tag{16}$$

where $\cos\psi = \cos\theta\cos\theta' + \sin\theta\sin\theta'\cos(\lambda - \lambda')$ is the *spherical distance* between the points $(1,\theta,\lambda)$ and $(1,\theta',\lambda')$. If $\theta = \theta'$ and $\lambda = \lambda'$ we get

$$P_n(1) = [P_n(\cos\theta)]^2 + 2 \sum_{m=1}^{n} \frac{(n-m)!}{(n+m)!} [P_{nm}(\cos\theta)]^2 \tag{17}$$

and as $| P_n(t) | \leqq 1$ we get the useful inequality:

$$| P_{nm}(\cos\theta) |^2 \leqq \frac{(n+m)!}{(n-m)!} \tag{18}$$

which is valid for all n and m. It is possible to introduce various inner products in subclasses of harmonic functions in such a way that they become Hilbert spaces. We shall in the next part of this section consider an example.

2.2. *A Hilbert Space of Potentials*

From now on we shall denote a point in $\mathbf{R}^3$ with a small letter, e.g. t and its spherical coordinates with (r_t,θ_t,λ_t).

As proved for example by *Weyl* (1931) the following relations hold:

$$\left.\begin{aligned} \frac{1}{4\pi}\int_\sigma \bar{R}_{nm}(\theta,\lambda)\bar{R}_{kl}(\theta,\lambda) &= \delta_{kn}\delta_{ml} \\ \frac{1}{4\pi}\int_\sigma \bar{R}_{nm}(\theta,\lambda)\bar{S}_{kl}(\theta,\lambda) &= 0 \\ \frac{1}{4\pi}\int_\sigma \bar{S}_{nm}(\theta,\lambda)\bar{S}_{kl}(\theta,\lambda) &= \delta_{kn}\delta_{ml} \end{aligned}\right\} \quad (1)$$

where $\int_\sigma$ denotes the integral over the unit sphere in $\mathbf{R}^3$, and

$$\delta_{jk} = \begin{cases} 1 \text{ if } j = k \\ 0 \text{ if } j \neq k \end{cases}$$

is the Kronecker symbol. It is because of these relations that $\bar{R}_{nm}$ and $\bar{S}_{nm}$ are denoted *fully normalized.* They constitute a complete orthonormal system in the Hilbert space of all square-integrable real functions on the unit sphere.

If we define the functions e_n^m for $n \in \mathbf{N}$, $m = -n,\ldots,-1,0,1,\ldots,n$

$$e_n^m(t) = \begin{cases} \left(\frac{R}{r_t}\right)^{n+1} \bar{R}_{nm}(\theta_t,\lambda_t) \quad \text{for} \quad m \geqq 0 \\ \left(\frac{R}{r_t}\right)^{n+1} \bar{S}_{n|m|}(\theta_t,\lambda_t) \quad \text{for} \quad m < 0 \end{cases} \quad (2)$$

we get

$$\frac{1}{4\pi R^2}\int_{\sigma_R} e_n^m e_k^l = \delta_{kn}\delta_{lm}, \quad (3)$$

where $\int_{\sigma_R}$ denotes integration over the sphere with radius R and center at origin.

If we now by U_0 denote all finite real linear combinations of the functions e_n^m

$$U_0 = \left\{ \sum_{n=1}^{N} \sum_{m=-n}^{n} a_{nm} e_n^m \mid N \in \mathbf{N}, a_{nm} \in \mathbf{R} \right\}, \tag{4}$$

the functions in U_o will all be regular and harmonic in M_R. We can define an inner product on U_0 $\langle , \rangle$, by:

$$\langle u, v \rangle = \frac{1}{4\pi R^2} \int_{\sigma_R} uv. \tag{5}$$

It is easily verified by means of (3) that, if

$$\left.\begin{aligned} u &= \sum_{n=1}^{N} \sum_{m=-n}^{n} a_{nm} e_n^m \\ v &= \sum_{n=1}^{N} \sum_{m=-n}^{n} b_{nm} e_n^m \end{aligned}\right\} \tag{6}$$

then

$$\langle u, v \rangle = \sum_{n=1}^{N} \sum_{m=-n}^{n} a_{nm} b_{nm}. \tag{7}$$

It is obvious that, if we define the norm $|| \; ||$, by

$$||u|| = \langle u, u \rangle^{1/2}, \tag{8}$$

U_0 becomes a pre-Hilbert space.

If we by U denote the completion of U_0 with respect to the norm (8), U becomes a Hilbert space. This Hilbert space can be shown to consist of all functions, regular and harmonic in M_R, for which the limit

$$\lim_{\varepsilon \to 0} \frac{1}{4\pi R^2} \int_{\sigma_{R+\varepsilon}} u^2 \tag{9}$$

exists and is finite. The inner product $\langle , \rangle$ on U becomes

$$\langle u, v \rangle_U = \lim_{\varepsilon \to 0} \int_{\sigma_{R+\varepsilon}} uv. \tag{10}$$

Obviously U is a real, separable Hilbert space and $(e_n^m \mid m, n \in \mathbf{N}, \mid m \mid \leqq n)$ is a complete orthonormal system in U. We shall later show that U has a reproducing kernel, but because this is a general property of Hilbert spaces consisting entirely of harmonic functions, we shall not discuss this property until section 4. In the same section we shall also consider other Hilbert spaces of potentials relevant to our problem.

2.3. *The Disturbing Potential*

The potential W is often divided into a *normal potential* N and a *disturbing potential* T:

$$W = N + T. \tag{1}$$

The normal potential N is in a way a first approximation to the potential, and the problem of determining the potential is then formulated as determining the disturbing potential T.

In choosing the coordinate system in such a way that the origin is situated at the gravity center of the earth, the z-axis coincides with the first principal axis of inertia, and so that we get a righthanded coordinate system, it is possible to deduce that *the coefficients to* e_n^m *for* $n = 1$, $n = 2$ and $m = 1$, -1 *in a series expansion of the potential must be zero* (Heiskanen & Moritz, 1967). The constant term in the series expansion of W is *the mass of the earth* and is so well known that we can include it entirely into the normal potential, which means that *it will not occur in the expansion of the disturbing potential.* We shall therefore formulate the problem of determining the disturbing potential as finding a harmonic function in the corresponding closed *subspace* of a Hilbert space of potentials. But as a closed subspace of a Hilbert space is again a Hilbert space, this will be of no importance for the theoretical considerations. Throughout the years many versions of the normal potential have been used, but a detailed description of this problem cannot be given here. The reader is referred to different books on physical geodesy, e.g. *Heiskanen & Moritz* (1967).

2.4. *Linear Functionals*

It is physically impossible to measure the value of the potential in any point of space. But it is possible to get some information about the potential by measuring different linear functionals of the disturbing potential as gravity anomalies, deflections of the vertical and others. We shall later see that these functionals are *bounded* i.e. *continuous* in all the relevant Hilbert space topologies.

Example 4.1. Gravity anomalies

The gravity anomaly in a point $t_0 \in M_R$, $\Delta g(t_0)$, corresponding to the disturbing potential T, can be found as

$$\Delta g(t_0) = -\left.\frac{\partial T(t_0)}{\partial r}\right|_{t=t_0} - \frac{2}{r_{t_0}} T(t_0) \tag{1}$$

(Heiskanen & Moritz, 1967). If T has the series expansion

$$T = \sum_{n=1}^{\infty} \sum_{m=-n}^{n} a_{nm} e_n^m(t_0), \tag{2}$$

(1) can easily be transformed to

$$\Delta g(t_0) = \frac{1}{r_{t_0}} \sum_{n=1}^{\infty} \sum_{m=-n}^{n} (n-1) a_{nm} e_n^m(t_0). \tag{3}$$

If we define

$$b_{nm}(t_0) = \frac{n-1}{r_{t_0}} e_n^m(t_0), \tag{4}$$

and use the inequality (18) in part 1 of this section, we get

$$b_{nm}(t_0)^2 \leqq \frac{(n-1)^2}{r_{t_0}^2} \left(\frac{R}{r_{t_0}}\right)^{2n+2} 2(2n+1) \tag{5}$$

and the sum

$$\sum_{n=1}^{\infty} \sum_{m=-n}^{n} b_{nm}(t_0)^2 \tag{6}$$

will then converge because $\frac{R}{r_{t_0}} < 1$.

If we then by g_{t_0} denote the element of the Hilbert space U defined in the previous part of this section with coefficients $b_{nm}(t_0)$ to e_n^m, it is obvious from (3) that

$$\Delta g(t_0) = \langle T, g_{t_0} \rangle. \tag{7}$$

From the Frechét-Riesz theorem on representation of bounded linear functionals on Hilbert spaces we then have that all the functions mapping a disturbing potential into the corresponding gravity anomaly in a point $t_0 \in M_R$ are bounded linear functionals on U, i.e. elements of *the dual space* to U, U^*.

3. Remarks Concerning Stochastic Models for the Variation of the Potential and Gravity Field of the Earth

3.1. *Historical Remarks*

The severe difficulties of determining the potential of the earth and the irregular behaviour of the gravity anomalies made *Kaula* (1959), in stead of determining the value of the gravity anomalies by interpolation between measured values, try to describe the gravity anomalies as a *second-order homogeneous random field on a sphere*, i.e. as a family of random variables $\{\Delta g(t), r_t = r\}$ with finite second-order moments, constant mean value function equal to zero, and a covariance function invariant with respect to the group of rotations of three-dimensional space around the origin:

$$E\{\Delta g(t)\Delta g(s)\} = \sum_{n=0}^{\infty} \sigma_n^2 P_n(\cos\psi(s,t)), \tag{1}$$

where $\psi(s,t)$ is the spherical distance between s and t. The coefficients σ_n^2 to the Legendre polynomials in (1) are known as "degree variances". (It has been shown by *Obukhov* (1947) that any rotation invariant positive definite continuous function on a sphere must be of the form (1)).

Kaula tried to estimate the first 30 values of $\sigma_n{}^2$ by means of a harmonic analysis of an empirical covariance function. Usual linear least squares prediction was then applied to predict the gravity anomalies in areas without measurements. The 1959 paper by Kaula was nothing but an "ad hoc"-adaption of time series analysis techniques, but contains many relevant ideas and valuable results.

Among other authors, *H. Moritz* (1963) developed a systematic theory for the statistical treatment of gravity measurements especially with respect to prediction of gravity anomalies. Moritz pointed out the very close analogy between interpolation and prediction. However, the approach of Moritz was not theoretically exact as regards the probabilistic background of the model, which we shall discuss later.

In later articles (e.g. Kaula 1963, 1966) several attempts were made to improve the values of the degree variances, and some of the basic theoretical concepts were discussed.

Although the physical and mathematical relationship between potential, gravity anomalies, deflections of the vertical and other functionals has been known among geodesists for years, *Krarup* (1969) was the first to exploit this in connection with the statistical analysis of these measurements in such a way that it became possible to treat all these functionals simultaneously.

In "A Contribution" Krarup points out that, in order to make sure that the predicted potential is a potential, it is not enough to model the process on a sphere, but you must define the process on the space outside a Bjerhammar sphere. Krarup considers *the values of the potential as the fundamental process and the gravity anomaly and other processes as results of some linear operation on the potential process.* Krarup proves that, if the predicted potential should be a potential, any rotation invariant covariance function must be of the form

$$c(s,t) = \sum_{n=0}^{\infty} A_n(2n+1)\left(\frac{R^2}{r_s r_t}\right)^{n+1} P_n(\cos\psi(s,t)), \tag{2}$$

with $A_n \geqq 0$.

Krarup draws the attention to the fact that the probabilistic background has not been treated in a satisfactory way. I think that Krarup is the first to mention the word probability measure! Among other things, Krarup points out that in the definition of the covariance function as

$$c(\psi) = \int_{\psi(s,t)=\psi} \int_{\sigma_R} \Delta g(s)\Delta g(t)\,dt\,ds \tag{3}$$

there is an implicit assumption of the process to be *ergodic*, as integration with respect to a probability measure has been changed by an integration with respect to the Haar measure on the rotation group, which by no means is a trivial assumption. In fact we shall prove that it is wrong!

Tscherning (1970) has made an attempt to apply the methods outlined by Krarup to the prediction of deflections of the vertical in Denmark from gravity data. Tscherning tries to specify some of the probabilistic assumptions and uses the interesting work of *Parzen* (1959) to describe the theoretical statistical ideas behind the analysis.

Finally, we shall mention the work of *Gaposchkin and Lambeck* (1970), where some improvements of the degree variances were made using satellite data.

3.2. *Motivating Remarks*

It is the main purpose of this work to build a stochastic model, which is theoretically exact, i.e. in which all the assumptions are explicitly stated and examined in a way that the model should be *without contradictions and approximations* where this is not necessary. I shall also try from the theoretical analysis to draw some consequences about the statistical analysis of the data, i.e. estimation and prediction.

It is my opinion that, in order to get a good model, we must try to use all information about the mathematical and physical properties of the potential.

It must then be very important to find a necessary and sufficient condition of the covariance function, which ensures that the sample functions are almost certain to be elements of some Hilbert space of potentials.

As it could be interesting to try to search for the potential outside another figure than a Bjerhammar sphere, e.g. an ellipsoid, I shall try to build up a general model which applies to any Hilbert space of harmonic functions.

We shall also describe the relation between measure-preserving transformations and rotation invariance because this gives an intuitively better understanding of the invariance assumptions of the model.

In this connection we shall investigate the ergodic properties of a similar process in order to deal with estimation problems.

First of all, however, we have to use an awful lot of mathematics. Maybe I will sometimes generalize more than necessary to solve the most relevant problems, but I think that it often gives a better understanding of the model.

4. Reproducing Kernel Hilbert Spaces

The theory of reproducing kernels has for two reasons turned out to be very useful in the treatment of this problem:

1) any Hilbert space of sufficiently well behaved functions (e.g. potentials) has a reproducing kernel,
2) the Hilbert space, spanned by a second order stochastic process, is isomorphic with some reproducing kernel Hilbert space.

The second property has been utilized for the first time by *Parzen* (1959, 1961). *Aronszajn* (1950) gives a detailed treatment of the theory and *Meschkowski* (1962) considers many applications. Also *Davis* (1963) has a short treatment of the basic results with special attention to approximation problems.

4.1. *Basic Properties and Definitions*

Let U be a Hilbert space consisting of functions u from M into the real or complex numbers: $u: M \to \mathbf{K}$ ($= \mathbf{R}$ or $\mathbf{C}$). By $\langle\ ,\ \rangle$, $||\ ||$ we denote respectively the inner product and norm in U. Let k be a function of two variables $k: M \times M \to \mathbf{K}$. For all $t \in M$ we shall by k^t denote the function $k^t: M \to \mathbf{K}$, defined as $k^t(s) = k(s,t)$ for $s \in M$.

Definition 1.1. The function k is said to be a *reproducing kernel* for the space U if

i) $\forall\ s \in M: k^s \in U$
ii) $\forall\ u \in U\ \forall s \in M: u(s) = \langle u, k^s \rangle$.

The property ii) is usually called the reproducing property.

From the definition is easily seen that k is Hermitian

$$k(s,t) = \langle k^t, k^s \rangle = \langle \overline{k^s, k^t} \rangle = \overline{k(t,s)} \tag{1}$$

Furthermore, $k(s,s) \geqq 0$ and for all $a_1, \ldots, a_n \in \mathbf{K}$, $t_1, \ldots, t_n \in M$

$$0 \leqq \langle \sum_{i=1}^{n} a_i k^{t_i}, \sum_{j=1}^{n} a_j k^{t_j} \rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i \bar{a}_j k(t_j, t_i). \tag{2}$$

The property (2) shall we express by saying that k is *positive*.

The following important theorem by Aronszajn gives a n. and s. condition of a Hilbert space having a reproducing kernel:

Theorem 1.2 (Aronszajn). *A necessary and sufficient condition that U have a reproducing kernel is that all the linear functionals l_s, $s \in M$, defined by*

$$l_s(u) = u(s), \ \forall u \in U \tag{3}$$

are bounded, i.e. *elements of the dual space U^*.*

Proof: Follows immediately from the definition and the Frechét-Riesz representation theorem.

Theorem 1.3. *If U has a reproducing kernel, the kernel is unique.*

Proof: Suppose k_1 and k_2 are reproducing kernels. From the definition we get $\forall s,t \in M$:

$$k_1(s,t) = k_1^t(s) = \langle k_1^t, k_2^s \rangle = \overline{k_2^s(t)} = k_2(s,t)$$

and therefore $k_1 = k_2$.

A property that makes the theory so important is that, if a Hilbert space has a reproducing kernel, the convergence in norm implies pointwise convergence. This means that the representations of the elements as sums are convergent in the usual sense, and not just in the strong topology.

Theorem 1.4. *Let U have a reproducing kernel, and let $u, u_1, u_2, \ldots$* be *elements of U, so that* $\lim_{n \to \infty} || u - u_n || = 0$. *Then*

$$\forall s \in M: \lim_{n \to \infty} u_n(s) = u(s), \tag{4}$$

where the convergence holds uniformly in every subset $M' \subseteq M$, so that k is bounded on the diagonal of $M' \times M'$.

Proof.
$$\begin{aligned} |u(s) - u_n(s)|^2 &= |\langle u - u_n, k^s \rangle|^2 \\ &\leqq || u - u_n || \cdot || k^s || = || u - u_n || \cdot k(s,s)^{1/2} \\ &\leqq C \cdot || u - u_n || \end{aligned}$$

where $| k(s,s)^{1/2} | < C$ for all $s \in M'$.

It could be interesting to see, if any positive hermitian function would be reproducing kernel for some Hilbert space. In fact, the answer is affirmative:

Theorem 1.5. *Let M be an arbitrary set, and let $k: M \times M \to \mathbf{K}(= \mathbf{R}$ or $\mathbf{C})$ be positive and Hermitian. Then there exists a Hilbert space, unique up to isomorphism, which has k as reproducing kernel.*

Proof. A rigorous proof is given by *Meschkowski*, 1962, p. 96, and I shall just sketch how to construct the space. Define U_0 as

$$U_0 = \left\{ \sum_{i=1}^{N} a_i k^{t_i} \,\middle|\, a_i \in \mathbf{K},\ t_i \in M,\ N \in \mathbf{N} \right\} \tag{5}$$

and the inner product on U_0 as

$$\left\langle \sum_{i=1}^{N} a_i k^{t_i}, \sum_{j=1}^{N} b_j k^{s_j} \right\rangle = \sum_{i=1}^{N} \sum_{j=1}^{N} a_i \bar{b}_j k(s_j, t_i) \tag{6}$$

and U_0 becomes a pre- Hilbert space, if the norm is defined as $\| \ \| = \langle \ , \ \rangle^{1/2}$. The space U generated by completion of U_0 will then be a Hilbert space with k as reproducing kernel.

4.2. *Some Special Results*

Separability: Let U be a reproducing kernel Hilbert space of functions on M.

Theorem 2.1. *U is separable if, and only if, there exists an orthonormal system e_n, $n \in \mathbf{N}$ in U, so that*

$$\forall s,t \in M: k(s,t) = \sum_{n=1}^{\infty} e_n(s)\overline{e_n(t)}.$$

The functions e_n will then constitute a complete orthonormal system in U.

Proof. Suppose U is separable. Then there exists a complete orthonormal system e_n, $n \in \mathbf{N}$. As $k_t \in U$ it can be written as

$$k^t(s) = \sum_{n \in \mathbf{N}} \langle k^t, e_n \rangle e_n(s) \tag{1}$$

But $\langle k^t, e_n \rangle = \overline{\langle e_n, k^t \rangle} = \overline{e_n(t)}$, and you have the result. Suppose e_n, $n \in \mathbf{N}$ is an orthonormal system, so that $k(s,t) = \sum_{n \in \mathbf{N}} e_n(s)\overline{e_n(t)}$. We shall

show that e_n, $n \in \mathbf{N}$ is complete. Assume that $u \in U$ is orthogonal to all the functions e_n, i.e. $\langle u, e_n \rangle = 0 \; \forall \, n \in \mathbf{N}$. Then

$$u(t) = \langle u, k^t \rangle = \sum_{n \in \mathbf{N}} \langle u, \overline{e_n(t)} e_n \rangle$$

$$= \sum_{n \in \mathbf{N}} \langle u, e_n \rangle \cdot e_n(t) = 0 \quad \forall t \in T.$$

As u was an arbitrary element of U, U must be separable.

The above theorem implies that, if $e_n, n \in \mathbf{N}$ and $f_n, n \in \mathbf{N}$ are complete orthonormal systems, then

$$\sum_{n \in \mathbf{N}} e_n(s)\overline{e_n(t)} = \sum_{n \in \mathbf{N}} f_n(s)\overline{f_n(t)} \quad \forall s, t. \tag{2}$$

Projections and subspaces:

Theorem 2.2. *If $U' \subseteq U$ is a closed subspace of U, then U' also has a reproducing kernel.*

Proof. The linear functionals $u \to u(s)$ must also be continuous, when they are restricted to U'.

Theorem 2.3. *If k' is the kernel of $U' \subseteq U$, and u is an element of U, then the projection of u on U' is found as*

$$Pu(t) = \langle u, k'^t \rangle. \tag{3}$$

Proof. U has a unique decomposition $u = u' + u''$, where $u' \in U'$ and $u'' \in U'^{\perp}$, and $u' = Pu$ is the projection of u on U'. Because $k'^t \in U'$, and $u'' \in U'^{\perp}$,

$$\langle u'', k'^t \rangle = 0 \, \forall t.$$

But then $Pu(t) = \langle u', k'^t \rangle = \langle u' + u'', k'^t \rangle = \langle u, k'^t \rangle$ q.e.d.

Linear functionals:

Theorem 2.4. *Let U have a reproducing kernel and let $u^* \in U^*$, be a bounded linear functional on U. Then the function*

$$h_{u^*}(t) = \overline{u^*(k^t)} \tag{4}$$

is an element of U and for any $v \in U$

$$u^*(v) = \langle v, h_{u^*} \rangle \tag{5}$$

Proof. According to the Frechét-Riesz theorem, there is a unique element $g_{u^*} \in U$, so that

$$u^*(v) = \langle v, g_{u^*} \rangle \; \forall \, v \in U$$

then $\overline{u^*(k^t)} = \overline{\langle k^t, g_{u^*} \rangle} = \langle g_{u^*}, k^t \rangle = g_{u^*}(t)$. But then $h_{u^*} = g_{u^*} \in U$ and the result follows.

4.3. *Examples of Reproducing Kernel Hilbert Spaces*

Example 3.1.

Let $M = \mathbf{N}$, and let

$$U = \left\{ u : \mathbf{N} \to \mathbf{R} \mid \sum_{n \in \mathbf{N}} | u(n) |^2 < \infty \right\} = \mathbf{R}^{\infty}$$

with the inner product

$$\langle u, v \rangle = \sum_{n \in \mathbf{N}} u(n) \, v(n). \tag{1}$$

Define

$$k(m,n) = \delta_{mn} = \sum_{i \in \mathbf{N}} e_i(m) \, e_i(n) \tag{2}$$

where $e_i(n) = \delta_{in}$ is a well-known orthonormal basis for U. It is not difficult to see that k is a reproducing kernel. In fact $k^m = e_m \in U$ and $u(m) = \langle u, e_m \rangle = \langle u, k^m \rangle$.

Example 3.2.

Let $M = [0,1]$ and let U consist of all functions f: $M \to \mathbf{R}$ in such a way that f is absolutely continuous with a square-integrable derivative Df, and $f(0) = 0$.

$$U = \left\{ f \mid \exists \, Df \in L_2([0,1]) : f(x) = \int_0^x Df(t) \, dt \; \forall \, x \right\}.$$

Define the inner product

$$\langle f, g \rangle = \int_0^1 Df(t) \cdot Dg(t) \, dt \tag{3}$$

and the function

$$k(s,t) = \min(s,t). \tag{4}$$

It is easy to see that

$$k^s = \left\{ \begin{array}{ll} t & \text{for } t \leqq s \\ s & \text{for } t > s \end{array} \right\} \tag{5}$$

is an element of U and

$$Dk^s = 1_{[0,s]} \tag{6}$$

We then have

$$\langle f, k^s \rangle = \int_0^1 Df(t)\, Dk^s(t)\, dt = \int_0^s Df(t)\, dt = f(s) \tag{7}$$

and thus $k(s,t)$ is the reproducing kernel for U.

Example 3.3.

Let $M = M_R$, the space outside a sphere with radius R in $\mathbf{R}^3$, define

$$\left. \begin{aligned} k(s,t) &= \sum_{n=0}^{\infty} \frac{2n+1}{n+1} \frac{R^{2n+1}}{(r_s\, r_t)^{n+1}} P_n(\cos\psi_{s,t}) \\ &= \frac{2}{L} + \frac{1}{R} \ln\left(\frac{r_s\, r_t (1 - \cos\psi_{s,t})}{R \cdot L + R^2 - r_s \cdot r_t \cos\psi_{s,t}} \right) \end{aligned} \right\} \tag{8}$$

where

$$L = \left(\left(\frac{r_s\, r_t}{R} \right)^2 - 2 r_s r_t \cos\psi_{s,t} + R^2 \right)^{1/2} \tag{9}$$

and $\psi_{s,t}$ is the spherical distance between s and t:

$$\cos\psi_{s,t} = \cos\theta_s \cos\theta_t + \sin\theta_s \sin\theta_t \cos(\lambda_s - \lambda_t) \tag{10}$$

$(r_s, \theta_s, \lambda_s)$ are the spherical coordinates of s. $k(s,t)$ is the famous Dirichlet kernel. The Hilbert space generated by this kernel can be shown to consist of all potentials u, so that

$$\| u \|^2 = \int_{M_R} \operatorname{grad} u \cdot \operatorname{grad} u < + \infty . \tag{11}$$

And the inner product is defined as

$$\langle u, v \rangle = \int_{M_R} \operatorname{grad} u \cdot \operatorname{grad} v. \tag{12}$$

Example 3.4.

The Hilbert space defined in section 2.2, consisting of harmonic functions, in such a way that

$$\lim_{\varepsilon \to 0} \int_{\sigma_{R+\varepsilon}} u^2 < +\infty. \tag{13}$$

The kernel is the Poisson kernel, see e.g. Krarup, 1969

$$k(s,t) = \sum_{n=0}^{\infty} (2n+1)\left(\frac{R^2}{r_s\, r_t}\right)^{n+1} P_n(\cos\psi_{s,t}) = \frac{r_s^2 r_t^2 - R^4}{R}\,\frac{1}{L^3}. \tag{14}$$

5. Some Fundamental Concepts of Probability Theory and Stochastic Processes

This section is primarily written for the reader, who has a basic knowledge of modern probability theory but who is unfamiliar with the more general theory of stochastic processes. It may, however, also be useful for others to read because the notation and terminology used in the rest of this work will be introduced here.

5.1. *Basic Definitions*

A probability space is a triplet $(\Omega, \mathscr{A}, P)$, where Ω is an arbitrary set, $\mathscr{A}$ is a boolean σ-algebra of subsets of Ω, and P is a probability measure, i.e. a σ-additive positive set function on $\mathscr{A}$, with $P(\Omega) = 1$. For a more detailed description see *Doob* (1953) or *Neveu* (1965).

A *random variable* is nothing but a measurable mapping of Ω into the set of real numbers, i.e. a mapping X, satisfying

$$X^{-1}(B) \in \mathscr{A}, \forall B \in \mathscr{B}, \tag{1}$$

where $\mathscr{B}$ is the σ-algebra of Borel subsets of $\mathbf{R}$, the real numbers. A random variable induces a probability measure on $(\mathbf{R}, \mathscr{B})$, which we shall denote by P^X, by

$$P^X(B) = P(X^{-1}(B)), \forall B \in \mathscr{B}. \tag{2}$$

The probability measure P^X is called the *distribution* of the random variable X.

Now let T be an arbitrary set, and let $(\Omega, \mathscr{A}, P)$ be a probability space.

Definition 1.1. A *stochastic process* is a mapping

$$\xi : T \times \Omega \to \mathbf{R} \tag{3}$$

so that for all $t \in T$ the mappings

$$\xi^t : \Omega \to \mathbf{R},\ \xi^t(\omega) = \xi(t,\omega),\ \forall\, \omega \in \Omega \tag{4}$$

are random variables.

The functions

$$\xi^\omega : T \to \mathbf{R},\ \xi^\omega(t) = \xi(t,\omega),\ \forall\, t \in T \tag{5}$$

for $\omega \in \Omega$ are usually called *the sample functions* of the process. Instead of *stochastic process* the word *random function* is often used. Sometimes we just write $\xi(t)$ instead of $\xi(t,\omega)$ when misunderstanding is impossible. It is often convenient to think of a stochastic process as an indexed set or family of random variables ξ^t, $t \in T$. Often T denotes time and the values ξ^t denote the random development of a system, e.g. the temperature in a room. The sample functions correspond to any particular realization of the process. The sample functions are elements of $\mathbf{R}^T$, the space of all mappings from T to $\mathbf{R}$.

The family of joint distributions of any finite set of random variables $\xi^{t_1}, \ldots, \xi^{t_n}$, $t_i \in T$ is called *the finite-dimensional distributions of the process*. A good deal of the probability structure of P is determined by the finite-dimensional distributions, but certainly not all of it. By $\mathscr{A}_\xi$ we denote the σ-algebra generated by sets of the form

$$\{\omega \mid (\xi(t_1,\omega), \ldots, \xi(t_n,\omega)) \in B_n\}, \tag{6}$$

where $t_1, \ldots, t_n \in T$, and B_n is any n-dimensional Borel set. The finite-dimensional distributions determine uniquely the probability of sets of the form (6) and therefore of all sets in $\mathscr{A}_\xi$. In order that all ξ^t should be random variables, $\mathscr{A}_\xi \subseteq \mathscr{A}$, but it is sometimes relevant to work with an $\mathscr{A}$ which is strictly bigger than $\mathscr{A}_\xi$, e.g. one could be interested in the probability of a set of the form

$$\{\omega \mid \xi(t,\omega) \leqq 1 \quad \forall\, t \in T^*\}, \tag{7}$$

where $T^* \subseteq T$. If T^* is not denumerable, the set (7) does not belong to $\mathscr{A}_\xi$, and the probability is not determined by the finite-dimensional distributions. I shall not treat these questions in detail, but refer to *Doob* (1953), Chapter II. The σ-algebra $\mathscr{A}_\xi$ is usually called the *σ-algebra spanned by ξ*.

Now, let $X \subseteq \mathbf{R}^T$ be a space of real-valued functions, so that all the sample functions of the process belong to X:

$$\forall\, \omega \in \Omega : \xi^\omega \in X. \tag{8}$$

Define $\mathscr{B}$ as the smallest σ-algebra of subsets of X that contains all *cylinder sets:*

$$\{x \in X \mid (x(t_1), \ldots, x(t_n)) \in B_n\} \tag{9}$$

where $t_1, \ldots, t_n \in T$ and B_n is any n-dimensional Borel set. Now $(X, \mathscr{B})$ is a measurable space, and the mappings

$$\tilde{\xi} : \Omega \to X, \ \tilde{\xi}(\omega) = \xi^\omega$$

are all measurable with respect to the σ-algebra $\mathscr{A}_\xi$. The mapping $\tilde{\xi}$ therefore induces a probability measure on $(X, \mathscr{B})$, which we shall denote by μ^ξ, by

$$\mu^\xi(B) = P\{\omega \mid \xi^\omega \in B\}, \ \forall B \in \mathscr{B}. \tag{10}$$

We shall call μ^ξ *the distribution of the process* ξ, or of *the random function* ξ. As $\tilde{\xi}$ is measurable with respect to $\mathscr{A}_\xi$, μ^ξ is determined from the finite-dimensional distributions.

5.2. *Construction of Random Functions*

It is often necessary to construct a random function from some family of finite-dimensional distributions, and one might ask, when it is possible to find an $(\Omega, \mathscr{A}, P)$, so that there exists a process on Ω with the given family of finite-dimensional distributions? The answer is contained in the fundamental theorem by *Kolmogorov* (see e.g. Doob, 1953, p. 609) saying that the following consistency conditions are necessary and sufficient. If $F_{t_1, \ldots, t_n}$ are the joint distribution function of $\xi^{t_1}, \ldots, \xi^{t_n}$ then

i) $F_{t_1, \ldots, t_n}(x_1, \ldots, x_n) = F_{t_{\pi(1)}, \ldots, t_{\pi(n)}}(x_{\pi(1)}, \ldots, x_{\pi(n)})$

for all permutations π of $1, \ldots, n$

ii) $\lim_{x_j \to \infty} F_{t_1, \ldots, t_n}(x_1, \ldots, x_n) =$
$F_{t_1, \ldots, t_{j-1}, t_{j+1}, \ldots, t_n}(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n)$.

If a family of finite-dimensional distributions satisfies i) and ii) we say that the family is *consistent*. As $(\Omega, \mathscr{A})$ can always be chosen $(\mathbf{R}^T, \mathscr{B})$, when $\mathscr{B}$ is the σ-algebra generated by the cylinder sets.

Now, if $X \subseteq \mathbf{R}^T$ is some subspace of real valued functions, when does a probability on $(X, \mathscr{B})$ exist, so that the process $\xi : T \times X \to \mathbf{R}$ defined as

$$\xi(t, x) = x(t) \quad \forall t \in T \tag{1}$$

has the given finite-dimensional distributions? This problem becomes relevant, when for example $T = \mathbf{R}$ and $X = C(T)$, the space of continuous functions on T, and you want to construct a process with continuous sample functions. Or if $T = M_R$ is the subset of $\mathbf{R}^3$ of points outside a sphere with radius R

$$M_R = \{(x,y,z) \in \mathbf{R}^3 \mid x^2+y^2+z^2 > R^2\},\ R \geqq 0 \tag{2}$$

and $X = \Pi_R$ or $X = U \subseteq \Pi_R$, where Π_R is the space of all functions regular and harmonic in M_R and U is some Hilbert space of potentials in Π_R. We want to ensure that all the sample functions are elements of X. We know that this is the case if, and only if, there exists a probability on $(X, \mathscr{B})$ so that the process defined by (1) has the given finite-dimensional distributions. Which conditions shall the finite-dimensional distributions satisfy? Usually it is very difficult to solve problems of this kind, and they are therefore often avoided and forgotten. In the book by *Cramér and Leadbetter* (1968), further examples of this kind are treated, and solutions are also given in some cases. Mostly only sufficient conditions are known. The following theorem is of no use at all, but maybe it gives some intuitive understanding of the problem.

Theorem 2.1. *Let P be a probability measure on $(\mathbf{R}^T, \mathscr{B})$ and let X be a subset of $\mathbf{R}^T$. If we by $\mathscr{B}_X$ denote the σ-algebra of sets*

$$\mathscr{B}_X = \{B \cap X \mid B \in \mathscr{B}\}, \tag{3}$$

(i.e. the cylinder sets of X), then the following two statements i) *and* ii) *are equivalent:*

i) *There exists a probability P_X* on *$(X, \mathscr{B}_X)$ so that $\forall B \in \mathscr{B}$: $P(B) = P_X(B \cap X)$*

ii) *X has outer P-measure* 1, i.e. $P^*(X) = \inf_{\{B \in \mathscr{B} \mid B \supseteq X\}} P(B) = 1$.

Proof: Suppose i) holds; then

$$P^*(X) = \inf_{\{B \in \mathscr{B} \mid B \supseteq X\}} P(B) = \inf_{\{B \in \mathscr{B} \mid B \supseteq X\}} P_X(B \cap X) = P_X(X) = 1,$$

which was to be proved. Now suppose ii) is true. If we define for any set $B \cap X \in \mathscr{B}_X$

$$P_X(B \cap X) = P(B). \tag{4}$$

This is easily seen to be a probability on $(X, \mathscr{B}_X)$ if, and only if, the above definition (4) has a meaning, i.e. is unique. Suppose $B_1 \cap X = B_2 \cap X$, where $B_1, B_2 \in \mathscr{B}$, then we have to show that $P(B_1) = P(B_2)$ But $B_1 \Delta B_2$ (Δ is the symmetric difference between B_1 and B_2, $B_1 \Delta B_2 = (B_1 \setminus B_2) \cup (B_2 \setminus B_1)$ must be contained in X^c, the complement to X:

$$B_1 \Delta B_2 \subseteq X^c \tag{5}$$

But then $(B_1 \Delta B_2)^c \supseteq X$ and $P(B_1 \Delta B_2)^c = 1 - P(B_1 \Delta B_2) \geqq P^*(X) = 1$ and therefore $P(B_1 \Delta B_2) = 0$. But as $B_1 \subseteq (B_1 \Delta B_2) \cup B_2$ and $B_2 \subseteq (B_1 \Delta B_2) \cup B_1$, $| P(B_1) - P(B_2) | \leqq P(B_1 \Delta B_2) = 0$, which means, that $P(B_1) = P(B_2)$ q.e.d.

The uselessness of the theorem is due to the fact that it is indeed very difficult to prove that ii) is fulfilled, when P is the probability on $(\mathbf{R}^T, \mathscr{B})$ defined by some consistent family of finite-dimensional distributions. However, I like the theorem!

5.3. *Second-order Random Functions*

Let $(\Omega, \mathscr{A}, P)$ be a probability space. The random variable $X: \Omega \to \mathbf{R}$ is said to have finite second-order moments if it is square-integrable, i.e. if

$$EX^2 = \int_\Omega X(\omega)^2 dP(\omega) < +\infty. \tag{1}$$

The space $L_2(\Omega, \mathscr{A}, P)$ of all square-integrable functions on Ω becomes a *Hilbert space*, defining the inner product $\langle , \rangle_2$ as

$$\langle X, Y \rangle_2 = EXY = \int_\Omega X(\omega) Y(\omega) dP(\omega). \tag{2}$$

Strictly speaking, it is a space of equivalence classes, where X and Y are said to be equivalent if $P\{\omega \mid X(\omega) = Y(\omega)\} = 1$. Stochastic processes, where $\xi^t \in L_2(\Omega, \mathscr{A}, P)$ for all $t \in T$, are called *second-order stochastic processes* or *second-order random functions*. These processes have always been of particular importance because of the correlation theory similar to the least squares theory known from classical mathematical statistics.

If ξ is a second-order random function, we define *the mean value function* m: $T \to \mathbf{R}$ as

$$m(t) = E\xi^t = \int_\Omega \xi(t, \omega) dP(\omega), \forall\, t \in T \tag{3}$$

and *the covariance function* $r : T \times T \to \mathbf{R}$ as

$$\left.\begin{aligned} r(s,t) &= E(\xi^s - m(s))\,(\xi^t - m(t)) \\ &= \operatorname{cov}(\xi^s, \xi^t), \quad s,t \in T. \end{aligned}\right\} \tag{4}$$

In the following we shall assume that $m(t) = 0\ \forall\, t \in T$, when the opposite is not explicitely remarked. This is no restriction to theoretical considerations as we can always consider the random variables $\xi^t - m(t)$ instead of ξ^t.

The covariance function is *symmetric* and *positive:*

$$\left.\begin{aligned} r(s,t) &= r(t,s)\ \forall\, s,t \in T \\ \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j r(t_i,t_j) &\geqq 0\ \forall\ t_1, \ldots, t_n \in T \\ &\forall\ a_1, \ldots, a_n \in \mathbf{R}. \end{aligned}\right\} \tag{5}$$

(5) follows from the fact, that

$$0 \leqq E\left|\sum_{i=1}^{n} a_i \xi^{t_i}\right|^2 = \sum_i \sum_j a_i a_j E\xi^{t_i}\xi^{t_j} = \sum_i \sum_j a_i a_j r(t_i,t_j). \tag{6}$$

But then theorem 1.5 from section 4 shows the existence of a Hilbert space, unique up to isomorphism, which has r as reproducing kernel. This Hilbert space can be chosen as the completion of the set

$$K_0(r) = \left\{ \sum_{i=1}^{n} a_i r^{t_i} \mid a_i \in \mathbf{R},\ t_i \in T \right\} \tag{7}$$

with respect to the inner product

$$\langle \textstyle\sum a_i r^{s_i}, \sum b_j r^{t_j} \rangle = \sum_i \sum_j a_i b_j r(s_i,t_j). \tag{8}$$

This Hilbert space we shall denote as K_ξ, *the kernel space of the process.*

There is another relevant Hilbert space representation of the process. By $L_2(\xi^t,\ t \in T)$ we mean the closed linear manifold, spanned by the random variables $\xi^t,\ t \in T$ considered as elements of $L_2(\Omega, \mathscr{A}, P)$. As this subspace of $L_2(\Omega, \mathscr{A}, P)$ is closed, it can be considered as a Hilbert space itself, which we shall denote as H_ξ, *the Hilbert space spanned by the process.* Now the interesting thing is that H_ξ and K_ξ are *isomorphic.* Define $J: H_\xi \to K_\xi$, as the extension by linearity and continuity of the isometric mapping $J^*: \{\xi^t,\ t \in T\} \to \{r^t, t \in T\}$ $J^*(\xi^t) = r^t$. It is no

problem to see that J^* can be extended to the one-to-one isometric operator J.

Now, this is the way in which the theory of reproducing kernel Hilbert spaces becomes relevant and useful to statistical inference on stochastic processes, especially prediction theory and related topics (see Parzen, 1961).

Now if a topology on T is defined, we say that ξ is *continuous in the mean* at $t_0 \in T$, if

$$E(\xi^t - \xi^{t_0})^2 \to 0 \quad \text{when } t \to t_0. \tag{9}$$

If this holds for all $t_0 \in T$ we just say that ξ is *continuous in the mean.* But as

$$E(\xi^t - \xi^{t_0})^2 = r(t,t) - 2r(t,t_0) + r(t_0,t_0), \tag{10}$$

continuity in the mean is equivalent to continuity of the covariance function at any diagonal point of $T \times T$. Now suppose that T is locally compact, and r is everywhere continuous. We already know that r^t is continuous everywhere on T. But as K_ξ is a reproducing kernel Hilbert space, any element of K_ξ is a pointwise convergent sum $\sum_{i=1}^{\infty} a_i r^{t_i}(t)$. Now let $t_0 \in T$, then there exists a compact neighbourhood U of t_0, and as r is continuous at the diagonal point, r is *bounded* on $U \times U$. But then the sum converges *uniformly*, and the limit must be continuous at t_0 (see section 4, Theorem 1.4). Hence, all the elements of K_ξ are continuous functions.

Thus we have seen that continuity in the mean is in natural way related to continuity of the functions in K_ξ. In the same way differentiability in the mean is closely related to differentiability of the functions in K_ξ. This is why one often avoids the difficulties in dealing with sample function properties and instead considers the corresponding properties of the function in K_ξ, as for example described by *Parzen* (1961). Another relevant reason is that *the predictions* all will be elements of the kernel space K_ξ. This, I think, is also the reason of letting the covariance function be a reproducing kernel for some Hilbert space of potentials, as suggested by *Krarup* (1969) in the statistical analysis of gravity data.

This is, however, a highly unsatisfactory approach, and I shall try to explain why. For example *continuity in the mean implies by no means sample function continuity*! In a Poisson-process the sample functions are with probability one step functions, and the process is anyhow continuous in the mean. Later I shall give some further examples of the difference

between the sample function properties and the properties of the functions in the kernel space of the process. Fortunately, it will be found that the difference in this particular "potential" problem is not so significant after all.

5.4. *Gaussian Random Functions*

A random function or a stochastic process is said to be *gaussian, if all the finite-dimensional distributions are multivariate gaussian.* There is an infinite number of ways to define the multivariate gaussian distribution law. I shall not deal with these definition in detail here, but refer to *Rao* (1965). Now, let $(X_1, \ldots, X_n)$ be an n-dimensional random variable. *The distribution of* $(X_1, \ldots, X_n)$ *is said to be multivariate gaussian, if any linear combination of the Xes has a one-dimensional gaussian distribution.* That this really is a meaningful definition is e.g. discussed in *Rao* (1965), Chapt. VIII. Now $X = (X_1, \ldots, X_n)$ is a measurable mapping of an $(\Omega, \mathscr{A}, P)$ into a Euclidean n-dimensional space V. The above definition then says that, whenever $v^* \in V^*$, the dual space of V, the mapping

$$v^* \circ X : \Omega \to \mathbf{R}, \tag{1}$$

is a gaussian distributed real random variable. Now we can define a function $\tilde{m}$: $V^* \to \mathbf{R}$, by

$$\tilde{m}(v^*) = E(v^*(X)); \tag{2}$$

$\tilde{m}$ is obviously a linear functional on V^*, $\tilde{m} \in V^{**}$. We shall call $\tilde{m}$ *the mean value functional.* But we can very well identify V^{**} with V, and identify $\tilde{m}$ with some point $m \in V$, which we shall call *the mean value of the distribution.* It has the property:

$$E(v^*(X)) = v^*(m) \quad \forall v^* \in V^*, \tag{3}$$

which implies that $m = (m_1, \ldots, m_n)$ is the "classical mean value vector",

$$EX_i = m_i \quad i = 1, \ldots, n\,. \tag{4}$$

In a similar way we define a function $\tilde{R}$: $V^* \times V^* \to \mathbf{R}$, by

$$\tilde{R}(u^*, v^*) = E(u^*(X - m))\ (v^*(X - m))\,. \tag{5}$$

$\tilde{R}$ is a positive symmetric bilinear form on $V^* \times V^*$, the *covariance form* of the distribution. We can in a unique way find the corresponding linear mapping or operator R on V^*, so that

$$\langle Ru^*, v^* \rangle_* = \tilde{R}(u^*, v^*) \forall u^*, v^* \in V^*. \tag{6}$$

Here $\langle , \rangle_*$ denotes the usual scalar product on the Euclidean space V^*. R is a positive, symmetric operator, which we shall call the *covariance operator* of the distribution. Now R is given by some matrix (r_{ij}), where the element in i-th row and j-th column is

$$r_{ij} = \tilde{R}(e_i^*, e_j^*) = \operatorname{cov}(X_i, X_j), \tag{7}$$

where e_i^*, $i = 1, \ldots, n$ are the unit vectors in V^*, corresponding to

$$e_i^* : u \to u_i, \quad u = (u_1, \ldots, u_n) \in U. \tag{8}$$

This is what is called the covariance matrix in most statistical literature. The covariance matrix has the property of being *positive:*

$$\sum_{ij} a_i a_j r_{ij} \geqq 0 \quad \forall a_1, \ldots, a_n \in \mathbf{R}. \tag{9}$$

Any positive matrix defines a covariance operator and a gaussian distribution together with any element of V as mean value. It seems to be most natural to think of the covariance as a bilinear form, but the properties of the distribution are closely connected with the covariance as an operator, and the statistical literature mostly deals with the covariance as a matrix. I hope this will not cause too much confusion to the reader. I have here tried to describe the multivariate gaussian law in a way that makes the generalization to Hilbert space almost obvious.

Now, let ξ be a gaussian random function. *The finite-dimensional distributions are completely determined by the mean value function and the covariance function,* as $(\xi^{t_1}, \ldots, \xi^{t_n})$ follows a multivariate gaussian distribution with mean value $(m(t_1), \ldots, m(t_n))$ and covariance matrix $(r(t_i,t_j))$.

If ξ is any second-order random function, the corresponding mean value function and covariance function define the finite-dimensional distributions of a gaussian random function. The idea of making a correlation analysis of data from any second-order random function is closely related to the hope that the distributions are not too far from the gaussian distributions. For theoretical convenience we shall therefore work with gaussian random functions instead of second-order random functions. If this is a bad simplification, it is bad to make a correlation analysis.

Any positive, symmetric function is a covariance function of some gaussian random function, which is easily seen from the fact that the gaussian distributions with mean value zero and $(r(t_i,t_j))$ as covariance matrix constitute a consistent family of finite-dimensional distributions.

We have now seen that the properties of a reproducing kernel and a covariance function are exactly the same, which of course can be utilized.

Example 4.1.

Let $T = [0, 1]$, $r(s,t) = \min(s,t)$, $m(s) \equiv 0$ be respectively index set, covariance function and mean value function of a gaussian random function. This is the famous *Wiener process*, used to describe the Brownian motion. The sample functions of the Wiener process can be chosen to be continuous, i.e. the continuous functions on T, $C([0, 1])$ has outer measure 1, by the probability measure on $\mathbf{R}^{[0,1]}$, defined by the finite-dimensional distributions. But the sample functions of the Wiener process indeed behave very wildly! It can be shown that the sample functions cross any level more than a finite number of times in any small interval. I am glad, I do not have to draw one of those! But note that the kernel space of the Wiener process entirely consists of functions which are *absolutely continuous* with square-integrable derivatives, as you can see in section 4, ex. 3.2. There is, in fact, a remarkable difference between the sample functions and the functions in the kernel space!

Example 4.2. *Pure white noise:*

Let $T = \mathbf{R}$, $r(s,t) = \delta_{st}$, $m(t) \equiv 0$. Here ξ^s and ξ^t are identically independently distributed gaussian random variables with mean value zero and variance 1. The sample functions behave as wildly as functions possibly can. The kernel space consists of functions that are identically zero except in a denumerable set of points, and in such a way that $\sum_{t \in \mathbf{R}} f(t)^2 < +\infty$. The sample functions and the functions in the kernel space do not look like each other at all!

Example 4.3.

Let $T = \mathbf{R}$, $r(s,t) \equiv 1$, $m(t) \equiv 0$. This is a trivial process, where the sample functions are identically constant equal to ξ^{t_0}, and ξ^{t_0} is normally distributed (0, 1). The kernel space is easily seen to consist of identically constant functions also. Here is an example where the sample function space and the kernel space are identical. But the process is not very interesting at all.

5.5. *Convergence Concepts and Limit Theorems*

Two fundamental convergence concepts are necessary for the following. First, a sequence of random variables $\xi_1, \xi_2, \ldots$ has to be defined on some probability space $(\Omega, \mathscr{A}, P)$. We say that the sequence ξ_n *converges to* ξ *in the mean,* if ξ is a random variable on $(\Omega, \mathscr{A}, P)$ and

$$E \mid \xi_n - \xi \mid^2 \xrightarrow[n \to \infty]{} 0. \tag{1}$$

This is the same as convergence in the topology on $L_2(\Omega, \mathscr{A}, P)$ and is of course only defined for $\xi_n \in L_2(\Omega, \mathscr{A}, P)$.

We say that ξ_n *converges to* ξ *almost surely* if

$$P\{\omega \mid \lim_{n \to \infty} \xi_n(\omega) = \xi(\omega)\} = 1. \tag{2}$$

Any of the two modes of convergence implies the convergence of the distribution functions F_n of ξ_n to that of ξ, F, in every continuity point of F (weak convergence).

Another mode of convergence frequently used in connection with stochastic processes is *convergence in probability*, when ξ_n is said to converge in probability if

$$\forall \varepsilon > 0 : P\{ \mid \xi_n - \xi \mid > \varepsilon\} \xrightarrow[n \to \infty]{} 0, \tag{3}$$

but we shall not use this concept much in the present work. The following theorem will be used frequently in the next section.

Theorem 5.1. *If ξ_n is a sequence of gaussian random variables and $\xi_n \to \xi$ almost surely, then ξ_n converges to ξ in the mean.* (In fact it is enough to assume that $\xi_n \to \xi$ in probability, and the same proof will go through).

Proof. If ξ_n converges to ξ almost surely, $\xi_n(\omega)$ is a Cauchy-sequence for almost all ω, i.e.

$$\lim_{n, m \to \infty} \mid \xi_n(\omega) - \xi_m(\omega) \mid = 0. \quad \text{a.s.} \tag{4}$$

Put $\Delta_{nm} = \xi_n - \xi_m$. Δ_{nm} converges to zero a.s. and therefore also weakly.

Δ_{nm} has Fourier transform

$$E e^{it\Delta_{nm}} = e^{ita_{nm} - 1/2 t^2 \sigma^2_{nm}}, \tag{5}$$

where $a_{nm} = E\Delta_{nm}$ and $\sigma_{nm}^2 = E(\Delta_{nm} - a_{nm})^2$. Then

$$Ee^{it\Delta_{nm}} \to 1 \tag{6}$$

which implies

$$ita_{nm} - 1/2t^2\sigma_{nm}^2 \to 0. \tag{7}$$

But then $a_{nm} \to 0$ and $\sigma_{nm}^2 \to 0$ and

$$E(\Delta_{nm}^2) = \sigma_{nm}^2 + a_{nm}^2 \to 0. \tag{8}$$

(8) says that $\xi_n - \xi_m$ is a Cauchy sequence in $L_2(\Omega, \mathscr{A}, P)$, and therefore ξ_n converges to ξ^* in the mean. But $\xi_n \xrightarrow{\text{a.s.}} \xi$ and $\xi_n \xrightarrow{\text{i.m.}} \xi^*$ implies $\xi = \xi^*$ a.s., and the theorem has been proved.

We shall state another theorem without proof:

Theorem 5.2. *Let* $\xi_1, \xi_2, \ldots$ *be mutually independent random variables with variances* $\sigma_1^2, \sigma_2^2, \ldots$. *Then, if* $\sum_{n=1}^{\infty} \sigma_n^2 = \sigma^2 < +\infty$ *and* $|\sum_{n=1}^{\infty} E\xi_n| < +\infty$, $\xi = \sum_{n=1}^{\infty} \xi_n$ *is convergent almost surely and also in the mean. Moreover* $E\xi = \sum_{n=1}^{\infty} E\xi_n$, $E\xi^2 - (E\xi)^2 = \sigma^2$. *Conversely, if* $\xi = \sum_{n=1}^{\infty} \xi_n$ *converges in the mean,* $\sum_{n=1}^{\infty} \sigma_n^2 = \sigma^2 < +\infty$ *and* $\sum_{n=1}^{\infty} E\xi_n$ *converges.*

For a proof, see Doob, (1953), p. 108 *Th.* 2.3.

Remark: If especially, $\sum_{n=1}^{\infty} E\xi_n$ is absolutely convergent, i.e. if the convergence does not depend on the order of summation, $\sum_{n=1}^{\infty} \xi_n$ is also absolutely convergent, as the conditions of the theorem are fulfilled for any ordering.

6. Gaussian Measures on Hilbert Spaces. Random Functionals

The exposition of the theory of gaussian measures on Hilbert spaces given here mostly follows the one given by *Rozanov* (1968). Also *Gelfand and Vilenkin* (1964) have an exposition of the theory of gaussian measures on topological vector spaces (Chapt. IV). The cited works all deal with *countably Hilbert spaces*, i.e. topological vector spaces with a sequence of inner products, whose corresponding norms are non-decreasing, to define the topology. But, as the theory can always in the end be reduced to just Hilbert spaces, I shall not introduce these concepts here. This makes it possible to prove the results without using too much sophisticated mathematics. I hope that the reader is familiar with the most common results of the theory of compact self-adjoint operators. In *Gelfand and Vilenkin* (1964) Chapt. I, most of these results can be found.

6.1. *Hilbert-valued Random Variables*

Let U be a real, separable Hilbert space, and let $(\Omega, \mathscr{A}, P)$ be a probability space. Denote by U^* the dual space to U, consisting of all bounded linear functionals on U.

Definition 1.1. A mapping $\xi: \Omega \to U$ is said to be a *Hilbert-valued random variable* if it is weakly measurable, i.e. if

$$\forall u^* \in U^* \; \forall B \in \mathscr{B} : \{\omega \mid u^*(\xi(\omega)) \in B\} \in \mathscr{A}, \tag{1}$$

where $\mathscr{B}$ is the Borel subsets of the real numbers, or equivalently if all the mappings $u^* \circ \xi : \Omega \to \mathbf{R}$, $u^* \in U^*$ are real random variables.

It is not difficult to see that the above definition involves that ξ should be measurable with respect to the σ-algebra $\mathscr{B}(U)$ generated by the sets

$$u^{*-1}(B), \; B \in \mathscr{B}, \; u^* \in U^*. \tag{2}$$

As usual, ξ induces a probability measure P^{ξ} on $(U, \mathscr{B}(U))$ by

$$P^{\xi}(B) = P\{\omega \mid \xi(\omega) \in B\}, \; B \in \mathscr{B}(U). \tag{3}$$

We shall call P^{ξ} the *distribution* of ξ.

Definition 1.2. A Hilbert-valued random variable is said to be *gaussian,* if the random variables $u^* \circ \xi$, $u^* \in U^*$ are all gaussian random variables. The corresponding distribution P^{ξ} is said to be *a gaussian measure* on U.

A Hilbert-valued random variable ξ induces a random function $\tilde{\xi} : U^* \times \Omega \to \mathbf{R}$, defined by

$$\tilde{\xi}(u^*, \omega) = u^*(\xi(\omega)). \tag{4}$$

The sample functions have the property

$$\tilde{\xi}^{\omega}(\lambda u^* + \mu v^*) = \lambda \tilde{\xi}^{\omega}(u^*) + \mu \tilde{\xi}^{\omega}(v^*), \; \lambda, \mu \in \mathbf{R} \; u^*, v^* \in U^*. \tag{5}$$

Therefore, we shall call the process $\tilde{\xi}$ a *linear random functional* on U^*. Another way to express this is to say that U can be identified with its double-dual U^{**}, and the mapping ξ can be regarded as a mapping from Ω to U^{**}.

If the Hilbert-valued random variable ξ is *gaussian*, the corresponding linear random functional is a gaussian stochastic process. To see this, we shall convince ourselves that all the finite-dimensional distributions are multivariate gaussian. Let $u_1^*, \ldots, u_n^* \in U^*$. We have to show that $\sum_{i=1}^{n} a_i u_i^* \circ \xi$ has a one-dimensional gaussian distribution. But $u^* = \sum_{i=1}^{n} a_i u_i^* \in U^*$, and

$$\sum_{i=1}^{\infty} a_i u_i^* \circ \xi = u^* \circ \xi. \tag{6}$$

But $u^* \circ \xi$ has a one-dimensional gaussian distribution, which follows from definition 1.2. Now to compare with section 5, the finite-dimensional distributions define a measure on $\mathbf{R}^{U^*}$, and it is here obvious, that the subspace of $\mathbf{R}^{U^*}$, U^{**}, has outer measure 1, if we identify U and U^{**}. Thus we can also regard the gaussian measure as a measure on U^{**}, constructed from a measure on $\mathbf{R}^{U^*}$.

The mean value function $\tilde{m}$ of this gaussian stochastic process, obviously satisfy the property of being *linear:*

$$\tilde{m}(\lambda u^* + \mu v^*) = \lambda \tilde{m}(u^*) + \mu \tilde{m}(v^*), \tag{7}$$

and the covariance function $\tilde{R}$ that of being bilinear:

$$\tilde{R}\Big(\sum_{i=1}^{n} a_i u_i^*, \sum_{j=1}^{m} b_j v_j^*\Big) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \tilde{R}(u_i^*, v_j^*). \tag{8}$$

All the sample functions of the process $\tilde{\xi}$ are continuous, as for any $\omega \in \Omega$, the mapping

$$\tilde{\xi}^\omega : U^* \to \mathbf{R},\ \tilde{\xi}^\omega(u^*) = u^*(\xi(\omega)) \tag{9}$$

is an element of U^{**}. But then the process $\tilde{\xi}$ is *continuous in the mean*, i.e.

$$E \mid u^*(\xi) - u_0^*(\xi) \mid^2 \to 0, \text{ when } \| u^* - u_0^* \|_* \to 0. \tag{10}$$

This follows from section 5.5. But

$$E \mid u^*(\xi) - u_0^*(\xi) \mid^2 = (\tilde{m}(u^* - u_0^*))^2 + \tilde{R}(u^* - u_0^*, u^* - u_0^*), \tag{11}$$

and it then follows from (10) that $\tilde{m}$ and $\tilde{R}$ are continuous. But this means that $\tilde{m} \in U^{**}$, and as we again can identify U and U^{**}, there is a unique element $m \in U$, so that

$$\tilde{m}(u^*) = u^*(m),\ \forall u^* \in U^*. \tag{12}$$

This element m, we shall call the *mean value* of the gaussian measure or distribution P^ξ, defined by (3).

Similarly there is a unique bounded operator R on U^*, see e.g. *Gelfand and Vilenkin* (1964) p. 10, so that

$$\tilde{R}(u^*, v^*) = \langle Ru^*, v^* \rangle_*, \tag{13}$$

where $\langle , \rangle_*$, denotes the inner product in U^*. As $\tilde{R}$ has to be positive and symmetric, R also becomes *positive* and *symmetric:*

$$\langle Ru^*, v^* \rangle_* = \langle u^*, Rv^* \rangle_*, \forall u^*, v^* \in U^*, \tag{14}$$

and

$$\langle Ru^*, u^* \rangle_* \geqq 0, \forall u^* \in U^*. \tag{15}$$

The operator R is called *the covariance operator* of the gaussian measure. As the finite-dimensional distributions of the random functionals are completely determined by m and R, it follows that the gaussian measure on $(U, \mathscr{B}(U))$ is completely determined.

We shall prove that R *is a compact operator.* Now, as R is positive, it has a unique square-root $R^{1/2}$. Let $(u_n^*)_{n \in \mathbf{N}}$ be a sequence of elements of U^*, weakly converging to zero:

$$\lim_{n \to \infty} u_n^*(u) = 0, \forall u \in U. \tag{16}$$

Define the sequence of random variables

$$\xi_n(\omega) = u_n^*(\xi(\omega)). \tag{17}$$

Because of (16), we have

$$\forall \omega \in \Omega : \xi_n(\omega) \xrightarrow[n \to \infty]{} 0. \tag{18}$$

But then $\xi_n(\omega)$ must converge to zero also in the mean (see section 5.5).

$$E\xi_n(\omega)^2 = \tilde{R}(u_n^*, u_n^*) = \langle Ru_n^*, u_n^* \rangle_* = \| R^{1/2} u_n \|_*^2 \xrightarrow[n \to \infty]{} 0. \tag{19}$$

Hence weak convergence of the sequence $(u_n^*)_{n \in \mathbf{N}}$ implies strong convergence of $(R^{1/2} u_n^*)_{n \in \mathbf{N}}$ which means that $R^{1/2}$ is *a compact operator,* and a fortiori R is a compact operator.

As R is positive, symmetric and compact, there exists a complete orthonormal system $(e_n^*)_{n \in \mathbf{N}}$ in U^*, so that $(e_n^*)_{n \in \mathbf{N}}$ are eigenfunctions of the operator R:

$$\forall n \in \mathbf{N} : Re_n^* = \lambda_n e_n^*. \tag{20}$$

Because R is positive, $\lambda_n \geqq 0$, $\forall n \in \mathbf{N}$, so if we put $\lambda_n = \sigma_n^2$, we have:

$$\langle Re_i^*, e_j^* \rangle_* = \sigma_i^2 \langle e_i^*, e_j^* \rangle_* = \left\{ \begin{array}{ll} \sigma_i^2 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{array} \right\} \tag{21}$$

If we define the random variables

$$\xi_n(\omega) = e_n^*(\xi(\omega)), \tag{22}$$

we have that ξ_i and ξ_j are *stochastically independent,* whenever $i \neq j$. If we define the elements $(e_n)_{n \in \mathbf{N}}$ of U by

$$\forall u \in U : e_n^*(u) = \langle u, e_n \rangle \tag{23}$$

$(e_n)_{n \in \mathbf{N}}$ becomes a complete orthonormal system in U, as we have used the natural inner product preserving mapping from a Hilbert space to its dual. But, as $\xi(\omega) \in U$, $\forall \omega \in \Omega$, the series

$$\sum_{n \in \mathbf{N}} \langle \xi(\omega), e_n \rangle^2 = || \xi(\omega) ||^2 < +\infty \; \forall \omega \in \Omega. \tag{24}$$

Using (23) and (22) we then have that

$$\sum_{n \in \mathbf{N}} e_n^*(\xi(\omega))^2 = \sum_{n \in \mathbf{N}} \xi_n(\omega)^2 < +\infty \; \forall \; \omega \in \Omega. \tag{25}$$

But then $\sum_n \xi_n^2$ must also be convergent in the mean, and theorem 5.2 of section 5 gives that

$$\sum_n E(\xi_n - E\xi_n)^2 = \sum_n \langle Re_n^*, e_n^* \rangle_*^2 = \sum_n \sigma_n^2 < +\infty. \tag{26}$$

This means that R is a *nuclear operator* (a compact self-adjoint operator is said to be nuclear, if $\sum_{n \in \mathbf{N}} \lambda_n < +\infty$, where λ_n are the eigenvalues).

Now we shall also prove the existence of a gaussian measure to any pair (m,R), where $m \in U$ and R is a positive, symmetric, nuclear operator on U^*.

First we choose a complete orthonormal system in U^*, $(e_n^*)_{n \in \mathbf{N}}$ of eigenvectors for R, so that

$$R(e_n^*) = \sigma_n^2 e_n^*, \tag{27}$$

and therefore

$$\langle Re_i^*, e_j^* \rangle_* = \sigma_i^2 \, \delta_{ij}. \tag{28}$$

We put

$$a_n = e_n^*(m), \tag{29}$$

which means that a_n are the coefficients of m with respect to the complete orthonormal system $(e_n)_{n \in \mathbf{N}}$ in U, defined by

$$\langle u, e_n \rangle = e_n^*(u) \; \forall u \in U. \tag{30}$$

Of course

$$\sum_{n \in \mathbf{N}} a_n^2 = || m ||^2 < +\infty. \tag{31}$$

Then a sequence $(\xi_n)_{n \in \mathbf{N}}$ of random variables exists, independently normally distributed (a_n, σ_n^2). But, as $\sum_n \sigma_n^2 < +\infty$ and $\sum_n a_n^2 < +\infty$, theorem 5.2 and the remark gives that the series

$$\sum_n | \xi_n - a_n | < +\infty \text{ for } \omega \notin \Lambda, \text{ when } P(\Lambda) = 0. \tag{32}$$

We can now change the values of $\xi_n(\omega)$ in Λ, so that

$$\xi_n(\omega) = a_n, \forall \omega \in \Lambda. \tag{33}$$

(Here we have assumed that the σ-algebra $\mathscr{A}$ in the underlying probability space is *complete*, i.e. contains all subsets of sets with probability zero, but this can always be done, see e.g. *Doob* (1953), Chapt. II). The distributions of the "new" random variables will be unaltered. But then also

$$\sum_{n \in \mathbf{N}} (\xi_n(\omega) - a_n)^2 < +\infty, \forall \omega \in \Omega \tag{34}$$

and because $\sum_{n \in \mathbf{N}} a_n^2 < +\infty$, also

$$\sum_{n \in \mathbf{N}} \xi_n^2(\omega) < +\infty \ \forall \omega \in \Omega \tag{35}$$

We can then define a mapping ξ from Ω into U, so that $\xi(\omega)$ is the element of U with coefficients $\xi_n(\omega)$ with respect to the orthonormal system $(e_n)_{n \in \mathbf{N}}$. It is obvious that, for any element $u_n^* \in U^*$, where $u_n^* = \sum_{i=1}^{n} \lambda_i e_i^*$, the random variable

$$u_n^*(\xi) = \sum_{i=1}^{n} \lambda_i e_i^*(\xi) = \sum_{i=1}^{n} \lambda_i \xi_i \tag{36}$$

has a gaussian distribution with mean value

$$Eu_n^*(\xi) = \sum_{i=1}^{n} \lambda_i E\xi_i = \sum_{i=1}^{n} \lambda_i a_i = u_n^*(m) \tag{37}$$

and variance

$$V(u_n^*(\xi)) = \sum_{i=1}^{n} \lambda_i^2 V(\xi_i) = \sum_{i=1}^{n} \lambda_i^2 \sigma_i^2 = \langle Ru_n^*, u_n^* \rangle_*. \tag{38}$$

But, as $\xi(\omega) \in U, \forall \omega \in \Omega$, we have that

$$u_n^*(\xi(\omega)) \underset{n \to \infty}{\longrightarrow} u^*(\xi(\omega)) \quad \text{when} \quad \| u_n^* - u^* \|_* \underset{n \to \infty}{\longrightarrow} 0. \tag{39}$$

If we then define $u^* = \sum_{i=1}^{\infty} \lambda_i e_i^*$ and $u_n^* = \sum_{i=1}^{n} \lambda_i e_i^*$, (39) is true. But then $u_n^*(\xi)$ also converges to $u^*(\xi)$ in the mean, and $u^*(\xi)$ must have a gaussian distribution with mean value

$$Eu^*(\xi) = \lim_{n \to \infty} Eu_n^*(\xi) = \lim_{n \to \infty} u_n^*(m) = u^*(m) \tag{40}$$

and variance

$$V(u^*(\xi)) = \lim_{n \to \infty} \langle Ru_n^*, u_n^* \rangle_* = \langle Ru^*, u^* \rangle_* . \tag{41}$$

We have then proved that $u^*(\xi)$ is a gaussian real random variable for any $u^* \in U^*$, and, consequently, that ξ is a Hilbert-valued gaussian random variable. The mean value and the covariance operator are uniquely determined from their values on a complete orthonormal system and must therefore be m and R.

We summarize the results:

Summary 1.1. *Let m be an element of a real, separable Hilbert space U, and let R be an everywhere defined operator on U^*. Then there is a gaussian Hilbert valued random variable ξ on U with m as mean value and R as covariance operator if, and only if, R is positive, symmetric and nuclear. The following results hold:*

i) $\forall u^* \in U^* : u^*(\xi)$ *has a gaussian distribution*
ii) $Eu^*(\xi) = u^*(m), \forall u^* \in U^*$
iii) $E(u^*(\xi) - u^*(m))(v^*(\xi) - v^*(m)) = \langle Ru^*, v^* \rangle_*, \forall u^*, v^* \in U^*$
iv) *There is a complete orthonormal system in* U^*, $(e_n^*)_{n \in \mathbf{N}}$, *so that* $\langle Re_i^*, e_j^* \rangle_* = \sigma_i^2 \delta_{ij}, Re_i^* = \sigma_i^2 e_i^*$
v) $E \| \xi \|^2 = \text{trace } (R) = \sum_{n=1}^{\infty} \sigma_n^2 < + \infty .$

6.2. *Gaussian Random Functions and Gaussian Measures on Reproducing Kernel Hilbert Spaces*

If U is a reproducing kernel Hilbert space, we can of course specialize the results. But it will also be proved here that, if ξ is a gaussian process, *the sample functions are elements of some reproducing kernel Hilbert space if, and only if, there is a gaussian measure* on the Hilbert space that in a natural way "generates" the process. Consequently, the necessary and sufficient conditions to be fulfilled by the covariance function that the sample functions be elements of a Hilbert space with reproducing kernel are simple. It also gives a very comfortable way of dealing with linear operations on the process.

Now, let U be a separable reproducing kernel Hilbert space, consisting of real valued functions on T. Denote the kernel by k. Let (m,R) be respectively an element of U and a positive, symmetric nuclear operator on U^*, together determining a gaussian measure on U, the distribution of some Hilbert-valued random variable ξ. Choose a complete ortho-

normal system in U of eigenvectors for R, $(e_n^*)_{n\in\mathbf{N}}$. From theorem 2.4, section 4, we get that the functions

$$e_n(t) = e_n^*(k^t),\ n \in \mathbf{N} \tag{1}$$

constitute the corresponding orthonormal system in U, and

$$\langle Re_i^*, e_j^* \rangle_* = \sigma_i^2 \langle e_i^*, e_j^* \rangle_* = \sigma_i^2 \langle e_i, e_j \rangle = \sigma_i^2 \delta_{ij}. \tag{2}$$

Now let $L_t \in U^*$ be the linear functionals

$$L_t(u) = u(t),\ \forall u \in U. \tag{3}$$

Let us consider the process

$$\{L_t(\xi),\ t \in T\} = \{\xi(t),\ t \in T\}. \tag{4}$$

We get that

$$E\xi(t) = L_t(m) = m(t), \tag{5}$$

and

$$r(s,t) = \operatorname{cov}(\xi(s), \xi(t)) = \langle RL_s, L_t \rangle_* = \langle R'k^s, k^t \rangle = R'k^s(t). \tag{6}$$

Here R' is the nuclear operator on U corresponding to R on U^*, which for example could be defined as

$$R'(e_n') = \sigma_n^2\, e_n'\ \forall n \in \mathbf{N} \tag{7}$$

and extended by linearity and continuity to U. But as

$$k(s,t) = \sum_{n\in\mathbf{N}} e_n(s) e_n(t). \tag{8}$$

(theorem 2.1, section 4),

$$r(s,t) = \sum_{n\in\mathbf{N}} \sigma_n^2 e_n(s) e_n(t). \tag{9}$$

The process (4) is a gaussian process with mean value function m and covariance function r, which means that all the sample functions are elements of U. It is in a very canonical way generated from the gaussian measure (m,R). All linear operations on the process (4) can be regarded as observations from the larger process

$$\{u^*(\xi),\ u^* \in U^*\}. \tag{10}$$

From (9) it is easily seen that r is the reproducing kernel for the Hilbert space:

$$H_r = \left\{ \sum_{i=1}^{\infty} a_i \sigma_i e_i \mid \sum_{i=1}^{\infty} a_i^2 < +\infty \right\} \tag{11}$$

with the inner product

$$\langle \sum_i a_i \sigma_i e_i, \sum_j b_j \sigma_j e_j \rangle_r = \sum_i a_i b_i. \tag{12}$$

It is easy to see that $H_r \subseteq U$, and, in general, $H_r \subset U$ (inclusions refer to the underlying sets), as e.g. $\sum_i \sigma_i e_i \in U$, as $\sum_i \sigma_i^2 < +\infty$, but $\sum_i \sigma_i e_i \notin H_r$. H_r is isomorphic to the Hilbert space *spanned by the process* (4). Let now m and r be mean value function and covariance function for some gaussian process ξ, whose sample functions belong to some reproducing kernel Hilbert space U with probability one and $m \in U$:

$$\exists \Lambda \in \mathscr{A} : P(\Lambda) = 0 : \forall \omega \notin \Lambda : \xi^\omega \in U. \tag{13}$$

We can change ξ in Λ, so that e.g. $\xi^\omega \equiv 0$ for $\omega \in \Lambda$, without changing m and r. As $\{L_t, t \in T\}$ is *total* in U^*, i.e. the closed linear manifold spanned by $\{L_t, t \in T\}$ is U itself, any element $u^* \in U^*$ can be regarded as a limit of elements u_n^* of the form

$$u_n^* = \sum_{i=1}^{n} a_i L_{t_i}. \tag{14}$$

It is easy to see that

$$E u_n^*(\xi) = \sum_{i=1}^{n} a_i L_{t_i}(m) = u_n^*(m) \tag{15}$$

and that

$$V(u_n^*(\xi)) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j r(t_i, t_j) \tag{16}$$

when $u_n^* \to u^*$,

$$\forall \omega \in \Omega : u_n^*(\xi^\omega) \to u^*(\xi^\omega). \tag{17}$$

But then the convergence (17) is also in the mean, and

$$\left.\begin{aligned} V(u_n^*(\xi)) \to V(u^*(\xi)) &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_i a_j r(t_i, t_j) \\ E(u_n^*(\xi)) \to E(u^*(\xi)) &= u^*(m). \end{aligned}\right\} \tag{18}$$

But now we have shown that the mapping

$$\omega \to \xi^\omega \tag{19}$$

defines a Hilbert-valued random variable on U, and, consequently, there is a complete orthonormal system in U, so that

$$r(s,t) = \sum_{n \in \mathbf{N}} \sigma_n^2 e_n(s) e_n(t) \tag{20}$$

where

$$\sum_{n \in \mathbf{N}} \sigma_n^2 < +\infty. \tag{21}$$

If, on the other hand, r has the expression (20) for some orthonormal system $\{e_n, n \in \mathbf{N}\}$, it is obvious that there is a gaussian measure on U with covariance operator R so that $r(s,t) = \langle RL_t, L_s \rangle_*$, as we can define

$$\bar{R}(e_n^*) = \sigma_n^2 e_n^* \tag{22}$$

and by linearity and continuity extend $\bar{R}$ to an operator R which is seen to be a covariance operator because of (21). We can summarize the results in the following theorem:

Theorem 2.1. *Let U be a real, separable Hilbert space with the reproducing kernel $k : T \times T \to \mathbf{R}$. Let r be a positive, symmetric function $r : T \times T \to \mathbf{R}$. The following statements are then equivalent:*

i) *There is a complete orthonormal system $(e_n)_{n \in \mathbf{N}}$ in U, so that*

$$r(s,t) = \sum_{n \in \mathbf{N}} \sigma_n^2 e_n(s) e_n(t) \; \forall s,t \in T, \; \sum_{n \in \mathbf{N}} \sigma_n^2 < +\infty.$$

ii) *There is a gaussian process with mean value 0 and covariance function r, the sample functions of which belong to U with probability one.*

iii) *There is a gaussian measure on U with covariance operator R, so that*

$$r(s,t) = \langle r_t, k_s \rangle = \langle RL_t, L_s \rangle_*.$$

We have constructed a method to get from a process $\{\xi^t, t \in T\}$ with sample functions in a reproducing kernel Hilbert space U to a random functional $\{u^*(\xi), u^* \in U^*\}$ in a way that makes the first process canonically embedded in the random functional by the equation $\xi^t = L_t(\xi)$. It is important that U has a r.k. as $L_t \in U^* \Leftrightarrow U$ has r.k. (Theorem 1.2, section 4). It is possible also to get from e.g. $\{\xi^t, t \in [0,1]\}$ where $\xi^\omega \in L^2[0,1]$ a.s. to a random functional $\{u^*(\xi), u^* \in U^*\}$. But here $\{\xi^t, t \in [0,1]\}$ will not be embedded in the functional.

It is also easy to see that $r(s,t) = k(s,t)$ implies $\sigma_n^2 = 1 \; \forall \, n \in \mathbf{N}$ in i), and therefore there is *no* gaussian process with mean value zero and covariance function r the sample functions of which belong to U with probability one.

7. Topological Groups and Their Representations

7.1. *Topological Groups and Haar Measure*

A *topological group* G is a group with a topology that makes the group operations continuous.

Let G in the following be a locally compact (Hausdorff) topological group, satisfying the second axiom of countability. We shall consider the transformations of G, $\delta(g)$ and $\gamma(g)$, defined for $g \in G$ by:

$$\delta(g): x \to xg \quad x,g \in G \tag{1}$$

$$\gamma(g): x \to gx \quad x,g \in G \tag{2}$$

$\delta(g)$, $g \in G$ are called the *right-translations* and $\gamma(g)$, $g \in G$ the *left-translations*. If μ is a measure on the Borel-sets $\mathscr{B}_G$ of G (the σ-algebra generated by the closed sets of G), μ is said to be *right-invariant* if

$$\mu(B) = \mu(\delta(g)\,(B)),\ \forall g \in G,\ \forall B \in \mathscr{B}_G, \tag{3}$$

and *left-invariant* if

$$\mu(B) = \mu(\gamma(g)\,(B)),\ \forall g \in G,\ \forall B \in \mathscr{B}_G. \tag{4}$$

The following theorem is fundamental:

Theorem 1.1. *Up to a constant factor there is exactly one right-invariant measure λ and exactly one left-invariant measure μ on G.*

λ is called the *right Haar measure* and μ the *left Haar measure*.

If G is *commutative*, $\lambda = \mu$.

But also, if G is just compact, $\lambda = \mu$ and λ is *finite*, i.e. $\lambda(G) < +\infty$. In fact, λ is finite if, and only if, G is compact. This means that on a compact group, there is a *unique* Haar *probability measure*, i.e. $\lambda(G) = 1$.

Example 1.2.

Let $G = (\mathbf{R}, +)$ the real numbers with addition as group composition. G is locally compact and commutative, and the Haar-measure is the Lebesgue-measure.

Example 1.3.

Let $G = (\{ z \in \mathbf{C} \mid |z| = 1 \}, \cdot)$ the complex unit circle with multiplication as group composition. G is compact and commutative, and the Haar probability measure is the uniform distribution over the unit circle in the complex plane.

Example 1.4.

Let $G = (\{e^{\frac{ip2\pi}{n}}, p \in \mathbf{Z}\}, .)$. G is finite, compact (discrete topology), and commutative. The Haar probability measure is "counting measure" assigning the probability $\frac{1}{n}$ to any of the n points of G.

The theory of Haar measure is very fundamental in many fields of probability and statistics. I shall refer to *Bourbaki* (1965), for proofs and details.

7.2. *Representations of Topological Groups*

Let G be a topological group and let U be a separable Hilbert space. By $O(U)$ we denote the unitary group on U, i.e. the group of all unitary operators.

Definition 2.1. A mapping $T: G \to O(U)$ is said to be a *representation* of G, if it is strongly continuous and homomorphic, i.e.

i) $\forall u \in U: g \to T(g)u$ is continuous (in the norm topology on U)
ii) $\forall g_1, g_2 \in G: T(g_1)T(g_2) = T(g_1 g_2)$.

The mappings $t_{u,v}: g \to t_{u,v}(g) = \langle T(g)u, v \rangle$ for $u, v \in U$ are called the *coefficients* of T. As the strong and weak topology on $O(U)$ are identical, it is the same as saying that all the coefficients should be continuous.

Note the following relations:

$$T(g^{-1}) = T(g)^{-1} = T^*(g); \; T(e) = I. \tag{1}$$

Here T^* is the adjoint operator to T, I is the identity on U, and e is the neutral element of G. Furthermore

$$t_{u,v}(g) = \langle T(g)u,v\rangle = \langle u,T(g^{-1})v\rangle = \overline{t_{v,u}(g^{-1})} \tag{2}$$

$$t_{u,v}(e) = \langle u,v\rangle. \tag{3}$$

Example 2.2.

Let $G = O(U)$, define $T(A) = A$, $\forall A \in O(U)$. T is a representation (trivial).

Example 2.3.

Let λ be right Haar measure on G. $L_2(G)$ is a separable Hilbert space, consisting of the functions on G, square-integrable with respect to λ. Define for $g \in G$ the mapping $T(g) = L_2(G) \to L_2(G)$ given by:

$$T(g)f(x) = f(xg^{-1}) \; \forall x \in G. \tag{4}$$

$T(g)$ is easily seen to be linear. $T(g)$ is unitary because of the right invariance of λ. T is a homomorphism:

$$\left.\begin{aligned} T(g_1)T(g_2)f(x) = T(g_1)f(xg_2^{-1}) = f(xg_2^{-1}g_1^{-1}) = f(x(g_1g_2)^{-1}) = \\ T(g_1g_2)f(x). \end{aligned}\right\} \tag{5}$$

$T(g)f$ is also continuous. I shall not prove this, but it is not difficult. T is a representation. This representation is called the *regular right representation* of G.

Sometimes we have to deal with different representations of the same group. If T is a representation, we shall denote the corresponding Hilbert space by U_T.

Definition 2.4. Let T be a representation of G. By $\dim T$ we mean the *dimension* of U_T.

Definition 2.5. Let S and T be representations of the same group G. S and T are said to be *equivalent* ($S \cong T$), if there exists an isometric one-to-one and onto operator $A: U_S \to U_T$, so that

$$\forall g \in G: \; AS(g) = T(g)A \tag{6}$$

Of course

$$S \cong T \Rightarrow \dim S = \dim T. \tag{7}$$

Example 2.6.

Let T be a representation of G. Define S by

$$U_S = U_T \tag{8}$$

$$S(g) = T(g_0 g g_0^{-1}),\ \forall g \in G \tag{9}$$

where g_0 is an arbitrary but fixed element of G. Obviously $S \cong T$, which is seen when defining $A = T(g_0^{-1})$:

$$AS(g) = T(g_0^{-1})T(g_0 g g_0^{-1}) = T(g g_0^{-1}) = T(g)A. \tag{10}$$

A subspace $F \subseteq U$ is said to be stable for some bounded operator A if $A(F) \subseteq F$.

If F is stable for all the operators $T(g)$, $g \in G$, where T is some representation, the same is true for $F^{\perp}$ and $F^{\perp\perp} = \bar{F}$. F is said to be *stable for* T. Let F be stable for T and closed. Let us consider the mapping $g \to T(g) \mid F$. ($T(g) \mid F$ is the restriction of $T(g)$ to F). As F is closed, it is a Hilbert space itself and the mapping $g \to T(g) \mid F$ is a representation of G into $O(F)$. This representation we shall denote $T \mid F$. We have $U_{T \mid F} = F$.

Definition 2.7. $T \mid F$ is called a *subrepresentation* of T. The subrepresentation is called *trivial* if $F = \{\mathbf{0}\}$ or if $F = U_T$, otherwise it is called *non-trivial*.

Definition 2.8. If T has only trivial subrepresentations and $U_T \neq \{\mathbf{0}\}$, T is called *irreducible*, otherwise *reducible*.

Example 2.9.

Let G be any group, and let $U_T = \mathbf{C}^2$. Define $T(g) = I$, $\forall g \in G$. Any subspace of $\mathbf{C}^2$ is stable for T, and T is therefore reducible. If $U_T = \mathbf{C}^1$, i.e. $\dim T = 1$, T is of course irreducible.

Let $(T_n, n \in \mathbf{N})$ be a family of representations of the same group G. We can construct the direct orthogonal sum of the corresponding Hilbert spaces $U = \bigoplus_{n \in \mathbf{N}} U_{T_n}$. We shall write an element u af U as $u = (u_n)_{n \in \mathbf{N}}$ where $u_n \in U_{T_n}$. Define for $g \in G$ $T(g)u$ by

$$T(g)(u_n)_{n \in \mathbf{N}} = (T_n(g)u_n)_{n \in \mathbf{N}}, \tag{11}$$

$T: G \to O(U)$ becomes a representation of G. We shall write $T = \bigoplus_{n \in \mathbf{N}} T_n$.

If T is a representation of G, and F is a proper stable subspace of U_T, $F \oplus F^\perp$ is isomorphic to U_T and $T \cong T \mid F \oplus T \mid F^\perp$, which is seen when using the isomorphism from U_T to $F \oplus F^\perp$. If $T|F$ or $T|F^\perp$ are reducible, we can again decompose it into two components and so on. If $T \cong \bigoplus_{n \in \mathbf{N}} T_n$, where T_n are all irreducible, we say that T has been decomposed into irreducible components.

Let F be a closed subspace of U. It is well known, that F is stable for an operator A, if and only if A and P_F (the projection operator corresponding to F) commute, we shall primarily consider the case, in which G is *compact*. The following theorem is then important:

Theorem 2.10. *Any representation of a compact group is direct sum of denumerably many irreducible representations of finite dimension.*

And as a corollary:

Corollary 2.11. *Any irreducible representation of a compact group has finite dimension.*

Any representation of a compact group then has the following decomposition

$$T = \bigoplus_{i=1}^{\infty} \bigoplus_{j=1}^{n_i} T_{ij} \tag{12}$$

where T_{ij} are irreducible and $T_{ij} \cong T_{ik}$. The sequence $n_1, n_2, \ldots$ is uniquely determined up to permutation, and the same is true for dim $(\bigoplus_{j=1}^{n_i} T_{ij})$. But T_{ij} is only uniquely determined when $n_i = 1$. (Note that the n_i's could be cardinal numbers of **N**).

Example 2.12.

Let G be the group from example 1.3, and let $T(g) = I$ on $U = \mathbf{C}^2$. T can be decomposed in any way you like as $2T_1$, where $T_1(g) = I$ on $\mathbf{C}$.

For proofs and details the reader is referred to *Dixmier* (1964).

7.3. *Irreducible Representations of Compact Groups*

Let T be an irreducible representation of a compact group G. This means that the dimension of U_T is *finite*. Now every $T(g)$ is a unitary linear mapping of U_T onto U_T. If dim $T = n$, there is a complete ortho-

normal system $e_1, \ldots, e_n$. In this orthonormal system, $T(g)$ is determined by a unitary matrix $(a_{ij}(g))$, where

$$a_{ij}(g) = \langle T(g)e_i, e_j \rangle. \tag{1}$$

But this is what we called *coefficients* of the representation.

$$a_{ij}(g) = t_{e_i, e_j}(g) \tag{2}$$

What becomes important to us are certain simple integration properties of these coefficients. Let μ be the Haar probability measure on G. Let $n = \dim T$, and let T be irreducible; then the following relations hold:

$$\int_G t_{e_i, e_j}(g)\, t_{e_k, e_l}(g^{-1})\, d\mu(g) = \frac{1}{n}\delta_{il}\delta_{jk} \tag{3}$$

Or, using (2), section 7.2

$$\int_G t_{e_i, e_j}(g)\, \overline{t_{e_l, e_k}(g)}\, d\mu(g) = \frac{1}{n}\delta_{il}\delta_{jk}. \tag{4}$$

Furthermore, if S and T are inequivalent irreducible representations of the same compact group G, and s and t are the corresponding coefficients:

$$\int_G s_{e_i, e_j}(g)\, t_{f_k, f_l}(g)\, d\mu(g) = 0 \tag{5}$$

where $e_1, \ldots, e_n$ and $f_1, \ldots, f_m$ are orthonormal systems in U_S, respectively U_T. What we shall in fact use, is a result about integration of reducible representations of a certain type.

Theorem 3.1. *Let T be a representation of a compact group G, so that $T \cong \bigoplus_{i=1}^{\infty} T_i$, where T_i and T_j are non-equivalent whenever $i \neq j$. Let e_{ij}, $i = 1,2, \ldots, j = 1, \ldots, n_i = \dim T_i$, be a complete orthonormal system in U_T so that e_{ij}, $j = 1, \ldots, n_i$ are orthonormal systems in U_{T_i}. Let $u, v, x, y \in U_T$ with $\langle u, e_{ij} \rangle = u_{ij}$, $\langle v, e_{ij} \rangle = v_{ij}$, $\langle x, e_{ij} \rangle = x_{ij}$, $\langle y, e_{ij} \rangle = y_{ij}$. Then:*

$$\int_G t_{u,v}(g)\, \overline{t_{x,y}(g)}\, d\mu(g) = \sum_{i=1}^{\infty} \frac{1}{n_i} \sum_{j=1}^{n_i} \sum_{k=1}^{n_i} u_{ij}\, y_{ik}\, \bar{v}_{ik}\, \bar{x}_{ij}. \tag{6}$$

Proof: The proof is just arithmetic calculation. Note that

$$\left.\begin{aligned} t_{u,v}(g) &= \langle T(g)u, v\rangle = \sum_{i=1}^{\infty}\sum_{j=1}^{n_i} u_{ij}\langle T(g)e_{ij}, v\rangle \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{n_i}\sum_{k=1}^{\infty}\sum_{l=1}^{n_i} u_{ij}\overline{v_{kl}}\langle T(g)e_{ij}, e_{kl}\rangle . \end{aligned}\right\} \tag{7}$$

But, as $T(g)(U_{T_i}) \subseteq U_{T_i}$ and $U_{T_i} \perp U_{T_k}$ for $i \neq k$, $\langle T(g)e_{ij}, e_{kl}\rangle = 0$ for $i \neq k$, and we get:

$$t_{u,v}(g) = \sum_{i=1}^{\infty}\sum_{j=1}^{n_i}\sum_{k=1}^{n_i} u_{ij}\bar{v}_{ik}t_{e_{ij},e_{ik}}(g) \tag{8}$$

and analogously:

$$\overline{t_{x,y}(g)} = \sum_{p=1}^{\infty}\sum_{q=1}^{n_p}\sum_{r=1}^{n_p} \bar{x}_{pq}y_{pr}\,\overline{t_{e_{pq},e_{pr}}(g)}. \tag{9}$$

As the coefficients are continuous on G, and $|\, t_{e_{ij},\,e_{kl}}(g)\,| \leqq 1, \forall g \in G$, there is no problem in changing the order of integration and summation, and we get:

$$\left.\begin{aligned} &\int_G t_{u,v}(g)\overline{t_{x,y}\,(g)}\,d\mu(g) = \\ &\sum_{i=1}^{\infty}\sum_{=1}^{n_i}\sum_{k=1}^{n_i}\sum_{p=1}^{\infty}\sum_{q=1}^{n_p}\sum_{r=1}^{n_p} u_{ij}\bar{v}_{ik}\bar{x}_{pq}y_{pr}\left(\int_G t_{e_{ij},e_{ik}}(g)\overline{t_{e_{pq},e_{pr}}(g)}\,d\mu(g)\right). \end{aligned}\right\} \tag{10}$$

But, as T_i and T_p are inequivalent for $i \neq p$, we can use (5) and remove all terms in the sum with $i \neq p$. Furthermore, we can use (4) and remove all terms with $j \neq q$ or $k \neq r$. All other integrals are equal to $\frac{1}{n_i}$, also from (4). We get:

$$\int_G t_{u,v}\overline{t_{x,y}} = \sum_{i=1}^{\infty}\sum_{j=1}^{n_i}\sum_{k=1}^{n_i} u_{ij}\bar{v}_{ik}\bar{x}_{ij}y_{ik}\frac{1}{n_i}, \tag{11}$$

which proves the result. Again, the reader is referred to *Dixmier* (1964) for proofs of (4) and (5).

8. Homogeneous Random Functionals and Unitary Representations

8.1. *General Case*

To draw statistical inference from observations of random functions, it is necessary to get some reduction of the model. One of the most important reductions in connection with random functions on the real axis is *stationarity*. A random function with $T = \mathbf{R}$ is said to be stationary, if the finite-dimensional distributions are invariant with respect to translations of the real axis, i.e. if the distributions of $\xi^{t_1}, \ldots, \xi^{t_n}$ and $\xi^{t_1+s}, \ldots, \xi^{t_n+s}$ are the same for any $s \in \mathbf{R}$. The real axis is a commutative locally compact Hausdorff topological group, and stationarity means that the translations of the group do not affect the distributions.

We shall here try to define a similar concept in the case of a gaussian random functional, which we shall call *homogeneity*. What we need is a group acting on the dual space to U, where U is a Hilbert space with a gaussian measure. We shall furthermore demand that this group consists of *unitary operators* on U^*. It is my opinion that otherwise the concept has no practical meaning, and in this particular context it is all that I have to use. It seems to be most convenient to study such groups of unitary operators as *representations* of groups, because it is my feeling that they will be given this way from the nature of the problem dealt with, as we shall see, making the model for the potential of the earth.

Now, let U be a real, separable Hilbert space, U^* its dual, and let m and R be respectively mean value and covariance operator for a gaussian measure on U. Let ξ be a gaussian Hilbert-valued random variable on U, with this gaussian measure as distribution. Let G be a topological group, (locally) compact, satisfying 2. axiom of countability, and let $T: G \to O(U^*)$ be a unitary representation of G. For $u_1^*, \ldots, u_n^* \in U^*$, the distribution of $(u_1^*(\xi), \ldots, u_n^*(\xi))$ is multivariate gaussian, and the same is true for $(T(g)u_1^*(\xi), \ldots, T(g)u_n^*(\xi))$ for any $g \in G$.

Definition 1.1. The random functional ξ is said to be *T-homogeneous* if, for any $u_1^*, \ldots, u_n^* \in U^*$ and any $g \in G$, *the distributions of* $(u_1^*(\xi), \ldots, u_n^*(\xi))$ *and* $(T(g)u_1^*(\xi), \ldots, T(g)u_n^*(\xi))$ *are identical.*

The mean value of $u^*(\xi)$ is $u^*(m)$ (Theorem 1.1, section 6). If ξ has to be T-homogeneous, we must have

$$T(g)u^*(m) = u^*(m), \forall g \in G. \tag{1}$$

But also the covariance of $u^*(\xi)$ and $v^*(\xi)$ must be left unchanged by $T(g)$:

$$\langle Ru^*, v^* \rangle_* = \langle RT(g)u^*, T(g)\, v^* \rangle_*. \tag{2}$$

But as $T(g)$ is unitary (2) is equivalent to

$$\left.\begin{aligned} \langle Ru^*, v^* \rangle_* &= \langle T(g)^{-1} RT(g)u^*, v^* \rangle_* \Leftrightarrow \\ \langle (R - T(g)^{-1} RT(g))u^*, v^* \rangle &= 0 \ \forall u^*, v^* \in U^*. \end{aligned}\right\} \tag{3}$$

But this cannot be true unless

$$(R - T(g)^{-1} RT(g))u^* = 0 \ \forall\, u^* \in U^*, \tag{4}$$

which again means that

$$R = T(g)^{-1}\, RT(g) \Leftrightarrow T(g)R = RT(g). \tag{5}$$

If, on the other side, R and $T(g)$ commute,

$$\langle RT(g)u^*, T(g)v^* \rangle = \langle T(g)\, Ru^*, T(g)\, v^* \rangle_* = \langle Ru^*, v^* \rangle_* \tag{6}$$

and the covariance between $u^*(\xi)$ and $v^*(\xi)$ is left unchanged by T.

By the canonical isomorphism between U^* and U, T is transformed into an equivalent representation $\tilde{T}: G \to O(U)$, where J is the isometry $J: U^* \to U$,

$$\tilde{T}(g)J = JT(g) \ \forall\, g \in G, \tag{7}$$

(1) is then equivalent to

$$\left.\begin{aligned} T(g)u^*(m) = u^*(m) &= \langle J(u^*), m \rangle \Leftrightarrow \\ \langle u^*, J^{-1}(m) \rangle_* &= \langle T(g)u^*, J^{-1}(m) \rangle_* = \langle J(T(g)u^*), m \rangle \\ &= \langle \tilde{T}(g)J(u^*), m \rangle = \langle J(u^*), \tilde{T}(g^{-1})m \rangle. \end{aligned}\right\} \tag{8}$$

But this is true if, and only if,

$$\tilde{T}(g)m = m \ \forall\, g \in G. \tag{9}$$

We have proved the following:

Theorem 1.2. *The random functional ξ is T-homogeneous if, and only if, the following two conditions are satisfied*

i) $RT(g) = T(g)R \quad \forall g \in G$

ii) $\tilde{T}(g)m = m \qquad \forall g \in G,$

(where $\tilde{T}(g) = JT(g)J^{-1}$ is the to $T(g)$ by the canonical isomorphism between a Hilbert space and its dual corresponding operator).

If we take a cylinder set C on U,

$$C = \{ u \in U \mid (u_1^*(u), \ldots, u_n^*(u)) \in B \} \tag{10}$$

where B is a Borel set in $\mathbf{R}^n$, the probability of C is determined by the distribution of $(u_1^*(\xi), \ldots, u_n^*(\xi))$.

$$P^{\xi}(C) = P\{\omega \in \Omega \mid (u_1^*(\xi(\omega)), \ldots, u_n^*(\xi(\omega)) \in B\}. \tag{11}$$

Let us consider the set $\tilde{T}(g)C$:

$$\tilde{T}(g)C = \{u \in U \mid (u_1^*(\tilde{T}(g)^{-1}u), \ldots, u_n^*(\tilde{T}(g)^{-1}u)) \in B\} \tag{12}$$

as

$$\left.\begin{aligned} u^*(\tilde{T}(g)u) &= \langle J(u^*), \tilde{T}(g)^{-1}u\rangle = \langle \tilde{T}(g)\, Ju^*, u\rangle \\ &= \langle J(T(g)u^*), u\rangle = T(g)u^*(u) \end{aligned}\right\} \tag{13}$$

we have

$$\tilde{T}(g)C = \{u \in U \mid (T(g)u_1^*(u), \ldots, T(g)u_n^*(u)) \in B\}. \tag{14}$$

If ξ is T-homogeneous, we get from (11) that

$$P^{\xi}(C) = P^{\xi}(\tilde{T}(g)C) \quad \forall g \in G. \tag{15}$$

Conversely, if (15) is true for any cylinder set C, ξ obviously is T-homogeneous. We therefore have the following theorem:

Theorem 1.3. *The random function ξ is T-homogeneous if, and only if, $\tilde{T}(g)$ is measure-preserving for any $g \in G$.*

Intuitively this means that elements u and v in U have equal probability if $u = T(g)v$. This is the most convenient way to think of homogeneity if you want to understand what this actually means.

We shall use a lemma:

Lemma 1.4. *Let R be a positive symmetric compact operator and A a bounded operator on a Hilbert space U. Then A and R commute if, and only if,*

$$\forall \lambda > 0\colon\ A((R - \lambda I)^{-1}(\mathbf{0})) \subseteq (R - \lambda I)^{-1}(\mathbf{0}). \tag{16}$$

Proof: Suppose $AR = RA$. Let $u \in (R - \lambda I)^{-1}(\mathbf{0})$, which means that $Ru = \lambda u$. But

$$R(Au) = A(Ru) = A(\lambda u) = \lambda(Au) \tag{17}$$

therefore $Au \in (R - \lambda I)^{-1}(\mathbf{0})$. Conversely, suppose $A(R - \lambda I)^{-1}(\mathbf{0})) \subseteq (R - \lambda I)^{-1}(\mathbf{0})$. There is a denumerable set $(\lambda_n, n \in \mathbf{N})$ of eigenvalues of R, ɔ: so that $(R - \lambda_n I)^{-1}(\mathbf{0}) \neq \{\mathbf{0}\}$. It is well known that for $\lambda_n \neq 0$ the dimension of $(R - \lambda_n I)^{-1}(\mathbf{0})$ is finite and $U = \bigoplus_{n \in \mathbf{N}} (R - \lambda_n I)^{-1}(\mathbf{0})$. If we by P_n denote the projection on $(R - \lambda_n I)^{-1}(\mathbf{0})$ we have that, if $u \in U$, $u = \sum_{n \in \mathbf{N}} P_n u$. We get

$$ARu = A(\sum_{n \in \mathbf{N}} RP_n u) = A(\sum_{n \in \mathbf{N}} \lambda_n P_n u) \tag{18}$$

$$RAu = R(\sum_{n \in \mathbf{N}} A(P_n u)) = \sum_{n \in \mathbf{N}} \lambda_n A P_n u = ARu. \tag{19}$$

The second equality in (19) follows from the fact that $A(P_n u) \in (R - \lambda_n I)^{-1}(\mathbf{0})$. If we think of R as a covariance operator and $A = T(g)$, we immediately get the following theorem:

Theorem 1.5. *Let ξ be a random functional with zero mean value and R as covariance operator with $\sigma_n{}^2$, $n \in \mathbf{N}$ as eigenvalues. Then ξ is T-homogeneous if, and only if, $T \mid (R - \sigma_n^2 I)^{-1}(\mathbf{0})$ is a subrepresentation of T.*

If we standardize $\sigma_1^2 = 0$, we have that $\dim (R - \sigma_n^2 I)^{-1}(\mathbf{0}) < +\infty$ for $n \geqq 2$. We now have the decomposition of T:

$$T = \bigoplus_{n=1}^{\infty} T_n \tag{20}$$

where $T_n = T \mid (R - \sigma_n^2 I)^{-1}(\mathbf{0})$ and $\dim T_n < +\infty$ for $n \neq 1$. T_n are not necessarily irreducible, but can be decomposed into irreducible components

$$T_n = \bigoplus_{i=1}^{\nu_n} T_{in} \tag{21}$$

with $\nu_n < +\infty$ and $\dim T_{in} < +\infty$ for $n \neq 1$.

If we take the special case of $R^{-1}(\mathbf{0}) = \{\mathbf{0}\}$ it is worth noting that the only possible representations giving a T-homogeneous functional are those which can be decomposed into denumerable many irreducible representations of *finite* dimension. We ought to make a summary of the results in this part:

Summary 1.6. *Let ξ be a random functional with mean value zero and covariance operator R. Let T be a representation of a group G. The following statements are then equivalent:*

i) *ξ is T-homogeneous*

ii) *$\forall g \in G\ \forall u^* \in U^*$: $u^*(\xi)$ and $T(g)u^*(\xi)$ have the same distribution*

iii) $\forall g \in G$: $RT(g) = T(g)R$

iv) $\forall B \in \mathscr{B}(U)$: $P^{\xi}(B) = P^{\xi}(\tilde{T}(g)B)$

v) *T has the decomposition*

$$T = \bigoplus_{n=1}^{\infty} \left(\bigoplus_{i=1}^{\nu_n} T_{in} \right)$$

with ν_n, $\dim T_{in} < +\infty$ if $n \neq 1$ and

$$\bigoplus_{i=1}^{\nu_n} T_{in} = T \mid (R - \sigma_n^2 I)^{-1}(\mathbf{0}).$$

(ii) *follows from the fact that the multivariate distributions are determined from the one-dimensional distributions*).

8.2. *The Case of a Compact Group*

Suppose T is a representation of a *compact* group. Then T has a decomposition into irreducible components of finite dimension

$$T \cong \bigoplus_{i=1}^{\infty} \bigoplus_{j=1}^{n_i} T_{ij}, \tag{1}$$

where

$$U_T \cong \bigoplus_{i=1}^{\infty} \bigoplus_{j=1}^{n_i} U_{T_{ij}} = U^*, \tag{2}$$

and T_{ij} and T_{kl} are equivalent if $i = k$ and non-equivalent otherwise. If R is the covariance operator for a T-homogeneous functional, we also have the decomposition:

$$T = \bigoplus_{n=1}^{\infty} T_n,\ T_n = T \mid (R - \sigma_n^2 I)^{-1}(\mathbf{0}). \tag{3}$$

T_n are not necessarily irreducible, but $\dim T_n < +\infty$ if $\sigma_n^2 \neq 0$. We can obviously choose a decomposition (1), so that

$$\forall\, i,j\ \exists \nu(i,j)\colon U_{T_{ij}} \subseteq U_{T_{\nu(i,j)}} \tag{4}$$

by first making the decomposition (4) and then decompose T_n into its irreducible components:

$$T = \bigoplus_{n=1}^{\infty} \Big(\bigoplus_{\nu(i,j)=n} T_{ij} \Big). \tag{5}$$

If especially $n_i = 1$, $\forall i \in \mathbf{N}$ we can first *uniquely* decompose T as $T = \bigoplus_{i=1} T_i'$. From (4) it is easily seen that all vectors in some particular $U_{T_i'}$ *must* be eigenvectors of R corresponding to the *same* eigenvalue. Choosing complete orthonormal systems for $U_{T_i'}$ we get a complete orthonormal system *of eigenvectors* of R. Thus building a model for some random functional, an invariance with respect to a compact group will immediately tell us which are the eigenvectors of R and which correspond to the same eigenvalue and will in this way give us the useful diagonalization of R. If R is supposed to be unknown and estimated, R is parametrized by its eigenvalues (σ_n^2). The above reduction also reduces the number of unknown eigenvalues, as some of them which correspond to the same invariant subspace have to be equal. However, if the Hilbert space is infinite-dimensional, the number of different eigenvalues will still be infinite, as all invariant subspaces have finite dimension.

8.3. *Ergodicity and Estimation*

If $\{ \xi^t,\ t \in \mathbf{R} \}$ is a stationary gaussian random function on the real axis with mean value function identically zero and unknown covariance function, one will often want to *estimate* the covariance function from empirical observations of a part of the process.

It is of course difficult to estimate anything from *one single sample function* but because of the stationarity the two-dimensional distribution of (ξ^t, ξ^s) and (ξ^{t+h}, ξ^{s+h}) is identical for any u. This can somehow be regarded as "repetitions" of the same experiment, and statistics suddenly get a meaning. The problem is that the "repetitions" are not necessarily independent, and the law of large numbers does not apply immediately.

Suppose we have observed ξ^t in some large interval, say from $-T$ to T. It would be natural to suggest the following the following estimator for $E\xi^{t+h}\,\xi^t = r(h)$:

$$\hat{r}(h) = \frac{1}{2(T-h)} \int_{-T+h}^{T-h} \xi^t \xi^{t+h} dt, \tag{1}$$

the natural generalization of an average. $\hat{r}(h)$ is of course unbiased

$$E\hat{r}(h) = \frac{1}{2(T-h)} \int_{-T+h}^{T-h} r(h) dt = r(h). \tag{2}$$

When does this estimator in some sense converge to the true value for $T \to \infty$? This is the contents of one of the many "statistical" ergodic theorems. It does, roughly speaking, converge if the ξ^t-s are not too much dependent. It especially converges if the process is an ergodic process (also sometimes denoted as "metrically transitive") see e.g. *Doob* (1953).

In most geodetic literature the covariance function is *defined* averaging over the sphere instead of averaging with respect to the underlying probability measure, which implicitly assumes some kind of ergodicity. *This assumption cannot be fulfilled*, which in fact is not difficult to prove. This is important. It means, that it is needless work to measure and measure gravity anomalies, deflections of the vertical and so on, in order to determine the covariance function, and that the evaluations of "uncertainty" of the covariance function determined are completely misleading.

Now, let ξ be a random functional with zero mean value and covariance operator R: $U^* \to U^*$. Let G be a *compact* group and $T: G \to O(U^*)$ a representation of G. Suppose ξ is T-homogeneous and T is a representation of the type with all irreducible components non-equivalent:

$$T = \bigoplus_{i=1}^{\infty} T_i \tag{3}$$

where $\dim T_i = n_i < +\infty$ and T_i and T_j are non-equivalent, whenever $i \neq j$. Now if $u^*, v^* \in U^*$ the distribution of $(T(g)u^*(\xi), T(g)v^*(\xi))$ is the same for all $g \in G$. If μ is the Haar-probability measure on G, the average corresponding to (1) is

$$\int_G T(g)u^*(\xi) \cdot T(g)v^*(\xi)d\mu(g). \tag{4}$$

This should be an estimator of

$$\langle Ru^*, v^* \rangle = \int_\Omega u^*(\xi(\omega))\, v^*(\xi(\omega))dP(\omega). \tag{5}$$

If we choose an orthonormal system $e^*_{ij}, j = 1, \ldots, n_i$ for U_{T_i}, we know that

$$\langle Re^*_{ij}, e^*_{kl} \rangle = \sigma_i^2 \delta_{ik}\delta_{jl} \tag{6}$$

as e^*_{ij} are eigenvectors of R and corresponding to the same eigenvalue for $j = 1, \ldots, n_i$. By the canonical isomorphism from U to U^*, $\xi \in U$ corresponds to $\xi^* \in U^*$ with coefficients $e^*_{ij}(\xi)$ to e^*_{ij}, and

$$T(g)u^*(\xi) = \langle T(g)u^*, \xi^* \rangle_*. \tag{7}$$

But this is what we called coefficients of the representation T

$$T(g)u^*(\xi) = t_{u,\xi^*}(g) \tag{8}$$

(see 7.2).

If we by u_{ij}, and v_{ij} denote the coefficients of u^* and v^* to e^*_{ij} we can apply theorem 3.1, section 7, noting that

$$\int_G T(g)u^*(\xi)T(g)v^*(\xi)d\mu(g) = \int_G t_{u^*,\xi^*}(g)\, t_{v^*,\xi^*}(g)d\mu(g). \tag{9}$$

But this theorem says that this integral is equal to

$$\sum_{i=1}^{\infty} \frac{1}{n_i} \sum_{j=1}^{n_i} \sum_{k=1}^{n_i} u_{ij}v_{ij}e^*_{ik}(\xi)^2. \tag{10}$$

We know that $e^*_{ik}(\xi)$ are independently normally distributed with zero mean value and variance σ_i^2. The random variable

$$\left.\begin{aligned} \hat{R}(u^*, v^*) &= \int_G T(g)u^*(\xi)T(g)v^*(\xi)d\mu(g) = \\ &= \sum_{i=1}^{\infty} \sum_{j=1}^{n_i} u_{ij}v_{ij} \frac{1}{n_i} \sum_{k=1}^{n_i} e^*_{ik}(\xi)^2 \end{aligned}\right\} \tag{11}$$

therefore has mean value

$$E\hat{R}(u^*, v^*) = \sum_{i=1}^{\infty} \sum_{j=1}^{n_i} \sigma_i^2 u_{ij} v_{ij} = \langle Ru^*, v^* \rangle \tag{12}$$

and variance

$$\operatorname{Var} \hat{R}((u^*, v^*)) = \sum_{i=1}^{\infty} \sum_{j=1}^{n_i} u_{ij}^2 v_{ij}^2 \operatorname{Var}(e_{ij}^*(\xi))^2 \frac{1}{n_i} \tag{13}$$

which follows from theorem 5.2, section 5.5. But $e_{ij}^*(\xi)^2$ has a chi-square distribution with scale parameter σ_i^2 and one degree of freedom, and therefore

$$\operatorname{Var}(e_{ij}^*(\xi))^2 = 2\sigma_i^4. \tag{14}$$

Insert this into (13) and we get

$$\operatorname{Var}(\hat{R}(u^*, v^*)) = 2 \sum_{i=1}^{\infty} \sigma_i^4 \sum_{j=1}^{n_i} u_{ij}^2 v_{ij}^2 \frac{1}{n_i}, \tag{15}$$

again using the same theorem. Note that $\operatorname{Var}(\hat{R}(u^*, v^*)) > 0$ if $u_{ij} v_{ij}$ are not zero for all i, so that $\sigma_i^2 \neq 0$. We have proved:

Theorem 3.1. *If ξ is a gaussian random functional with zero mean value and covariance operator R, which is T-homogeneous, where T is a representation $T = \bigoplus_{i=1}^{\infty} T_i$ with $\dim T_i < +\infty$ and T_i and T_j non-equivalent if $i \neq j$, the estimator*

$$\hat{R}(u^*, v^*) = \int_G T(g) u^*(\xi) T(g) v^*(\xi) d\mu(g) \tag{16}$$

is unbiased

$$E\hat{R}(u^*, v^*) = \langle Ru^*, v^* \rangle_* \tag{17}$$

and has positive variance if there is an i and j, so that σ_i^2, or $u_{ij} v_{ij}$ are different from zero.

$$\operatorname{Var}(\hat{R}(u^*, v^*)) = 2 \sum_{i=1}^{\infty} \sigma_i^4 \sum_{j=1}^{n_i} u_{ij}^2 v_{ij}^2 \frac{1}{n_i} \tag{18}$$

where u_{ij}, v_{ij} are coefficients of u^* and v^* to an orthonormal system (e_{ij}^*) with e_{ij}^*, $j = 1, \ldots, n$ as an orthonormal system in U_{T_i} and $R(e_{ij}^*) = \sigma_i^2 e_{ij}^*$. Especially, if $\langle Ru^*, v^* \rangle_* \neq 0$ the variance will be positive.

9. The Model

9.1. *The Hilbert Space*

Building the model we have to choose a convenient Hilbert space of potentials. It does not matter which, it just has to be big enough and nice to work with. We shall choose the Hilbert space of section 2.2. We defined it first giving an orthonormal system in U

$$e_n^m(t) = \left\{ \begin{array}{ll} \left(\frac{R}{r_t}\right)^{n+1} \bar{R}_{nm}(\theta_t, \lambda_t) & \text{for } m \geqq 0 \\ \left(\frac{R}{r_t}\right)^{n+1} \bar{S}_{n|m|}(\theta_t, \lambda_t) & \text{for } m < 0 \end{array} \right\} \quad (1)$$

where $n = 1,2, \ldots$ and $m = -n, \ldots, n$, and then constructed the corresponding pre-Hilbert space of *finite* linear-combinations of e_n^m.

To show that it is a Hilbert space, suppose (a_{nm}) is a sequence of numbers so that $\sum a_{mn}^2 < +\infty$. We want to show that $\sum_{n=1}^{\infty} \sum_{m=-n}^{n} a_{nm} e_n^m(t)$ is convergent for any $t \in M_R$, uniformly on every compact subset of M_R, and that the limit is a harmonic function, i.e. satisfies the Laplace-equation. The convergence is easy to show along the same lines as example 4.1, section 2. The limit can be formally differentiated term by term, and the resulting series is seen to be absolutely and uniformly convergent on compact subsets, which means that the limit is the corresponding differential quotient of the series. This way we can see that it is possible to apply the Laplacian operator term by term, and as all terms are harmonic, the limit is also harmonic. Hence the completion of the pre-Hilbert space also consists of harmonic functions.

We mentioned in section 4 that the function

$$k(s,t) = \sum_{n=1}^{\infty} (2n+1) \left(\frac{R^2}{r_s r_t}\right)^{n+1} P_n(\cos\psi_{s,t}) \quad (2)$$

was the reproducing kernel for this Hilbert space, which we shall now prove. The addition formula (16), section 2.1, gives

$$\left.\begin{aligned}&\sum_{m=-n}^{n} e_n^m(s)e_n^m(t) = \\ &\left(\frac{R^2}{r_s r_t}\right)^{n+1}(2n+1)\left(2\cdot\sum_{m=1}^{n}\frac{(n-m)!}{(n+m)!}P_{nm}(\cos\theta_s)P_{nm}(\cos\theta_t)\right. \\ &\left.(\cos m\lambda_s\cos m\lambda_t + \sin m\lambda_s\sin m\lambda_t) + P_n(\cos\theta_s)P_n(\cos\theta_t)\right) \\ &= \left(\frac{R^2}{r_s r_t}\right)^{n+1}(2n+1)P_n(\cos\psi_{s,t}).\end{aligned}\right\} \tag{3}$$

So we have that

$$k(s,t) = \sum_{n=1}^{\infty}\sum_{m=-n}^{n} e_n^m(s)e_n^m(t). \tag{4}$$

To ensure that $k(\cdot,t) = k^t \in U$, we just have to show that $\sum_{n=1}^{\infty}\sum_{m=-n}^{n} e_n^m(t)^2 < +\infty$ which is quite trivial to do.

The reproducing property can also be verified immediately if $u = \sum a_{mn}e_n^m$, we have that $k^t = \sum e_n^m(t)e_n^m$ and

$$\langle u, k^t\rangle = \sum a_{mn}e_n^m(t) = u(t). \tag{5}$$

U^*, the dual space to U, is spanned by the linear functionals L_t, $t \in M_R$

$$L_t(u) = u(t)\,\forall\, u \in U, \tag{6}$$

and L_t and k^t are corresponding elements by the canonical isometric mapping J from a Hilbert space to its dual:

$$J(k^t) = L_t. \tag{7}$$

It is worth noting that although k is a positive symmetric kernel it could not be the covariance function of a stochastic process with all sample functions in U as discussed in section 6.2. It is, however, not serious at all in this particular problem.

If we by U_ε denote the corresponding Hilbert space of functions harmonic in $M_{R+\varepsilon}$, it has kernel function

$$\left.\begin{aligned}&\mathrm{k}_\varepsilon(s,t) = \sum_{n=1}^{\infty}\sum_{m=-n}^{n}{}_\varepsilon e_n^m(s)\,{}_\varepsilon e_u^u(t) = \\ &\sum_{n=1}^{\infty}\left(\frac{(R+\varepsilon)^2}{r_s r_t}\right)^{n+1}(2n+1)P_n(\cos\psi_{s,t}).\end{aligned}\right\} \tag{8}$$

It is easy to see that

$$k(s,t) = \sum_{n=1}^{\infty} \sum_{m=-n}^{n} \left(\frac{R}{R+\varepsilon}\right)^{2n+2} {}_{\varepsilon}e_n^m(s)\,{}_{\varepsilon}e_n^m(t). \tag{9}$$

And because

$$\sum_{n=1}^{\infty} \left(\frac{R}{R+\varepsilon}\right)^{2n+2} (2n+1) < +\infty \; \forall \varepsilon > 0 \tag{10}$$

it follows from section 6.2 that there is a gaussian process with sample functions all in U_ε. But then the sample functions are harmonic in $M_{R+\varepsilon}$ for any $\varepsilon > 0$ and therefore also harmonic in M_R.

This means that the main difficulties and problems in connexion with choosing a Hilbert space lie in the *boundary conditions* of the sample functions and not in the harmonicity. As we have already chosen to place a Bjerhammar sphere *in the interior of the earth,* these problems have no physical meaning, as the physically existing potential necessarily satisfies boundary conditions on the boundary of the earth and not on some convenient mathematical concept as the Bjerhammar sphere.

9.2. *The Group Representation*

We have a natural transformation group suitable for a reduction of the problem. Actually, if we think of the earth as a sphere we cannot see what is up and what is down (as the potential of rotation is out of the problem). This can be expressed by saying that a rotation of space around the origin must not change anything in the problem.

The group of rotations of three-dimensional space around the origin or the orthogonal group, together with its representations, has been studied by Weyl and others as one of the classical groups. In the rest of this part I shall refer to *Gelfand and Shapiro* (1956), whenever a result is stated without proof or without details.

To get a brief look on the topological nature of the rotation group, you can think of a rotation as a vector which has its direction along the axis of rotation and whose length $[0, \pi]$ is the angle of rotation, the rotation group being mapped into a sphere with radius π, so that opposite points correspond to the same rotation. This is the topological structure of the group: a sphere with opposite points identified. The group is *compact*, but not commutative.

Any rotation can be built up from rotations about the x-axis and z-axis. If we by $X(\theta)$, $0 \leqq \theta \leqq \pi$ denote the rotation about the x-axis through

the angle θ, and by $Z(\varphi)$, $0 \leqq \varphi \leqq 2\pi$ denote the rotation about the z-axis through the angle φ, any rotation is of the form

$$g(\varphi_2,\theta,\varphi_1) = Z(\varphi_2)X(\theta)Z(\varphi_1). \tag{1}$$

$(\varphi_2, \theta, \varphi_1)$ are called the *Eulerian angles* of the rotation. Different triplets $(\varphi_2,\theta,\varphi_1)$ correspond to different rotations except for $\theta = 0$ and $\theta = \pi$.

These Eulerian angles give a convenient expression of the Haar probability measure on the rotation group. If f is a continuous function on the group, the integration formula becomes

$$\int_G f(g)d\mu(g) = \frac{1}{8\pi^2}\int_0^{2\pi}\int_0^{\pi}\int_0^{2\pi} f(\varphi_2, \theta, \varphi_1)\sin\theta\, d\varphi_1 d\theta d\varphi_2. \tag{2}$$

As $\sin\theta d\theta d\varphi_1$ is the surface element of the unit sphere at the point with spherical coordinates $(1, \theta, \varphi_1)$, this integral is an integral with respect to the product measure of the surface measure on a sphere and arc-measure on a circle.

The matrices corresponding to $Z(\varphi)$ and $X(\theta)$ become:

$$Z(\varphi): \begin{Bmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{Bmatrix} \tag{3}$$

and

$$X(\theta): \begin{Bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{Bmatrix} \tag{4}$$

If L_t, $t \in M_R$ is the element of U^* determined by

$$L_t(u) = u(t)\,\forall\, u \in U, \tag{5}$$

we shall to any $g \in G$ (G is the rotation group) define $T(g)$: $U^* \to U^*$ by

$$T(g)(L_t) = L_{g(t)} \tag{6}$$

and $\tilde{T}(g)$: $U \to U$, by

$$\tilde{T}(g)(k^t) = k^{g(t)}, \tag{7}$$

where $g(t)$ is the picture of t by the rotation $g \in G$. This means that if $u \in U$

$$\tilde{T}(g)u(t) = \langle \tilde{T}(g)u, k^t \rangle = \langle u, \tilde{T}(g)^{-1} k^t \rangle = u(g^{-1}(t)). \tag{8}$$

$\tilde{T}(g)$ is obviously a transformation of U which is linear and bounded, and it is inner-product preserving (unitary) because of the rotation invariance of the kernel of U. $T(g)$ can also be extended to U by linearity and continuity. T and $\tilde{T}$ are obviously homomorphisms of G.

To show the continuity of T it is enough to show it at the identity of G, because if

$$\| \tilde{T}(g)u - u \| \underset{g \to e}{\to} 0 \quad \forall u \in U \tag{9}$$

then

$$\| \tilde{T}(g)u - \tilde{T}(g_0)u \| = \| \tilde{T}(g_0^{-1}g)u - u \| \underset{g \to g_0}{\to} 0 \tag{10}$$

as $g_0^{-1}g \underset{g \to g_0}{\to} e$ because of the continuity of the group operations.
It is also enough to show

$$\| \tilde{T}(g)k^t - k^t \| \underset{g \to e}{\to} 0 \quad \forall t \in M_R \tag{11}$$

as any u of U can be approximated by $\sum_{i=1}^{n} a_i k^{t_i}$, and

$$\left.\begin{aligned} \| \tilde{T}(g)u - u \| \leqq \| \tilde{T}(g)u - \tilde{T}(g)u_n \| + \| \tilde{T}(g)u_n - u_n \| + \| u_n - u \| \\ = 2 \cdot \| u - u_n \| + \| \tilde{T}(g)u_n - u_n \|. \end{aligned}\right\} \tag{12}$$

We have

$$\| \tilde{T}(g)k^t - k^t \|^2 = 2(k(t,t) - k(t, g^{-1}(t))) \tag{13}$$

and the right side of (13) tends to zero when g tends to e because of the continuity of the kernel. The continuity of T follows from the isomorphism $J: U \to U^*$.

If we define

$$V_n = \text{span} \{ e_n^m, m = -n, \ldots, n \}, \tag{14}$$

V_n is *invariant* under $\tilde{T}$. This can be seen by the following argument: if $u \in V_n$, i.e. u is a spherical harmonic of degree n, u has the series expansion:

$$u(t) = \left(\frac{R}{r_t} \right)^{n+1} \left(\sum_{m=0}^{n} a_{nm} \bar{R}_{nm}(\theta_t \lambda_t) + \sum_{m=1}^{n} b_{nm} \bar{S}_{nm}(\theta_t, \lambda_t) \right) \tag{15}$$

and

$$\left.\begin{aligned}\tilde{T}(g^{-1})u(t) = u(g(t)) = \\ \left(\frac{R}{r_{g(t)}}\right)^{n+1}\left(\sum_{m=0}^{n} a'_{nm}\bar{R}_{nm}(\theta_{g(t)},\lambda_{g(t)}) + \sum_{m=1}^{n} b'_{nm}\bar{S}_{nm}(\theta_{g(t)},\lambda_{g(t)})\right)\end{aligned}\right\} \quad (16)$$

but because $r_{g(t)} = r_t$, $\tilde{T}(g^{-1})u$ also has an expression containing r_t only as $\frac{1}{r_t^{n+1}}$ and must therefore also be an element of V_n.

We have mentioned earlier (section 2.2) that the functions $\bar{R}_{nm}$ and $\bar{S}_{nm}$ also played a role as orthonormal systems in the Hilbert space H_2 of square-integrable functions on the unit sphere. One of the classical representations of the rotation group, studied by Weyl and also treated in the earlier cited work by Gelfand and Shapiro, is defined by setting $T^*(g)f(t) = f(g^{-1}(t))$ for $f \in H_2$. It is here known that $\bar{R}_{nm}$, $n = 0, \ldots,$ n and $\bar{S}_{nm}$, $m = 1, \ldots, n$ together constitute a orthonormal system for a subspace $L_n \subset H_2$, which contains *no proper invariant subspace,* i.e. the subrepresentation $T^* \mid L_n$ is *irreducible.* As $\tilde{T} \mid V_n$ acts on the functions $\bar{R}_{nm}$ and $\bar{S}_{nm}$ exactly as $T^* \mid L_n$ it follows that $\tilde{T} \mid V_n$ is *irreducible.*

So we have the decomposition

$$\tilde{T} = \bigoplus_{n=1}^{\infty} \tilde{T} \mid V_n \quad (17)$$

and this decomposition is unique as $\dim V_n = 2n+1 \neq \dim V_m$ if $m \neq n$ and therefore $\tilde{T} \mid V_n$ and $\tilde{T} \mid V_m$ are non-equivalent.

By the isomorphism between U and U^* where $V_n \leftrightarrow V_n^*$ we also get

$$T = \bigoplus_{n=1}^{\infty} T \mid V_n^*, \quad (18)$$

and the same remarks hold.

9.3. *The Gaussian Measure*

We shall now define a gaussian measure on the Hilbert space U. First we have to get a mean value. If we substract a suitable normal potential from the potential of the earth and only consider the anomalous potential, we can very well assume the mean value to be zero.

We can say a lot about the covariance structure of the measure if we demand T-homogeneity. Demanding this kind of homogeneity is very

natural, as it follows from section 8.1 that this is equivalent to demanding $T(g)$ to be measure-preserving for any g, i.e. *sets of potentials which can be carried into each other by a rotation of space have the same probability.* Maybe it is a little too much as the earth is not a sphere, but this kind of homogeneity is closely related to the concept of the Bjerhammar sphere, and if we should just have invariance with respect to rotation about the z-axis, we should also use a Bjerhammar ellipsoid and everything would get unneccessarily complicated.

As $T \mid V_n^*$ are subrepresentations of T, elements of V_n^* are *eigenvectors* for the covariance operator R corresponding to the eigenvalue $\sigma_n{}^2$, and

$$\langle R e_n^{m*}, e_k^{r*} \rangle_* = \sigma_n^2 \delta_{kn} \delta_{rm} \tag{1}$$

when e_n^{m*} is the linear functional mapping a potential u into its coordinate to the basis vector e_n^m:

$$e_n^{m*}(u) = \langle u, e_n^m \rangle. \tag{2}$$

(The natural isomorphism has been used again). But as

$$\mathrm{Var}(e_n^{m*}(\xi)) = \langle R\, e_n^{m*}, e_n^{m*} \rangle = \sigma_n^2, \tag{3}$$

σ_n^2 can be interpreted as the variance of the coefficients of the potential to the normalized spherical harmonics of degree n. Hence the name "degree variances".

Another way to think of the model is that it is put together by denumerably many variance components V_n with variance σ_n^2 and dim $V_n = 2n + 1$.

In section 2.3 it was mentioned that the coefficients of the potential to e_n for $n = 1$, $n = 2$, and $m = 1, -1$ were *known to be zero.* Therefore we must have $\sigma_1^2 = 0$. If we want T-homogeneity we must also have $\sigma_2^2 = 0$ and assume the coefficients to e_2^n known for *all* $m = -2, \ldots, 2$. Therefore *we must include the best known value of these into the normal potential.* This has not been done in earlier attempts to give a statistical treatment of the problem, which in my opinion is wrong. It is not easy to estimate the practical importance of this "change" in the model, but why make unnecessary theoretical inconsistencies?

In order that the covariance operator should have finite trace, we must demand

$$\sum_{n=1}^{\infty} \sum_{m=-n}^{n} \sigma_n^2 = \sum_{n=3}^{\infty} \sigma_n^2 (2n+1) < +\infty. \tag{4}$$

This gaussian measure induces, as shown in section 6.2 a gaussian process ξ^t, $t \in M_R$, where ξ^t denotes the value of the potential at the point t with covariance function:

$$\left.\begin{aligned} E\xi^s\xi^t &= \sum_{n=1}^{\infty} \sigma_n^2 \sum_{m=-n}^{n} e_n^m(s) e_n^m(t) \\ &= \sum_{n=3}^{\infty} \sigma_n^2 (2n+1) \left(\frac{R^2}{r_s r_t}\right)^{n+1} P_n(\cos\psi_{s,t}). \end{aligned}\right\} \quad (5)$$

The degree variances σ_n^2 are *unknown* and have to be *estimated.* We never have more than a finite number of observations, and in my opinion it *makes no sense at all* to draw statistical inference on an *infinite number of unknown parameters from a finite number of observations.*

This means that we somehow have to reduce the problem further than we have done at the present stage. The best way to do so is to get a *physical interpretation* of σ_n^2 and deduce something about its structure. In fact, this is not easy, and I am not able to do it.

Another approach is to find an expression for σ_n^2 containing *few* parameters which *fits data satisfactorily*; this will be done in the last sections of the present work.

9.4. *The Linear Functionals*

The linear functionals which shall be considered here are

1) gravity anomalies
2) deflections of the vertical
3) height anomalies.

The gravity anomalies as linear functionals were discussed in section 2.4. It follows that the covariance between gravity anomaly in a point s and gravity anomaly in t is given as:

$$\left.\begin{aligned} E(\Delta g(s)\Delta g(t)) &= \sum_{n=1}^{\infty} \sum_{m=-n}^{n} b_{nm}(s) b_{nm}(t) \sigma_n^2 \\ &= \sum_{n=3}^{\infty} (2n+1)(n-1)^2 \sigma_n^2 \frac{R^{2n+2}}{(r_s r_t)^{n+2}} P_n(\cos\psi_{s,t}). \end{aligned}\right\} \quad (1)$$

The deflections of the vertical $\xi(t_0)$ and $\eta(t_0)$ at the point t_0 resulting from an anomalous potential u is given by:

$$\xi(t_0) = \frac{-1}{r_{t_0}\gamma_{t_0}} \frac{\partial u(t)}{\partial \theta_t}\bigg|_{t=t_0} \tag{2}$$

and

$$\eta(t_0) = \frac{-1}{\cos\theta_{t_0} \cdot r_{t_0} \cdot \gamma_{t_0}} \frac{\partial u(t)}{\partial \lambda_t}\bigg|_{t=t_0} \tag{3}$$

where γ_{t_0} is normal gravity at t_0 (Heiskanen & Moritz, 1967). Defining $c_{nm}(t_0)$ and $d_{nm}(t_0)$ as

$$c_{nm}(t_0) = \frac{-1}{r_{t_0}\gamma_{t_0}} \frac{\partial e_n^m(t)}{\partial \theta_t}\bigg|_{t=t_0} \tag{4}$$

and

$$d_{nm}(t_0) = \frac{-1}{r_{t_0}\gamma_{t_0}\cos\theta_{t_0}} \frac{\partial e_n^m(t)}{\partial \lambda_t}\bigg|_{t=t_0}, \tag{5}$$

we have, if $u = \sum_{n=1}^{\infty} \sum_{m=-n}^{n} a_{nm} e_n^m$,

$$\xi(t_0) = \sum_{n=1}^{\infty} \sum_{m=-n}^{n} a_{nm} c_{nm}(t_0) \tag{6}$$

and

$$\eta(t_0) = \sum_{n=1}^{\infty} \sum_{m=-n.}^{n} a_{nm} d_{nm}(t_0). \tag{7}$$

Noting that

$$\frac{\partial e_n^m(t)}{\partial \theta_t} = \left(\frac{R}{r_t}\right)^{n+1} \cdot \cos m\lambda_t \frac{\partial P_{nm}(\cos\theta_t)}{\partial \theta_t} \cdot k_{nm} \tag{8}$$

if $m \geqq 0$ and

$$\frac{\partial e_n^m(t)}{\partial \theta_t} = \left(\frac{R}{r_t}\right)^{n+1} \sin m\lambda_t \frac{\partial P_{n|m|}(\cos\theta_t)}{\partial \theta_t} \cdot k_{nm} \tag{9}$$

if m < 0, where $k_{nm} = \sqrt{2(2n+1)\frac{(n-m)!}{(n+m)!}}$, it is not difficult to see that

$$\left|\frac{\partial e_n^m(t)}{\partial \theta_t}\right|^2 \leqq Q(n) \cdot e_n^{m2}(t) \tag{10}$$

where Q is a polynomial in n. Similarly

$$\frac{\partial e_n^m(t)}{\partial \lambda_t} = -e_n^{-m}(t) \cdot m. \tag{11}$$

This easily gives

$$\sum_{n=1}^{\infty} \sum_{m=-n}^{n} c_{nm}^2(t) < +\infty \tag{12}$$

and

$$\sum_{n=1}^{\infty} \sum_{m=-n}^{n} d_{nm}^2(t) < +\infty. \tag{13}$$

If we therefore define $c_{t_0}, d_{t_0} \in U$ by

$$\langle c_{t_0}, e_n^m \rangle = c_{nm}(t_0) \tag{14}$$

and

$$\langle d_{t_0}, e_n^m \rangle = d_{mn}(t_0), \tag{15}$$

we get

$$\xi(t_0) = \langle u, c_{t_0} \rangle \tag{16}$$

and

$$\eta(t_0) = \langle u, d_{t_0} \rangle \tag{17}$$

and therefore that the linear functionals, mapping u into its corresponding deflections of the vertical, are elements of U.

The height anomalies are given by

$$\xi(t_0) = \frac{u(t_0)}{\gamma_{t_0}} \tag{18}$$

and are elements of U^* because U has a reproducing kernel and

$$\xi(t_0) = \langle u, \frac{k^{t_0}}{\gamma_{t_0}} \rangle. \tag{19}$$

From these relations it is very easy to get expressions of covariance between different functionals of the linear functionals of the potential.

If we note however that the coefficients $u_{nm}(t_0)$ to e_n^m of the element in U corresponding to $u_{t_0}^* \in U^*$ were found as

$$u_{nm}(t_0) = u_{t_0}^*(e_n^m) \tag{20}$$

we remark that

$$\langle Ru_{s_0}^*, v_{t_0}^* \rangle = u_{s_0}^*(v_{t_0}^*(r^{\cdot})) \,. \tag{21}$$

Here r is the "potential" covariance function given by (5) and (21) means that we first apply $v_{t_0}^*$ on r^t for any fixed $t \in M_R$ and then apply $u_{s_0}^*$ on $v_{t_0}^*(r^t)$ considered as a function on t. This has certainly importance for the computation of covariances.

10. Estimating the Covariance Function

10.1. *The Observations*

The observations available consist mainly of about 1.000.000 measurements of gravity anomalies and 100.000 deflections of the vertical. The observations are not homogeneously scattered on the surface, but are concentrated in densely populated areas. Deflections of the vertical are missing on the oceans. In the Kaula papers mentioned earlier, there is a detailed description of the observations.

Because of the enormous amount of gravity anomaly measurements it is natural to use these to get an empirical covariance function for the gravity anomalies and to determine the "degree variances" σ_n^2 from this one. But of course the practical problems in dealing with one million observations spread over the earth in an inconvenient way are not negligible. All these problems have been discussed intensively by Kaula and I shall not here do this again.

Besides, I do not find it worth while, which is a consequence of the non-ergodicity, as the next section will show.

10.2. *Uncertainty of the Empirical Covariance Function*

In *Kaula* (1959), the empirical covariance function is computed in the following way. First, the gravity anomalies are changed to mean square anomalies in squares of about 300×300 n.m. In squares without measurements, the anomalies are predicted from a very simple model of the Markov-type. All products of these values are computed and classified according to spherical distance. The average of products with approximately the same spherical distance is then used to estimate the covariance between gravity anomalies which are apart by this distance:

$$\hat{C}(\theta) = \frac{1}{N} \sum_{\psi s, t = \theta} \Delta g(s) \cdot \Delta g(t) \tag{1}$$

where N is the number of terms in the sum. Thus $\hat{C}(\theta)$ should be a reasonable estimate of

$$C(\theta) = \sum_{n=1}^{\infty} A_n P_n(\cos\theta) \tag{2}$$

where A_n are unknown parameters. To get a guess on the variance of $\hat{C}(\theta)$, Kaula by a heuristic argument comes to the result that it is possible to treat the products in the sum as if they were independent, as their dependence cannot be so very great at all. This gives that

$$\mathrm{Var}(\hat{C}(\theta)) \simeq \frac{\sigma^2}{N} \tag{3}$$

where

$$\sigma^2 = \sum_n A_n P_n(\cos\theta)_{\theta=0} = \mathrm{Var}(\Delta g) = C(0). \tag{4}$$

This means that if the sum contains, say, about 3000 products, this variance is negligible, and we can say that we actually have the right value.

But (3) is not true! Suppose we knew gravity anomaly continuously all over the earth. As estimator for the covariance function one should suggest

$$\hat{C}(\theta) = \frac{1}{8\pi^2 R^3} \int_{s \in k} \int_{t \in k_s(\theta)} \Delta g(s) \cdot \Delta g(t) d\sigma_{\theta,s}(t) d\sigma(s), \tag{5}$$

where k is a sphere with centre at the origin and radius R, and $k_s(\theta)$ is a circle on the sphere with "spherical radius" θ. $\sigma_{\theta,s}(t)$ is arc-measure on the circle $k_s(\theta)$, and $\sigma(s)$ is surface-measure on the sphere k.
Define

$$\tilde{C}(\psi) = \int_{x \in G} \Delta g(x(s)) \Delta g(x(t)) d\mu_H(x), \tag{6}$$

where $\psi_{s,t} = \psi$ and μ_H is Haar-probability measure on the rotation group G. We shall show that $\tilde{C}(\theta) = \hat{C}(\theta) \forall \theta$, and we shall then use the results from section 8.3 about the variance of $\tilde{C}(\theta)$.

If we put $f(x) = \Delta g(x(s_0)) \cdot \Delta g(x(t_0))$, where s_0 has spherical coordinates $(R,0,0)$ and t_0 has spherical coordinates $(R,\psi,0)$, we have from (2) section 9.2

$$\int_G f(x) d\mu_H(x) = \frac{1}{8\pi^2} \int_0^{2\pi} \int_0^{\pi} \int_0^{2\pi} f(\varphi_2, \theta, \varphi_1) \sin\theta \, d\varphi_1 d\theta \, d\varphi_2 \tag{7}$$

where $f(\varphi_2,\theta,\varphi_1)$ is $f(x)$ where x has eulerian angles $(\varphi_2,\theta,\varphi_1)$. Because of the invariance of μ_H

$$\tilde{C}(\psi) = \int_G (f(x))\, d\mu_H(x)) \tag{8}$$

as we can always bring s and t into s_0 and t_0 by means of a rotation. Using the formulas (3) and (4) from section 9.2 it is possible to realize that

$$f(\varphi_2,\theta,\varphi_1) = \Delta g\left(R,\theta,\frac{3\pi}{2}+\varphi_2\right)\cdot \Delta g(r_{\varphi_1}(y)) \tag{9}$$

where

$$r_{\varphi_1} = Z(\varphi_2)X(\theta)\, Z(\varphi_1)\, X(-\theta)\, Z(-\varphi_2) \tag{10}$$

and

$$y = Z(\varphi_2)X(\theta)\, (R,\psi,0) \tag{11}$$

(X and Z are the elementary rotations defined by (3) and (4) in section 9.2).

But then

$$\left.\begin{aligned} &\frac{1}{8\pi^2}\int_0^{2\pi}\int_0^{\pi}\int_0^{2\pi} f(\varphi_2,\theta,\varphi_1)\sin\theta\, d\varphi_1\, d\theta\, d\varphi_2 \\ &= \frac{1}{8\pi^2}\int_0^{2\pi}\int_0^{\pi} \Delta g\left(R,\theta,\frac{3\pi}{2}+\varphi_2\right)\int_0^{2\pi} \Delta g(r_{\varphi_1}(y))\, d\varphi_1 \sin\theta\, d\theta\, d\varphi_2 \end{aligned}\right\} \tag{12}$$

r_{φ_1} carries the point $\left(R,\theta,\frac{3\pi}{2}+\varphi_2\right)$ into itself:

$$\left.\begin{aligned} r_{\varphi_1}\left(R,\theta,\frac{3\pi}{2}+\varphi_2\right) &= Z(\varphi_2)X(\theta)Z(\varphi_1)X(-\theta)\left(R,\theta,\frac{3\pi}{2}\right) \\ &= Z(\varphi_2)X(\theta)Z(\varphi_1)(R,0,0) \\ &= \left(R,\theta,\frac{3\pi}{2}+\varphi_2\right). \end{aligned}\right\} \tag{13}$$

Therefore r_{φ_1} must be a rotation about the axis through the origin and $\left(R,\theta,\frac{3\pi}{2}+\varphi_2\right)$.

By rotating the point $\left(R,\theta+\frac{\pi}{2},\frac{3\pi}{2}+\varphi_2\right)$ one can see by some calculation that r_{φ_1} is a rotation of angle φ_1. But then $r_{\varphi_1}(y)$ must move along

a circle on the sphere with centre $\left(R,\theta,\frac{3\pi}{2}+\varphi_2\right)$ and spherical radius ψ with constant velocity. But $\left(R,\theta,\frac{3\pi}{2}+\varphi_2\right)$ will pass through all points on the sphere and $\sin\theta d\theta d\varphi_2$ is proportional to the volume of the surface element on the sphere. Hence

$$\tilde{C}(\psi) = \int_G f(x)\,d\mu_H(x) = C_1 \int_{s\in k}\ \int_{t\in k_s(\psi)} \Delta g(s)\Delta g(t)\,d\sigma_{\psi,s}(t)\,d\sigma(s) \quad (14)$$

and therefore

$$\tilde{C}(\theta) = \hat{C}(\theta). \quad (15)$$

And we get from section 8.3

$$\mathrm{Var}(\hat{C}(\theta)) = 2\sum_{n=1}^{\infty}\frac{\sigma_n^4}{2n+1}\left(\sum_{m=-n}^{n}(b_{nm}(r,0,0))^2(b_{nm}(r,\theta,0))^2\right), \quad (16)$$

where $b_{nm}(r,\theta,\lambda) = \frac{n-1}{r}e_n^m(r,\theta,\lambda)$, see section 2.4. But as $P_{nm}(t) = (1-t^2)^{\frac{m}{2}}\frac{d^mP_n(t)}{dt^n}$ we get

$$P_{nm}(1) = \left\{\begin{array}{ll} 0 & \text{if } m \neq 0 \\ 1 & \text{if } m = 0. \end{array}\right\} \quad (17)$$

Using (16) we get

$$\mathrm{Var}(\hat{C}(\theta)) = 2\sum_{n=1}^{\infty}\frac{\sigma_n^4}{2n+1}(b_{n0}(r,0,0))^2(b_{n0}(r,\theta,0))^2 \quad (18)$$

and as

$$b_{n0}(r,\theta,\lambda) = \frac{n-1}{r}\left(\frac{R}{r^2}\right)^{n+1}\sqrt{2n+1}P_n(\cos\theta), \quad (19)$$

$$\mathrm{Var}(C(\theta)) = 2\cdot\left(\frac{R}{r^2}\right)^4\sum_{n=1}^{\infty}(n-1)^4(2n+1)\sigma_n^4\left(\frac{R}{r}\right)^{4n}(P_n(\cos\theta))^2. \quad (20)$$

Except in trivial cases this variance is fairly large and is not to be neglected. The sum (1) converges to the integral (5) when the number of points becomes infinite and the maximum distance between two points tends to zero (a Riemann procedure). It therefore follows that the variance of the sum *does not tend to zero* as (3) would suggest.

This means that, even if we knew gravity all over the earth, we would not be able to find the true value of the covariance function. But if we knew gravity, we could find the true potential. Somehow it seems that it is much more difficult to find the covariance function than to find the potential. This fact gives the impression that somehow the problem is not suited for a statistical treatment. However, if the alternative is to use an arbitrary kernel for collocation it cannot be a bad idea to use a kernel that in a way *reflects the observed behaviour of the actual potential of the earth*, which will be the case when using some estimated covariance function.

But it seems to be less important to get very exact values of the empirical covariance function and the degree variances. As in section 9.3, the conclusion must be that it is desirable to get a covariance function which is easy to handle and which is not too far away from the empirical one.

10.3. *The Empirical Covariance Function*

Data used to compute the empirical covariance function are those prepared by *Kaula* (1966), and reproduced in *Gaposhkin and Lambeck* (1970), consisting of 300×300 *n.m.* mean anomalies. The normal potential used corresponds to a reference ellipsoid with $f = 1/298.255$ (see Gaposhkin and Lambeck, 1970, p. 49) with the modification that the values of the coefficients of the potential to spherical harmonics up to the degree $n = 2$ were included. These coefficients were taken from table 14 in the same paper. The empirical covariance function was computed and is plotted in figure 2. One may notice that, for spherical distances greater than 30°, the gravity anomalies seem to be approximately independent.

In figure 3 the empirical covariance function of the gravity anomalies refer to a normal potential without the above modification. In fact the two covariance functions are almost equal, so it seems to have no practical importance to include these extra coefficients.

This covariance function includes very few observations of the short-range values. To get a better idea of the short-range part of the covariance function (which in fact is the most important) another sample of about 16.000 $1° \times 1°$ mean gravity anomalies has been used. The corresponding covariance function has been taken directly from *Kaula*, (1966) and is shown in table I (fig. 4).

It is remarkable that the variance in the short-range sample is about 711 mGal^2, and the variance in the long-range sample about 270 mGal^2.

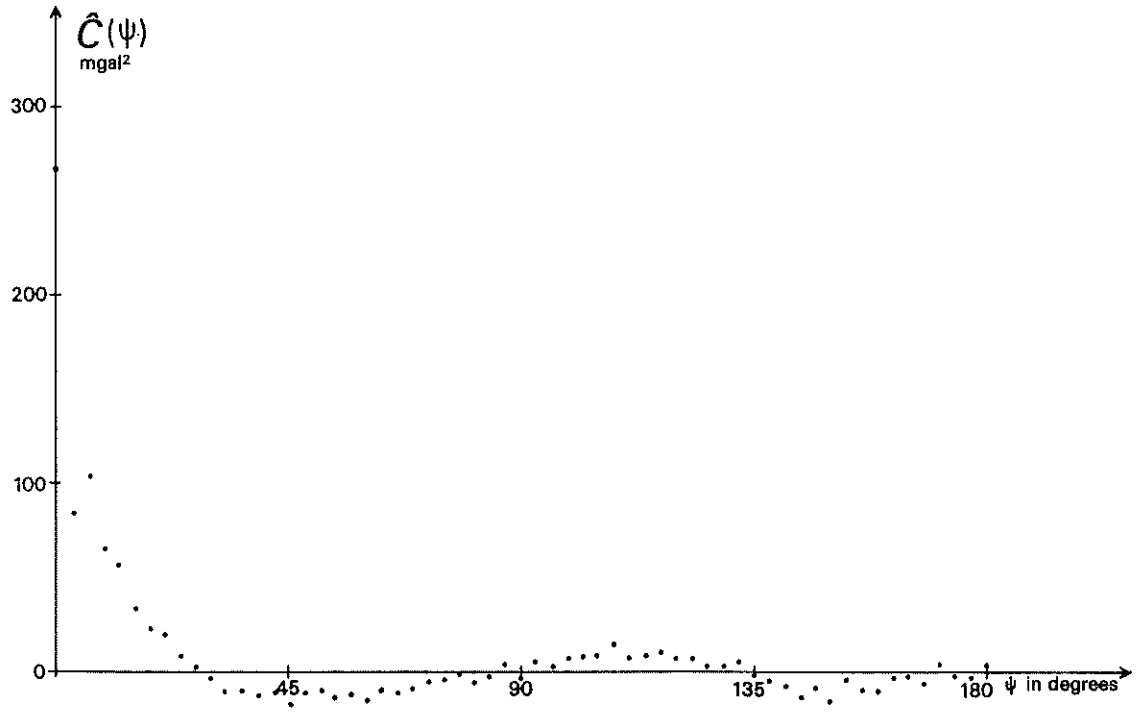

Fig. 2. Empirical covariance function of gravity anomalies referring to normal potential with terms of degree $\leqq 2$ included.

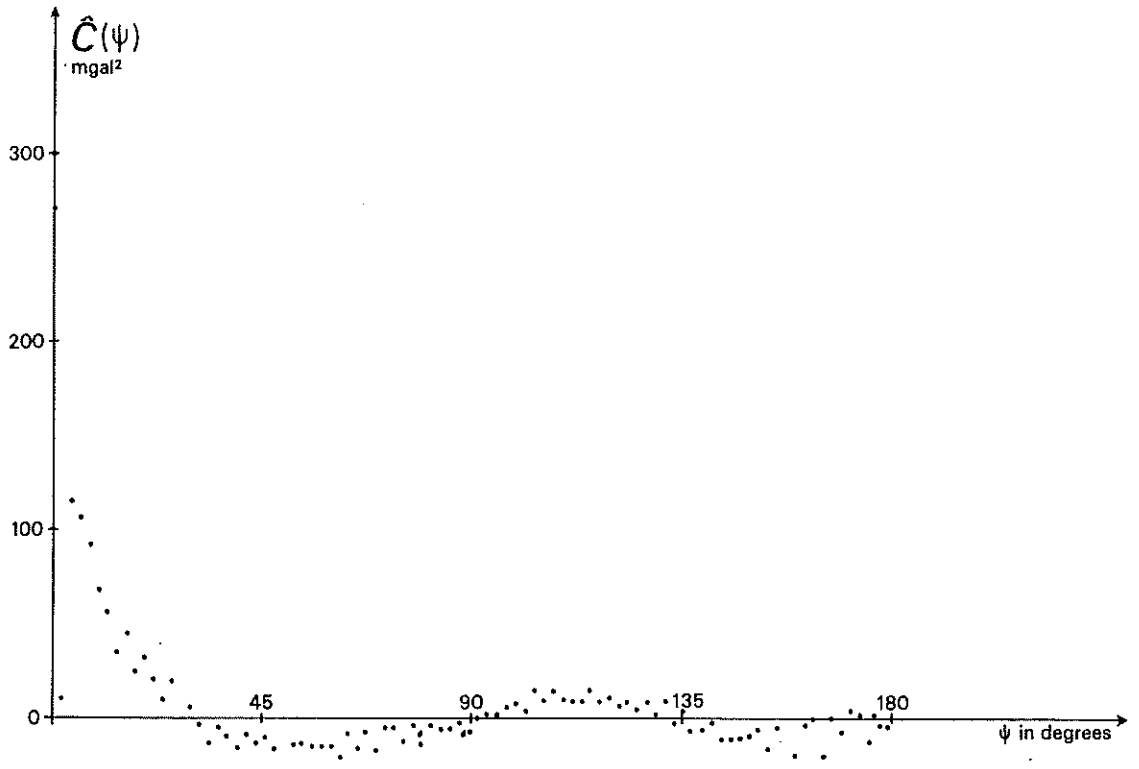

Fig. 3. Empirical covariance function of gravity anomalies referring to normal potential without modification.

Table I. Kaula's short-range empirical covariance function of gravity anomalies

argument °	cov.fct. mGal²	number of obs.
0.00	711	16331
0.92	421	6803
1.32	272	35225
2.35	196	43674
3.35	162	33864
4.34	97	19414

This, however, is easy to explain, as the global sample consists of mean anomalies in squares of 300×300 *n.m.*, which considerably reduces the variance; furthermore some of the values in the long-range sample are predicted from a very simple model.

Of course this is very unpleasant because one does not have the slightest idea of what happens to the covariance function from the long-range sample, and it seems to me impossible to have confidence in it. In my opinion it is more adequate to use only correct measurements of gravity anomaly and then ignore the fact that in large areas no measurements are present. We shall, however, not panic, because what we want from the empirical covariance function is just something to look at.

In *Gaposhkin and Lambeck*, (1970), a set of values of the "degree variances" corresponding to the gravity covariance function is listed. As the tendency of these are slightly decreasing towards a constant, one could hope that they could be described as $A \cdot \dfrac{n-1}{n-2}$ for $n \geqq 3$ and 0 otherwise, with good approximation. This corresponds to a potential covariance function

$$r(s,t) = \sum_{n=3}^{\infty} \frac{A}{(n-1)(n-2)} \left(\frac{R^2}{r_s r_t}\right)^{n+1} P_n(\cos\psi_{st}). \tag{1}$$

This can be obtained as a closed expression

$$\left.\begin{aligned} r(s,t) = A\left(\frac{R^2}{r_s r_t}\right)^3 \left(P_2(\cos\psi_{st})\left(1 + \ln\frac{2}{a}\right) + \frac{\sin^2\psi_{st}}{4}\right) \\ - \left(\frac{R^2}{r_s r_t}\right)^2 \cos\psi_{st} \ln\frac{2}{a} + \left(\frac{R^2}{r_s r_t}\right)\frac{b}{2}\left(3\left(\frac{R^2}{r_s r_t}\right)\cos\psi_{st} - 1\right) \end{aligned}\right\} \tag{2}$$

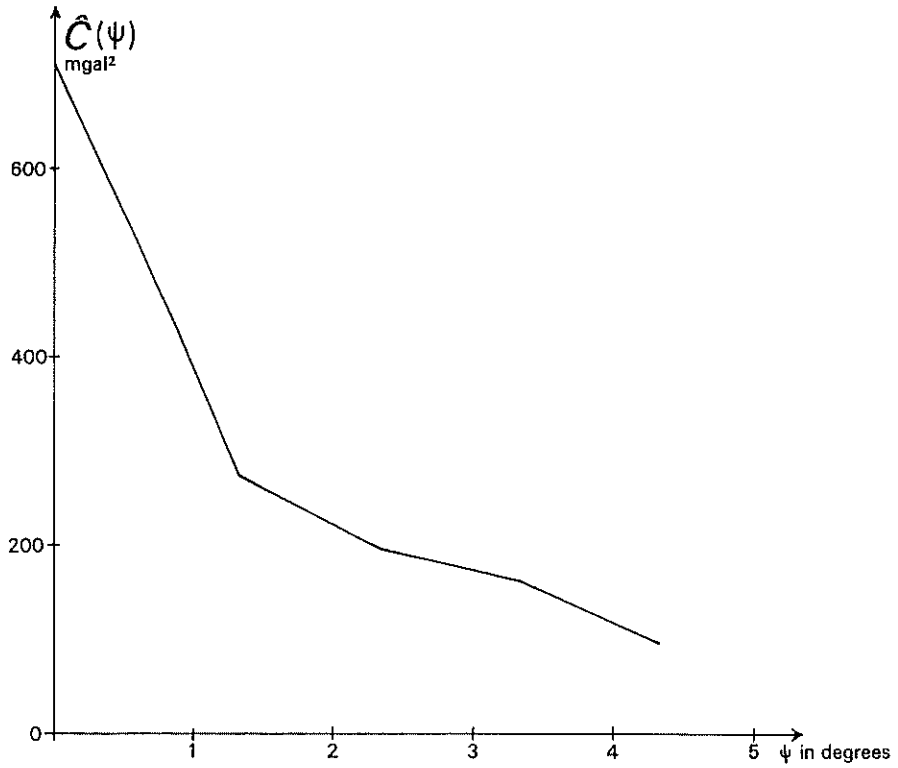

Fig. 4. Short-range empirical covariance function (Table I).

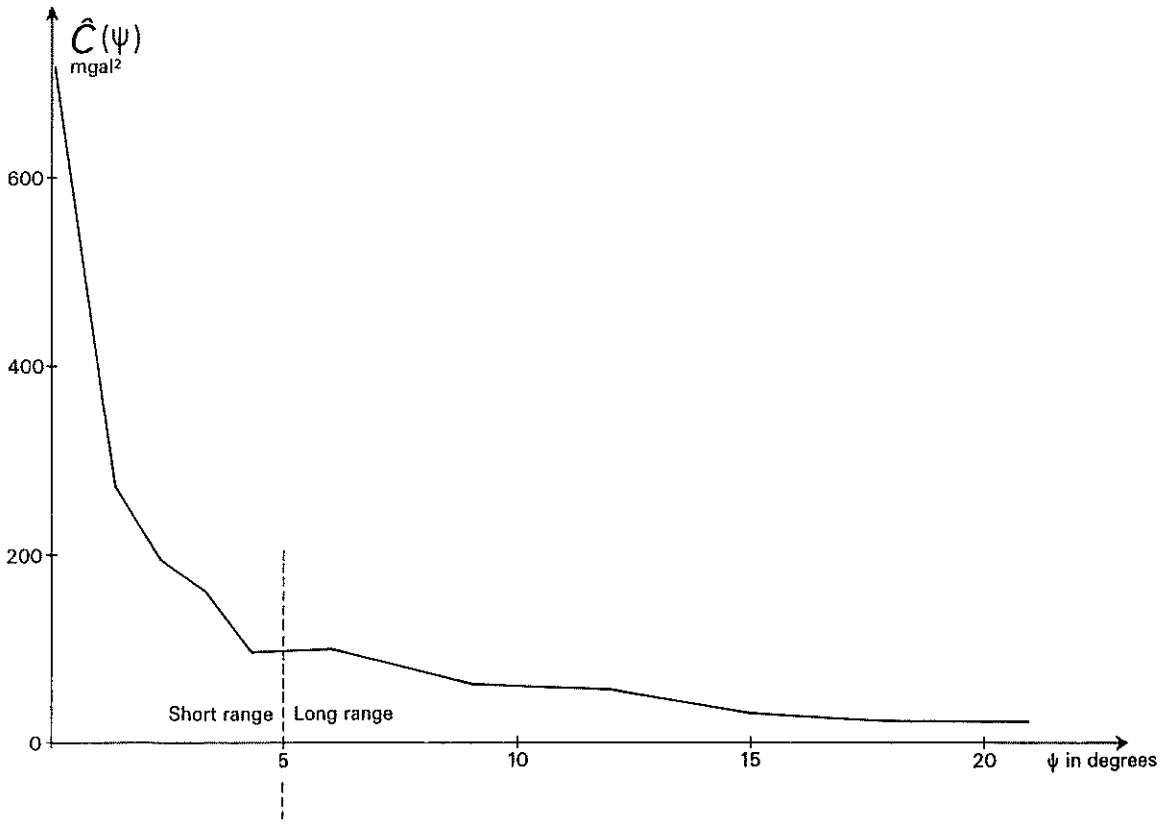

Fig. 5. Mixed empirical covariance function of gravity anomalies.

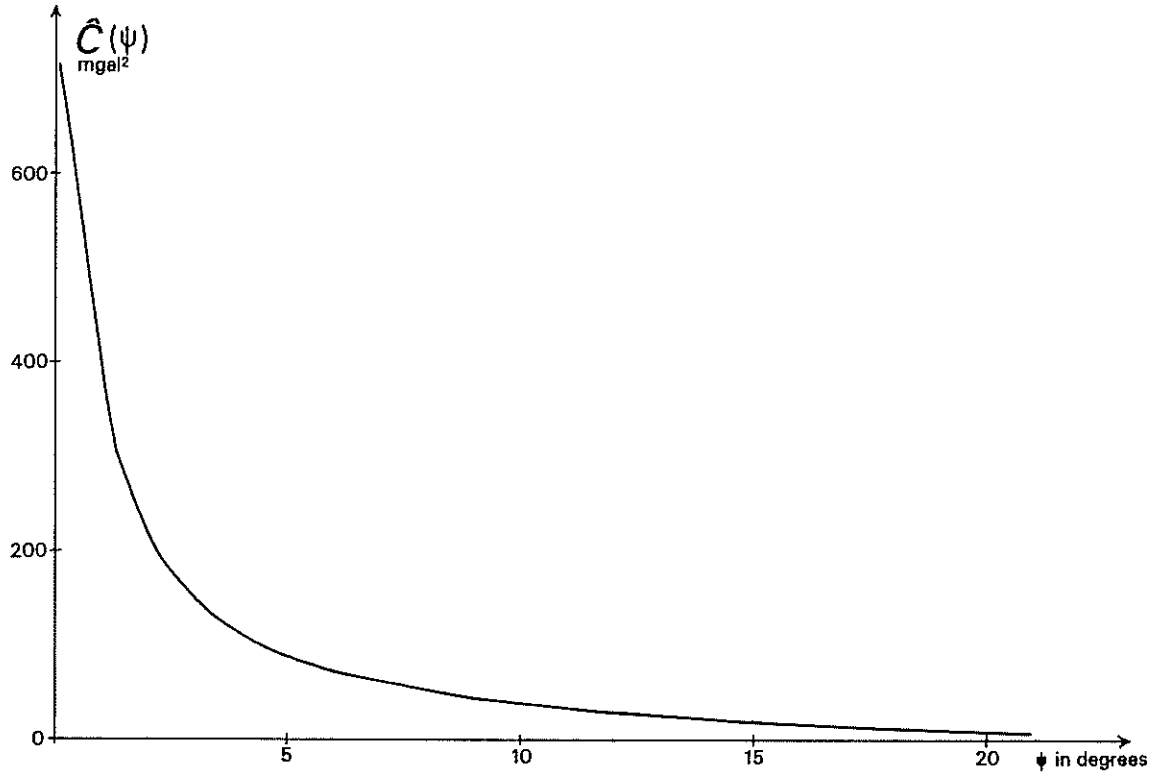

Fig. 6. Parametric covariance function of gravity anomalies.
$A = 7.84888 \quad R/r = 0.994.$

where

$$\left.\begin{aligned} a &= 1 - L - \left(\frac{R^2}{r_s r_t}\right)\cos\psi_{st} \\ b &= 1 + L - \left(\frac{R^2}{r_s r_t}\right)\cos\psi_{st} \\ L &= \left(1 - 2\left(\frac{R^2}{r_s r_t}\right)\cos\psi_{st} + \left(\frac{R^2}{r_s r_t}\right)^2\right)^{\frac{1}{2}}. \end{aligned}\right\} \qquad (3)$$

Maybe it looks complicated, but it is really nothing for an electronic computer and much easier than an infinite series. This model corresponds to

$$\sigma_n^2 = \begin{cases} A\dfrac{1}{(n-1)(n-2)(2n+1)} & \text{if} \quad n \geqq 3 \\ 0 & \text{if} \quad n \leqq 2. \end{cases}$$

The covariance functions between other types of functionals can be derived (see 9.4) "just" by differentiation of the expression (3). It should then be adequate to choose R and A to make this parametric expression look like the empirical one. The fit seemed to be the best for $A = 7.84888$, and $R/r = 0.9945$, making R be the mean radius of the earth. In fig. 5,

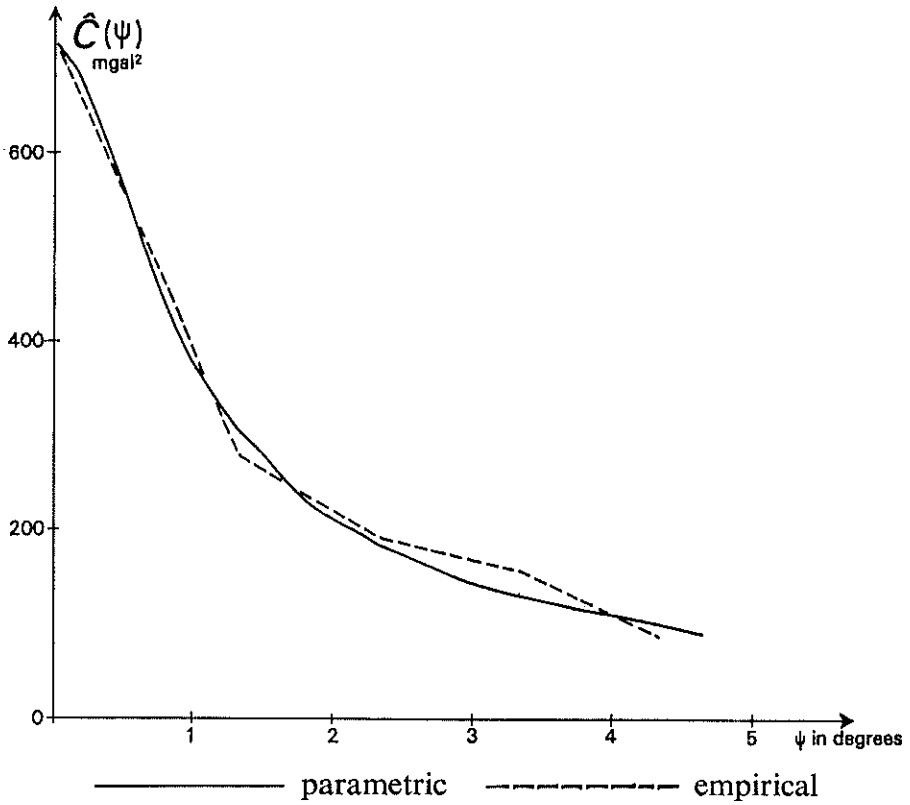

Fig. 7a. Comparison of short-range parts of parametric and empirical covariance function of gravity anomalies.

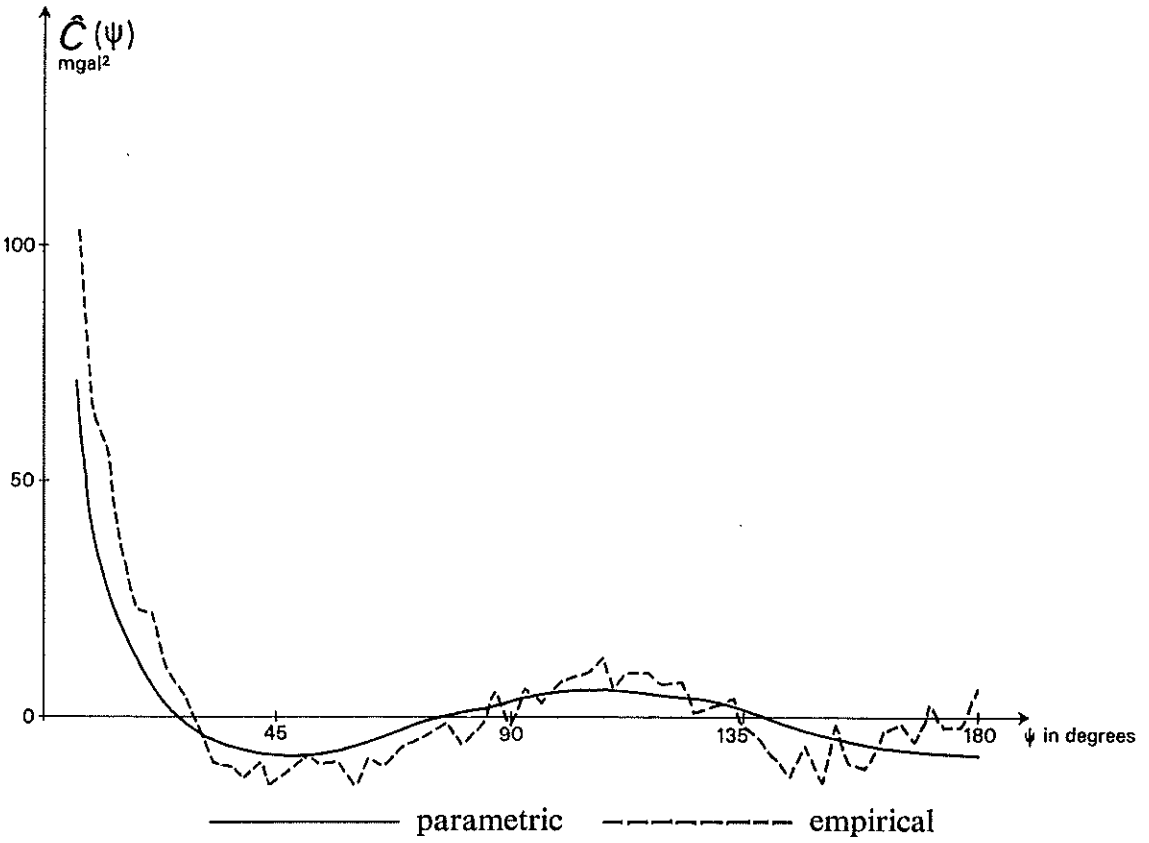

Fig. 7b. Comparison of long-range parts of parametric and empirical covariance function of gravity anomalies.

Table II. Parametric covariance functions between different linear functionals

ψ °	Grav. Anom. Grav. Anom. mGal²	Grav. Anom. Defl. Ver. mGal·arcsec	Grav. Anom. Height Anom. mGal · m	Defl. Ver. Defl. Ver. arcsec²	Defl. Ver. Height Anom. arcsec · m	Height Anom. Height Anom. m²
0	711.00	0.00	220.82	18.37	0.00	302.89
5	86.12	20.44	136.78	1.77	14.68	274.96
10	36.23	13.05	93.53	0.73	17.85	230.47
15	17.43	9.99	62.86	0.15	18.98	180.43
20	7.15	8.01	38.72	− 0.29	18.77	129.24
25	0.73	6.43	19.30	− 0.62	17.52	80.09
30	− 3.44	5.03	3.88	− 0.87	15.50	35.40
35	− 6.08	3.74	− 7.93	− 1.04	12.91	− 3.02
40	− 7.58	2.56	− 16.41	− 1.13	9.98	− 33.95
45	− 8.19	1.48	− 21.83	− 1.15	6.89	− 56.71
50	− 8.08	0.52	− 24.49	− 1.11	3.82	− 71.13
55	− 7.42	− 0.31	− 24.74	− 1.02	0.93	− 77.48
60	− 6.34	− 0.99	− 22.97	− 0.88	− 1.64	− 76.45
65	− 4.96	− 1.50	− 19.57	− 0.71	− 3.79	− 69.02
70	− 3.41	− 1.86	− 15.00	− 0.51	− 5.44	− 56.45
75	− 1.79	− 2.06	− 9.68	− 0.30	− 6.54	− 40.16
80	− 0.20	− 2.10	− 4.04	− 0.10	− 7.08	− 21.66
85	1.27	− 1.99	1.50	0.10	− 7.07	− 2.45
90	2.55	− 1.76	6.59	0.28	− 6.55	16.03
95	3.57	− 1.43	10.92	0.43	− 5.58	32.48
100	4.28	− 1.02	14.24	0.55	− 4.24	45.79
105	4.66	− 0.55	16.36	0.63	− 2.63	55.10
110	4.71	− 0.06	17.18	0.68	− 0.85	59.81
115	4.41	0.44	16.66	0.68	0.98	59.63
120	3.79	0.90	14.84	0.64	2.76	54.56
125	2.89	1.32	11.83	0.56	4.39	44.86
130	1.76	1.66	7.79	0.46	5.78	31.08
135	0.44	1.92	2.94	0.33	6.85	13.98
140	− 0.99	2.08	− 2.47	0.18	7.55	− 5.53
145	− 2.47	2.13	− 8.16	0.03	7.84	− 26.37
150	− 3.92	2.07	− 13.85	− 0.13	7.71	− 47.44
155	− 5.29	1.91	− 19.24	− 0.27	7.17	− 67.60
160	− 6.51	1.66	− 24.08	− 0.41	6.25	− 85.78
165	− 7.52	1.32	− 28.10	− 0.52	5.00	− 101.02
170	− 8.27	0.92	− 31.13	− 0.60	3.49	− 112.52
175	− 8.74	0.47	− 33.01	− 0.65	1.79	− 119.67
180	− 8.90	0.00	− 33.65	− 0.67	0.00	− 122.09

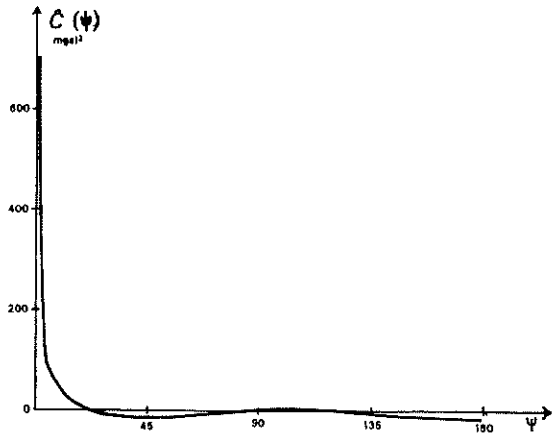

Fig. 8a. gravity anomaly × gravity anomaly.

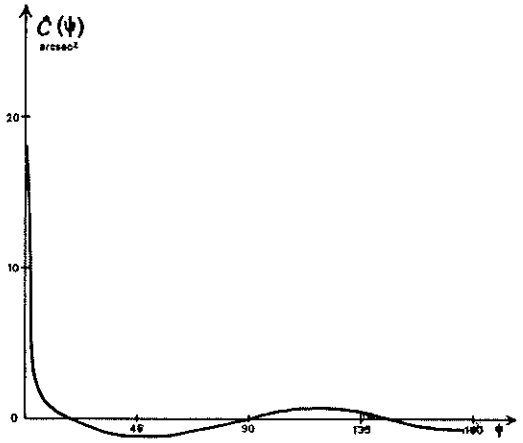

Fig. 8b. gravity anomaly × deflection of the vertical.

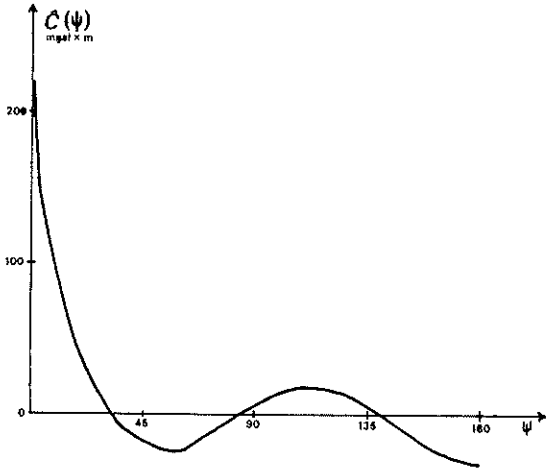

Fig. 8c. gravity anomaly × height anomaly.

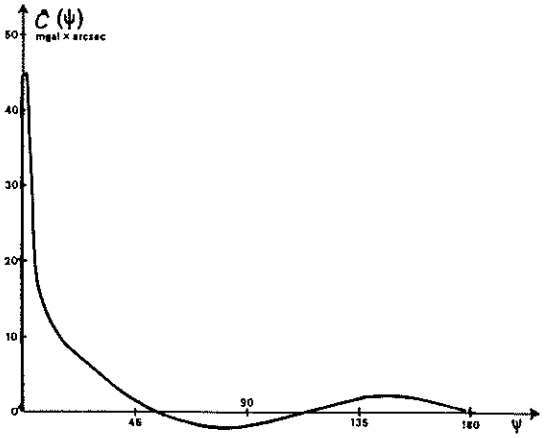

Fig. 8d. deflection of the vertical × deflection of the vertical.

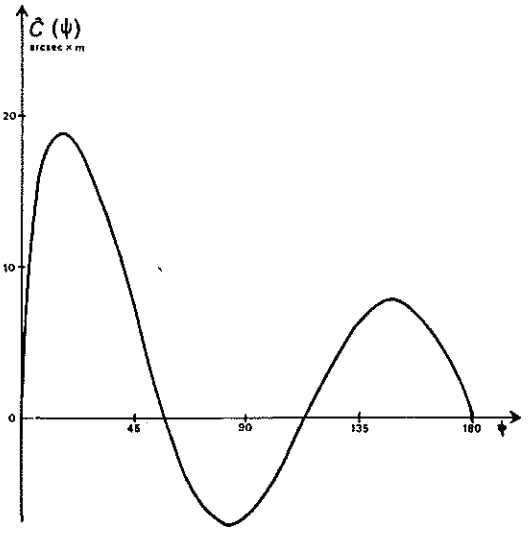

Fig. 8e. deflection of the vertical × height anomaly.

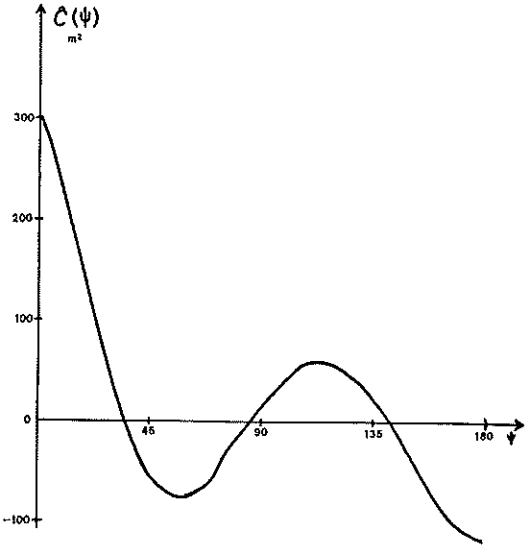

Fig. 8f. height anomaly × height anomaly.

the first part of a mixed empirical covariance function is made using the short-range covariance function from 0° until 5° and the long-range from 5° to 180°. The same part of the parametric covariance function is shown in fig. 6. To get a good comparison of the parametric and the empirical covariance functions, the short-range and long-range parts have been compared in two curves, and the result can be seen in fig. 7a–7b. The fit is remarkable! One could get the impression that there might be some physical reality behind this version of σ_n^2. But at least I can see no reason at all for using the empirically determined degree variances instead of these parametric degree variances. In table II, the covariances between different types of linear functionals are shown using the parametric expression, and the corresponding curves can be seen in fig. 8a–8f.

11. Results and Final Comments

11.1. *General Problem of Prediction*

Let $\xi: T \times \Omega \to \mathbf{R}$ be a second-order random function with known covariance function r and mean value identically zero. Suppose we have observed the process on some subset $T' \subseteq T$, where T' is finite. We want to have a good guess on $\xi(t_0,\omega)$ for $t_0 \notin T'$ from the observed values of ξ. It is an idea to seek the linear combination of the $\xi(t)$'s in T' that minimizes mean square error, i.e. find a_i, so that

$$E|\sum_{i=1}^{n} a_i \xi^{t_i} - \xi^{t_0}|^2, (T' = \{t_1, \ldots, t_n\}) \tag{1}$$

is at minimum. It is easy to see that this $\sum_{i=1}^{n} a_i \xi^{t_i}$ is the *projection* of ξ^{t_0} on the closed subspace spanned by the random variables $\{\xi^t, t \in T'\}$, considered as elements of H_ξ, the *Hilbert space spanned by the process* (see section 5). This random variable we shall call the *prediction of* ξ^{t_0} *from* $\{\xi^t, t \in T'\}$, and we shall write $E^*\{\xi^{t_0} | T'\}$. If we define

$$\mathscr{M}_{T'} = \overline{\text{span}}\{\xi^t, t \in T'\}, \tag{2}$$

we have

$$\forall u \in \mathscr{M}_{T'}: ||\xi^{t_0} - E^*\{\xi^{t_0} | T'\}|| \leqq ||\xi^{t_0} - u|| \tag{3}$$

It is natural to generalize this to T', when T' is an arbitrary subset of T, and *define the prediction of* ξ^{t_0} *as the projection on* $\mathscr{M}_{T'}$ defined by (2). This prediction problem has been treated by many authors, e.g. *Doob* (1953) and *Grenander and Rosenblatt* (1956). *Parzen* (1959, 1961) gave a very general treatment of the problem, applying reproducing kernel Hilbert space theory. We shall here not consider the general theory, but just use the results directly relevant to the practical purposes of this problem, i.e. T' finite.

The important thing is that, as $H_\xi = \mathscr{M}_T$ is *isomorphic to the kernel*

space of the process K_ξ, the prediction problem can be solved in K_ξ and then transformed to H_ξ later. As ξ^t corresponds to r_t, $\mathscr{M}_{T'}$ corresponds to

$$M_{T'} = \overline{\text{span}}\{r^t, t \in T'\}. \tag{4}$$

The prediction of ξ^{t_0} from T' corresponds to the projection of r^{t_0} on $M_{T'}$. The coefficients a_i are determined by the property

$$\langle r^{t_j}, r^{t_0} - \sum_{i=1}^{n} a_i r^{t_i} \rangle_k = 0, \; j = 1, \ldots, n \tag{5}$$

which gives the system of linear equations

$$\sum_{i=1}^{n} a_i r(t_i, t_j) = r(t_0, t_j), j = 1, \ldots, n. \tag{6}$$

If a_i, $i = 1, \ldots, n$ are solutions to (6), it follows that

$$E^*\{\xi^{t_0} | T'\} = \sum_{i=1}^{n} a_i \xi^{t_i}. \tag{7}$$

If we consider $E^*\{\xi^t | T'\}$ as a function of t, the coefficients a_i depend on t,

$$E^*\{\xi^t | T'\} = \sum_{i=1}^{n} a_i(t) \xi^{t_i}. \tag{8}$$

From (6) it is evident that, if $(r(t_i, t_j))$ is *regular* (as matrix),

$$a_i(t) = \sum_{j=1}^{n} r^{ij} r(t_j, t) \tag{9}$$

where r^{ij} are the elements of the inverse matrix to $(r(t_i, t_j))$. Hence

$$E^*\{\xi^t | T'\} = \sum_{i=1}^{n} \sum_{j=1}^{n} r^{ij} r(t_i, t) \xi^{t_i} = \sum_{j=1}^{n} \Big(\sum_{i=1}^{n} \xi^{t_i} r^{ij} \Big) r(t_j, t). \tag{10}$$

We see that having *observed* the values $\xi^{t_i} = m_i$, $i = 1, \ldots, n$, the predictor as a function of t, is given by:

$$\sum_{j=1}^{n} b_j r^{t_j}, \tag{11}$$

when $b_j = \sum_{i=1}^{n} m_i r^{ij}$ are solutions to the system of linear equations

$$\sum_{j=1}^{n} b_j r(t_i,t_j) = m_i, \quad i = 1, \ldots, n. \tag{12}$$

(12) is the most convenient problem to solve seen from a computational point of view. The mean square error of prediction is given by

$$\left.\begin{aligned} \| E^*\{\xi^t \mid T'\} - \xi^t \|_H^2 &= \| \sum_{i=1}^{n} a_i r^{t_i} - r^t \|_k^2 = \\ \langle r^t - \sum_{i=1}^{n} a_i r^{t_i}, r^t \rangle_k &= \\ r(t,t) - \sum_{i=1}^{n} a_i r(t_i,t) = r(t,t) - \sum_{i=1}^{n} \sum_{j=1}^{n} r^{ij} r(t_i,t) r(t_j,t), & \end{aligned}\right\} \tag{13}$$

which is seen from (6). It is not difficult to see that the prediction requires *the inversion of the matrix* $(r(t_i,t_j))$.

The prediction has many properties of the conditional expectation, and if ξ is a *gaussian* process, the two concepts are equivalent. The idea of minimizing mean square error of a linear combination of the observed values has first of all some optimal properties, when the process is not too far away from being gaussian. This is again an argument for considering only gaussian processes here.

11.2. *Prediction of Linear Functionals*

From the formulas in 11.1 it is easy to see that a constant factor on the covariance function does not affect the predictions. If we in our problem consider the covariance function to be determined by the values of σ_n^2 given by (4), section 10.3 and the value of R corresponding to $\frac{R}{r} = 0.9945$, we can suggest the only unknown parameter to be A, which does not affect the predictions. Suppose we have observed $u_1^*, \ldots, u_N^*$ linear functionals, whose covariance matrix would be

$$\langle Ru_i^*, u_j^* \rangle = A \cdot r_{ij} = A \cdot R \tag{1}$$

and suppose further that R is *regular*. The maximum-likelihood estimate of A would then be

$$A = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} u_i^* r^{ij} u_j^* \tag{2}$$

where r^{ij} is the elements of R^{-1}. This also requires inversion of R. As we have actually observed linear functionals where $N \simeq 10^6$, the inversion cannot practically be carried through. On the GIER-computer it is possible to carry through this inversion when N is not much bigger than 100. We have tried to make a prediction from 27 different linear functionals situated as sketched in fig. 9. We have predicted the two components of the deflection of the vertical and the gravity anomalies in 6 points. The points were chosen where the actual value of the functionals were known. In table III a comparison between measured and predicted values together with the prediction error is shown. It looks reasonable.

Of course, it does not show more than that the computing procedure seems to work.

11.3. *Final Comments*

The parametric model used in the present work has the obvious deficiency, that the values of $\sigma_n^2 = \dfrac{A}{(n-1)(n-2)(2n+1)}$ have no physical interpretation, or more correctly, I have not been able to understand the physical meaning of σ_n^2. But the "goodness of fit" is remarkable. Therefore one should think of the prediction as an *algorithm* and not as an optimal procedure having its background in a correct statistical model.

The most important property of an algorithm is that it works and that it is easy to handle. This algorithm satisfies the above properties. Of course, it would be better, if it was possible to make predictions from a larger number of observations. One could say, that this is a problem of getting a very large and fast computer. However, it would be convenient, if it was possible to find a simpler algorithm. This suggests an approach completely different from the statistical one here. But I do not know, if anyone has the idea yet.

I think, that the most important problem to solve is connected with the degree variances. This can only be done if somebody has a very deeply theoretical understanding of what is going on in the earth and what this means to the potential.

It seems to me that it is dangerous to give too much attention to the non-theoretical statistical approach to physical geodesy. Somehow the statistical method (i.e. the prediction theory, harmonic analysis of the covariance function etc.) makes you forget about the relevant problem because it "disappears" in computation and formulas. I think that the non-ergodicity of the process, which I have proved here, is merely due

Fig. 9.

to the fact that problems of understanding physical mechanisms seldom can be solved by statistical methods. An exact theoretical study of the probabilistic background of the method, as I have tried to carry through here, will often tell you, how the problems have been changed and hidden in places, where they are difficult to discover.

Table III. Prediction of linear functionals

a) gravity anomalies

long. °	lat. °	measured mGal	predicted mGal	diff. mGal	standard error
56 44	12 40	29.30	14.98	14.32	12.6
57 17	9 48	− 4.01	− 9.38	5.30	10.2
56 30	10 32	− 8.21	− 4.58	− 3.63	14.6
56 39	10 01	− 7.33	− 8.80	1.47	11.5
56 27	8 48	0.51	− 13.84	14.35	18.5
55 44	12 30	− 11.77	3.10	− 14.87	18.4

b) deflections of the vertical

long. °	lat. °	measured arcsec	predicted arcsec	diff. arcsec	standard error
57 17	9 49	ξ − 3.46	− 2.95	− 0.51	1.7
		η − 0.26	0.48	− 0.74	2.3
56 39	10 00	ξ − 0.57	− 2.25	1.68	2.1
		η 4.15	− 0.09	4.24	4.0
56 30	10 34	ξ − 0.44	− 2.35	1.91	2.0
		η 2.31	− 0.37	2.68	4.5
56 26	8 48	ξ 1.93	− 1.51	3.44	2.0
		η 1.89	1.38	0.51	1.2
55 44	12 30	ξ − 1.54	− 1.54	0.00	1.6
		η 2.97	2.11	0.86	1.7
56 41	12 51	ξ − 2.65	− 2.35	0.30	2.3
		η 3.13	4.79	− 1.66	1.7

Literature

ANDERSSON, S. A.: Inducerede repræsentationer af lokal kompakte grupper. København, 1970.

ARONSZAJN, N.: The Theory of Reproducing Kernels. Trans. Am. Math. Soc., Vol. 68, 1950.

BOURBAKI, N.: Intégration, Chap. VII, 1965.

CRAMÉR, H. AND M. R. LEADBETTER: Stationary and Related Stochastic Processes. Wiley, 1968.

DAVIS, P. J.: Interpolation and Approximation. Blaisdell, 1963.

DIXMIER, J.: Les C*-algebres et leurs représentation. Gauthier-Villars, Paris, 1969.

DOOB, J. L.: Stochastic Processes. Wiley, 1953.

GAPOSCHKIN, E. M. AND K. LAMBECK: Smithsonian Standard Earth (II). Smiths. Astroph. Obs. Sp. Rep. 315, 1970.

GELFAND, I. M. AND Z. SHAPIRO: Representations of the Group of Rotations of 3-dimensional space. Am. Math. Soc. Transl. Ser. 2. Vol. 2, pp. 207–316, 1956.

GELFAND, I. M. AND N. YA. VILENKIN: Generalized Functions. Vol. 4. Application of Harmonic Analysis. 1964.

GRENANDER, U. AND M. ROSENBLATT: Statistical Analysis of Stationary Time Series. Almqvist and Wiksell, Stockholm, 1956.

HEISKANEN, W. A. AND H. MORITZ: Physical Geodesy. Freeman, 1967.

KAULA, W. M.: Statistical and Harmonic Analysis of Gravity. Techn. Rep. Army Map Service, Wash., D.C., 1959.

KAULA, W. M.: Determination of the Earth's Gravitational Field. Rev. of Geophysics. Vol. 1, p. 507, 1963.

KAULA, W. M.: Tests and Combination of Satellite Determinations of the Gravity Field with Gravimetry. Journ. of Geophysical Research. Vol. 71, p. 5303, 1966.

KRARUP, T.: A contribution to the mathematical foundation of physical geodesy. Geodætisk Institut Meddelelse No. 44, København, 1969.

MESCHKOWSKI, H.: Hilbertsche Räume mit Kernfunktion. Springer, 1962.

MORITZ, H.: Interpolation and Prediction of Point Gravity Anomalies. Ann. Aca. Sci. Fenn. Ser. A. III Geol.-Geog. Vol. 69, 1963.

NEVEU, J.: Mathematical Foundations of the Calculus of Probability. Holden-Day, 1965.

OBUKHOV, A. M.: Statistically Homogeneous Random Fields on a Sphere. Uspekhi Mat. Nauk, Vol. 2, No. 2, pp. 196–198, 1947.

PARZEN, E.: Statistical Inference on Time Series by Hilbert Space Methods. Techn. Rep. No. 23. University of Stanford, 1959.

PARZEN, E.: An Approach to Time Series Analysis. Ann. Math. Stat. Vol. 32, No. 4, 1961.

RAO, C. R.: Linear Statistical Inference and its Applications. Wiley, 1965.

ROZANOV, YU. A.: Gaussovskie beskonechnomernye respredeleniya. Trudy Mat. Inst. im. Steklova. Izd, "NAUKA", Moskva, 1968 (Infinitedimensional gaussian distributions).

STERNBERG, W. J. AND T. L. SMITH: The Theory of Potential and Spherical Harmonics. Un. Toronto Press, 1946.

TSCHERNING, C. C.: Bestemmelse af lodafvigelser ved collokation. København, 1970.

WEYL, H.: The Theory of Groups and Quantum Mechanics. Dover, 1931.