LSE CPNSS

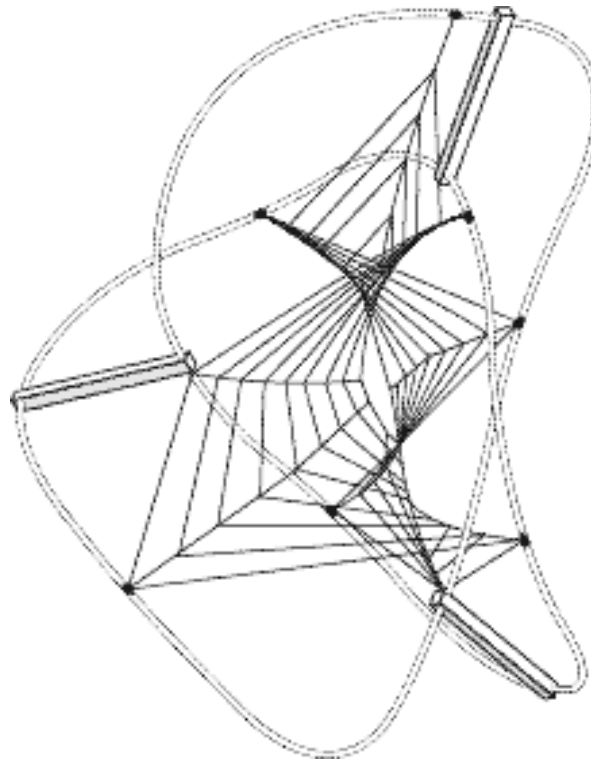# Centre for Philosophy of Natural and Social Science

Discussion Paper Series

DP 76/05

When Evidence is not in the Mean

Marcel Boumans

Editor: Max Steuer

When Evidence is not in the Mean*

Marcel Boumans

When Evidence is not in the Mean[1]

Introduction

> 'It is that kind of a web of observations that he next goes about comparing
> with another one like it, made by a different observer in a different country
> with different instruments and at different times. […] Now I leave it to you
> to judge how much confidence may be placed in deductions made from
> such methods of investigation.' (Galileo 1967, 310)

The crucial problem in metrology is to evaluate the accuracy of the measurement
results. Accuracy is defined in the *International Vocabulary of Basic and General
Terms in Metrology* (IVM 1993) as:

> **3.5 accuracy of measurement.** Closeness of the agreement between the
> result of a measurement and a true value of the measurand (IVM 1993,
> 24)

and can thus not be a quantitative concept, or as is explicitly noted: '"Accuracy" is
a qualitative concept' (IVM 1993, 24). Remarkably, precision is not defined in the
Vocabulary, only closely related concepts as

> **3.6 repeatability (of results of measurements).** Closeness of the
> agreement between the results of successive measurements of the same
> measurand carried out under the same conditions of measurement (IVM
> 1993, 24)

and

3.7 reproducibility (of results of measurements). Closeness of the agreement between the results of measurements of the same measurand carried out under changed conditions of measurement (IVM 1993, 24)

A definition of precision can be found in a statistical dictionary, for example:

> **Precision.** A term applied to the likely spread of estimates of a parameter in a statistical model. Measured by the standard error of the estimator; this can be decreased, and hence precision increased, by using a larger sample size. (Everitt 1998, 260)

The essential difference between 'precision' on the one hand, and 'repeatability' and 'reproducibility' on the other hand, is that in the definition of the latter two the 'conditions of measurement' are explicitly mentioned. These conditions include: principle of measurement, method of measurement, observer, measuring instrument, reference standard, location, conditions of use, and time (IVM 1993, 24).

This IVM dictionary was prepared by a joint working group consisting of experts appointed by: BIPM (International Bureau of Weights and Measures), IEC (International Electrotechnical Commission), IFCC (International Federation of Clinical Chemistry), ISO (International Organization for Standardization), IUPAC (International Union of Pure and Applied Chemistry), IUPAP (International Union of Pure and Applied Physics), and OIML (International Organization of Legal Metrology).

> '[E]very measurement is tainted by imperfectly known errors, so that the significance which can give to the measurement must take account of this uncertainty. We must therefore express with precision that self-same impreciseness.' (IVM 1993, 4)

However, comparing this 1993 version of the Vocabulary with the draft of the 3rd edition (2004), one will find a change of vocabulary. Accuracy has disappeared, and Precision is now introduced into the Vocabulary and defined as:

> **2.35 precision.** Closeness of agreement between quantity values obtained by replicate measurements of a quantity, under specified conditions (IVM 2004, 23)

This definition is similar to repeatability as defined in IVM 1993.

Reason for the disappearance of accuracy in the 3rd edition is a change of approach in metrology, from a Classical (Error) Approach to an Uncertainty Approach. The Classical Approach took it for granted that a measurand can ultimately be described by a single true value, but that instruments and measurements do not yield this value due to additive "errors", systematic and random. In the new Uncertainty Approach, the notion of error no longer plays a role, there is finally only one uncertainty of measurement. It characterizes the extent to which the unknown value of the measurand is known after measurement (IVM 2004, 2). So, instead of evaluating measurement results in terms of errors, it is now preferred to discuss measurement in terms of uncertainty. Uncertainty is defined as:

> **2.11 uncertainty.** Parameter that characterizes the dispersion of the quantity values that are being attributed to a measurand, based on the information used (IVM 2004, 16)

For the evaluation of uncertainty two types are distinguished. Type A evaluation: by a statistical analysis of quantity values obtained by measurements under repeatability conditions; and Type B evaluation: by means other than a statistical

analysis of quantity values obtained by measurement. Precision as defined in statistics is similar to Type A evaluation.

Precision or Type A evaluation can be objectively established for any chosen metric, they are considered to be quantitative concepts. However, accuracy or Type B evaluation depends much more on qualitative knowledge of the measurand itself and cannot assessed in the same objective way. This acknowledgement that objective standards are not enough for evaluating measurement results is admitted in the *Guide to the Expression of Uncertainty in Measurement* (GUM 1993, 8):

> 'Although this *Guide* provides a framework for assessing uncertainty, it cannot substitute for critical thinking, intellectual honesty, and professional skill. The evaluation of uncertainty is neither a routine task nor a purely mathematical one; it depends on detailed knowledge of the nature of the measurand and of the measurement. The quality and utility of the uncertainty quoted for the result of a measurement therefore ultimately depend on the understanding, critical analysis, and integrity of those who contribute to the assignment of its value.'

This is a remarkable position taken by one of the most important bastions of objectivity and positivism. However, it is less surprising if one would go back more than a century ago when these two characteristics of the (modern) scientific image were not that dominant as they became in the 20th century. In this paper the works of two Dutch scientists, Christophorus H.D. Buys Ballot (1817–1890) and Corneille L. Landré (1838-1905), will be studied, who had to deal with a problem mentioned in the quotation at the beginning of this introduction. They had to deal with errors in the observations coming from different sources at different times.

The Mean

At first sight, the works of the meteorologist Buys Ballot and those of the actuary Landré have nothing in common. However, a closer reading of their works show that both men shared a critical view on the applicability of the arithmetical mean as the standard method to treat measurement errors. They found that this method, most appropriate for dealing with errors of measurements produced by precise measuring instruments under similar circumstances, was not applicable to deal with errors of unreliable instruments or in cases where observations are only available at different moments of time, one at each moment. In meteorology and actuarial science, simply averaging the measurement results did not lead to more accuracy. Therefore both men developed alternative 'calculi of observations' in which effects of varying measurement circumstances were explicitly taken account of.

This paper discusses and compares different calculi of observations (data) that aim at providing accurate information about facts about phenomena. The origin of this problem can be found in Galilei Galileo's *Dialogue Concerning the Two Chief World Systems*, where he discussed methods of determining the position of a celestial body, the new star of 1572 (Klein 1997, 149, 151 and Maistrov 1974, 30-34). Twelve observations were made, all of which gave conflicting positions. The problem was deciding which position was the correct position.

> 'Simp[licio]. I should judge that all were fallacious, either through some fault of the computer or some defect on the part of the observers. At best I might say that a single one, and no more, might be correct; but I should not know which one to choose.' (Galileo 1967, 281)

Klein lists the conclusions at which the three characters in the *Dialogue* eventually arrive and which form the basic assumptions of a theory of errors uphold till today (Klein 1997, 151):[2]

1. Errors were inevitable.

    '[T]here is some error in every combination of these observations. This I believe to be absolutely unavoidable, for the observations used in every investigation being four in number ([…] made by different observers in different places and with different instruments), anybody who knows anything about matters will say that it cannot be that no error will have fallen in among the four. Especially when we know that in taking a single polar elevation with the same instrument in the same place and by the same observer (who may have made it many times), there will be a variance of a minute or so […].' (Galileo 1967, 289-90)

2. There is no bias to overestimation or underestimation.

    'I do not doubt that they are equally prone to err in one direction and the other.' (291)

3. Small errors were more probable than larger ones.

    '[F]or granted that these were wise and expert men, one must believe that they would be more likely to err little than much.' (290)

Maistrov (1974, 33) mentions also a fourth conclusion:

4. The greatest number of measurements ought to be concentrated around the true value.

    'And among the possible places, the actual place must be believed to be that in which occur the greatest number of distances, calculated on the most exact observations.' (293)

---

[2] She also mentions a fourth conclusion: 'The size of the errors depended upon the precision of the instrument' (151), which does not appear in her Table 6.2 *Analysis of errors in measurement* summarizing the contributions of several authors, among those of Galileo. Maistrov 1974 discusses a similar conclusion: 'the size of the instrumental errors, so to speak, must not be reckoned from the outcome of the calculation, but according to the number of degrees and minutes actually counted on the instrument' (Galileo 1967, 293). This does not refer to the precision of an instrument but to the sensitivity of an instrument.

Thus, Galileo arrived at the conclusion that errors in measurements are inevitable, that the errors are symmetrically distributed, that the probability of error increases with the decrease of the error size, and that the majority of observations cluster around the true value (see also Maistrov 1974, 33).

Simplicio believed that truth lay only in a single actual observation. This belief was a common one in Galileo's days, and had still to be argued against in the eighteenth century, when it became more and more accepted that the arithmetical mean of observations was an appropriate method to treat observational errors. For example, Klein quotes Thomas Simpson (1710-1761) who wrote in 1755 (also quoted by Stigler1986, 90):

> 'My Lord, It is well known to your Lordship that the method practised by astronomers, in order to diminish the errors arising from the imperfections of instruments, and of the organs of sense, by taking the Mean of several observations, has not been so generally received, but that some persons, of considerable note, have been of opinion, and even publicly maintained, that one single observation, taken with due care, was as much to be relied on as the Mean of a great number.' (Simpson 1755, 82-3)

Simpson showed that

> 'the taking of the Mean of a number of observations, greatly diminishes the chances for all the smaller errors, and cuts off almost all possibility of any great one: which last consideration, alone, seem sufficient to recommend the use of the method, not only to astronomers, but to all others concerned in making of experiments of any kind (to which the above reasoning is equally applicable). And the more observations or experiments there are made, the less will the confusion be liable to err, provided they admit of being repeated under the same circumstances.' (92-3)

In other words, this method of taking the mean of observations is based on the assumption that by averaging the observations, the observational errors will cancel each out.

So, though the arithmetical mean was considered for a long while already as a method to obtain true values, it was not clear why it produces the best results, that is, why the errors will cancel out (the reasons Galileo gave were not considered to be valid because they are connected to a human agency). A crucial step to the justification of this method was made by Carl Friedrich Gauss (1777-1855) in his *Theoria motus corporum coelestium*. The main topic of Gauss's book was an investigation of the mathematics of planetary orbits. At the end, he added a section on the combination of observations. The innovative approach was that Gauss assumed that the possible values of errors ε have probabilities given by a curve $\varphi(\varepsilon)$.[3] Gauss noted that *a priori* he could only make general statements about $\varphi(\varepsilon)$: It would be at a maximum at $\varepsilon = 0$; it would be symmetric; and it would be zero outside the range of possible errors. Instead of imposing further conditions directly, he assumed the conclusion. He adopted as a postulate that when any number of equally good direct observations of an unknown magnitude are given, the most probable value is their arithmetic mean, and subsequently he proved that the distribution must have the form of what would later be called the Gaussian, or normal, curve:

$$\varphi(\varepsilon) = \frac{h}{\sqrt{\pi}} e^{-h^2 \varepsilon^2}$$

for some positive constant h, where h could be viewed as a measure of precision of observation. He then showed how in the more general situation this error distribution led to the method of least squares.

---

[3] See for an exposition of his argument Stigler 1986 and Whittaker and Robinson [1924] 1944.

Stigler emphasizes that Gauss's argument was essentially both circular and non sequitur (Stigler 1986, 141). In outline its three steps ran as follows: The arithmetic mean (a special, but major, case of the method of least squares)[4] is only 'most probable' if the errors are normally distributed; the arithmetic mean is 'generally acknowledged' as an excellent way of combining observations so that errors may be taken as normally distributed; finally, the supposition that errors are normally distributed leads back to least squares. It was Laplace who cut this circularity with the central limit theorem (Laplace 1809). Laplace showed how this theorem could provide a better rationale for Gauss's choice of $\varphi(\varepsilon)$ as an error curve: If the sums of errors are considered, then the limit theorem implied they should be approximately distributed as the normal curve $\varphi(\varepsilon)$. Note that in the central limit theorem it is assumed that the circumstance in which the observations are made are identical and that the normal curve is a good approximation if the number of observations is very large.

In his *History of statistics*, Stigler posits explicitly that: 'The method of least squares was the dominant theme – the leitmotif – of nineteenth-century mathematical statistics. In several respects it was to statistics what the calculus had been to mathematics a century earlier' (Stigler 1986, 11). Ten years after Legendre's 1805 appendix – where the method of least squares was first introduced by Adrien Marie Legendre (1752-1833) -, the method of least squares was a standard tool to deal with measurement errors in astronomy and geodesy in France, Italy, and Prussia. By 1825 the same was true in England. Both sciences were driven from the theory of probability as applied to the choice of the best value to be adopted for the measure of a physical quantity when there are a large number of independent determinations, equally trustworthy so far as skill and care are concerned, yet differing from one another within the limits of actual measurement. The method of least squares was a method to assess the measurements obtained with precise instruments. However, a science like

---

[4] In the case of observations y_i of only one value x, the squared errors, $\Sigma_i\varepsilon_i^2 = \Sigma_i(y_i - x)^2$, is 'least' (minimal) when: $\frac{1}{n}\Sigma_i y_i = x$.

meteorology, which concerned itself very little with the repeated measurement of physical quantities that were supposed to be without variation and where measurements were not equally trustworthy, required other methods for obtaining meteorological facts. And in actuarial science, where no instruments are used at all, it happened to be that a century after its introduction the method of least squares was still subject of discussion.

The method of least squares is based on the theory of the properties of great numbers of observations. However, this theory is used in meteorology in an entirely different manner. It is applied to observations which are designed and intended to register changes in the quantity which is measured in determining for example, daily, monthly, seasonal or secular changes in the meteorological elements such as pressure, temperature, humidity, rainfall, sunshine or any other. It therefore borrows from the science of probabilities the methods which have been developed to deal with errors, and applies them to such cases as the variation of annual temperature or rainfall in the measurement of which no errors at all are assumed. The differences are assumed to be due to natural causes which are certainly real but which are unknown to those who make the observations: 'the most important causes are always to be sought in the variations (deviations)' (Buys Ballot 1850, 45). These deviations are the differences between the measurement of meteorological element at a specific place and date and its so-called 'normal'. The idea of a normal for a period of years is that the mean value for every period of the same length of a very long series would give the normal and that necessarily would imply perfect recurrence at the completion of the period. Buys Ballot formulated the aims of meteorology in twenty-one propositions.[5] The fourteenth proposition noted that: 'Moreover, the averages cover all influences that do not manifest themselves each year on the same date; deviations make them conspicuous' (Buys Ballot 1851, 2). Research based on this principle that deviations of normals would teach us much more

---

[5] These were presented in the introduction of the first yearbook of the Dutch meteorological institute (Buys Ballot 1851). A survey of these propositions can be found in KNMI 1954, 20-4.

than the direct observations led to his famous meteorological law. Buys Ballot had to deal with measurements of unreliable instruments collected from different places all over the world. Deviations were trustworthier; they eliminated the deficiencies of these instruments.

In actuarial science, where one often has to deal with time series, not every observation has the same significance to determine the true value of e.g. the birth rate or death rate at a certain moment of time. Instead of taking simply an arithmetical mean, which assumes equal significance of each observation, the true values are determined by a weighted average. To choose among all possible weighting systems, Landré insisted on the concurrent observation of a certain property having the greatest weight. This requirement is not to be found anymore in current graduation methods.

Because of the different kinds of problems both men had to deal with, they developed alternative types of calculi of observations for the fields they worked in. What these problems were and how they treated them will be discussed in the next sections. Each will be introduced by a brief biography.[6]

Brief Biography of Christophorus Henricus Didericus Buys Ballot
Born October 10, 1817 at Kloetinge, Netherlands. Initially, he started studying literature at the university of Utrecht (BA Literature, June 1838), but soon after he also took courses in mathematics and physics. His main interest was mathematics, than came astronomy (BA Mathematics and Philosophy magna cum laude, September 1839). He finished his doctoral thesis in 1844. July 1845 lector mineralogy and geology, 1846 lector theoretical chemistry, November 1847 extraordinary professor in mathematics, 1854 founder and first director of the Royal Dutch Meteorological Institute, April 1855 member of the Royal Academy of Sciences, 1857 professor in mathematics and in 1867 professor in physics. He

---

[6] For an extensive biography of Buys Ballot, I refer to Van Everdingen 1953. Biographies of Landré are to be found in three memoriams written by G.J.D. Mounier (1906a), his daughter Henriette F. Landré (1906) and M.C. Paraira (1905-1907).

is famous for the meteorological law he formulated in 1857: In the Northern Hemisphere, if one stands with his back to the wind, the area of low pressure is to the left. In the Southern Hemisphere the reverse is true. 1873 first chairman of the International Meteorological Committee, a precursor of the World Meteorological Organization. Died February 3, 1890.

On the Great Importance of Deviations from the Mean State

In a paper 'On the great importance of *Deviations* from the mean state of the Atmosphere for the Science of Meteorology', Buys Ballot explicated his method to uncover the main causes of temperature and pressure. This method was based on four 'propositions':

'I. The average temperature which prevails at any certain place is not that which is generated there by the action of the sun, &c., and which would depend simply on the latitude and the elevation of the ground, but is remarkably changed by the influences of other regions, particularly by the action of the winds.

II. That average temperature, such as it is obtained anywhere from observations during a series of years, for the different months or days of the year, will by no means always prevail at those places for the determined month or day of each single year. On the contrary, observations give generally great variations; and it is precisely the magnitude of these variations which it is of the utmost importance to learn.

III. What we asserted regarding the temperature in proposition II. applies equally to all meteorological indications; it is of great importance to become acquainted with those variations of the barometer, and of the force and direction of the wind.

IV. The most efficient means for prognosticating the weather are, the employment of the electric telegraphs and of self-registering instruments, because they facilitate and make possible a tabular union of the variations mentioned in II. and III.' (Buys Ballot 1850, 43)

From these propositions he deduced eight different kinds of averages: The 'mean theoretical temperature of a place' for a long period of time (e.g. a year or a season), indicated by Θ, and for a shorter period of time (e.g. a month or a day), indicated by θ. The 'mean temperature deduced from long series of observations', indicated by MT for a long period of time, and mt for a shorter period of time. In his later publications, Buys Ballot called the latter 'normals'. And, finally, the 'observed mean temperature', for a long period of time OT, and for a shorter period of time ot. The values of the theoretical temperatures Θ and θ of a particular place are only determined by the latitude and altitude of that place, and thus are the temperatures if that place would not receive warmth from and emit warmth to the surrounding environment. Theoretically we should infer these values from 'the warmth which emanates from the sun to us every day; from the warmth which every day and night issues from beneath the surface of earth; and from the warmth that is produced by animals, consumed by plants, lost by radiation, given by condensation of vapour' (44). MT and mt are the averages of series of observations of the temperature made at different locations on the same altitude and latitude. So, MT and mt could be used as estimates of Θ and θ respectively. The value of e.g. MT so obtained would, however, be somewhat greater than Θ, 'because there is more air drawing towards the north than towards the south over a whole parallel circle, that southern air at the same time being warmer; and also because, near the equator, the latent heat which is employed in the vaporizing of water is greater than that which is freed by rain; and, on the contrary, in higher latitudes there is more freed than expended on the formation of vapour' (44). The difference between Θ and MT should be ascribed to the influence of the wind during a season, similar the difference between θ and mt should be ascribed to the influence of the wind during a day.

MT should, thus, be considered as the 'equilibrium state of temperature at a determined season of the year at that place'. The differences OT-MT and ot-mt should then be ascribed to circumstances as the wind having not its usual

direction in that period of time, or the distribution of temperature at the surrounding places being different from usual.

> 'Thus it is necessary that we know the variations, not only for the place itself for which we desire to explain the temperature, but also for the surrounding places, since the variations at the first place must be explained partly from the variations at the latter. The most important causes are always to be sought in the variations (deviations); it is from those that we must derive the exhibition of the state of temperature, not from the absolute observed temperature.' (Buys Ballot 1850, 45)

These conclusions also applied to barometer readings: 'the deviations again are the greatest importance, especially as here the theoretical state is known for every latitude' (45).

The method Buys Ballot implicitly proposed in this paper was the method of residues, as called by Whewell ([1847] 1967). This method should be applied when a combination of influences are operating at the same time:

> 'When we have, in a series of changes of a variable quantity, discovered *one* Law which the changes follow, detected its argument, and determined its magnitude so as to explain most clearly the course of observed facts, we may still find that the observed changes are not fully accounted for. When we compare the results of our Law with the observations, there may be a difference, or as we may term it, a *Residue*, still unexplained. But this Residue being thus detached from the rest, may be examined and scrutinized in the same manner as the whole observed quantity was treated at first: and we may in this way detect in *it* also a Law of change. If we can do this, we must accommodate this new found Law as nearly as possible to the Residue to which it belongs; and this being done, the difference of our Rule and of the Residue itself, forms *a Second Residue*.

This Second Residue we may again bring under our consideration; and may perhaps in *it* also discover some Law of change by which its alterations may be in some measure accounted for. If this can be done, so as to account for a large portion of this Residue, the remaining unexplained part forms a *Third Residue*; and so on.' (Whewell [1847] 1967, 409-10)

When comparing this method with the Method of Means, Whewell notes that both methods are opposite to each other: 'For the Method of Residues extricates Laws from their combination, *bringing them into view in succession*; while the Method of Means discovers each Law, not by bringing the others into view, but by *destroying their effect* through an accumulation of observations' (411).

Whewell referred to Herschel's *Discourse on the study of natural philosophy* for a treatment of this method in a wider sense.

'Complicated phenomena, in which several causes concurring, opposing, or quite independent of each other, operate at once, so as to produce a compound effect, may be simplified by subducting the effect of all the known causes, as well as the nature of the case permits, either by deductive reasoning or by appeal to experience, and thus leaving, as it were, a *residual phenomenon* to be explained.' (Herschel 1836, 156)

Buys Ballot's Principle

Buys Ballot aimed at a worldwide network of observations. His first experience with such network was acquired during his student days, when he and his study friend Frederick Wilhem C. Krecke were involved in a project run by Richard van Rees, their professor mathematics and physics at the University of Utrecht (see Van Everdingen 1953, 24, and Van Lunteren 1998, 226). From 1839 till 1843, both students assisted Van Rees in making meteorological observations on each

21st of the months March, June, September and December (Van Everdingen 1953, 27, and Buys Ballot 1882, 2).

John Herschel, England's leading scientist at that time, initiated this project when he stayed in South Africa (Cannon 1961). From 1835 worldwide on more than sixty locations meteorological observations were carried out on the 21st of the months March, June, September and December. Participants were expected to do these observations each hour during a period of 36 hours according to a number of prescribed instructions. These instructions were published as 'Instructions for making and registering meteorological observations at various stations in Southern Africa and other countries in the South Seas, as also at sea'. In 1838 he ceased this project because he considered this worldwide network too coarse-meshed (Van Lunteren 1998, 219). However, he asked Quetelet to carry out this project on a smaller scale in Belgium. This Belgium network soon spread across whole Europe.

Herschel's instructions were meant as 'the means of rendering their observations most available for useful purposes, and comparable with each other, and, with those intended to be referred to as standards' (Herschel 1836, 138). The instructions consisted of three parts: 'I. General Recommendations and Precautions', 'II. Of the Times of Observations and Registry', and the by far largest part 'III Of Meteorological Instruments, and first of the Barometer and its attached Thermometer'. What is striking about these instructions is that they are very detailed, in the sense of that they stretch to every tiny detail of measurement readings. For example, one was instructed 'before reading off, give a few taps on the instrument, enough to make the upper end of the column of quicksilver *shake* visible, as the mercury is apt to adhere to the glass and give erroneous readings' (141).

Twelve years later, soon after the establishment of a meteorological observatory at Utrecht, Buys Ballot published a call 'to all friends of meteorology' for

participation in a project of providing observations from across The Netherlands (Buys Ballot, 1848). The nature of the instructions he gave in this call were strikingly opposite to those by Herschel. The observations didn't need to be as precise as they were in astronomy, and only three observations on 'easy' hours are welcome, and even observations twice a day would make him very content. Moreover, these observations do not to be taken on the same hours each day. To underpin his instructions, Buys Ballot even quoted an opponent as starting point for explicating them:

> 'For me it is impossible to suppress a feeling of distrust, whenever I consider these immense series of observations, from which one hopes to achieve knowledge about laws and causes of the phenomena in our atmosphere. Nobody is going to assert that *those* will fully meet the requirements of precision, and also that no greater correctness is now already achievable and necessary; but one has also to note the proportional minor knowledge, that is inferred from these and the uncertainty to which it is also rightly subjected.' (Van Willigen, quoted in Buys Ballot 1848, 380)

Van der Willigen replied to Buys Ballot's call with an extensive discussion of the sense and meaning of precision in the natural sciences (Van Lunteren 1998, 231): 'It cannot be said often enough: precision and progress in science are to the physicist words of the same meaning' (Van der Willigen, quoted in Van Lunteren 1998, 231).

Only a year before this exchange Van der Willigen had received his PhD from the Leiden astronomer Frederik Kaiser (1808-1872). Kaiser was internationally renowned for his precise measurements (Beek 2004). In general, this strive for precision was not uncommon in those days, because, particularly in astronomy, one's reputation as a scientist depended on it. In astronomy, the quality of the

observations revealed the quality of the researcher, correctness and preciseness were the permits for a successful scientific career.

However, in meteorology, as Buys Ballot perceived it, the number and spread of observations were more important than more preciseness. In his propositions published in 1851, he explained why:

> '12. After all, it is good that at some observatories very extensive observations are made that fill bulky volumes: that is important for *Climatology*, however it is preposterous to think that with them one meets the needs of *Meteorology*. The weather situation at one place depends on those at the surrounding places; thus observations at one place cannot learn us more than that one already likely thinks to gather from the judicious connection of the indication of barometer, thermometer and wind vane.
> 13. Because in a new perspective long continued observations become important, when they are linked to those of the surrounding places, is it now infinite better to make simple observations at hundred places not too far from each other, than at ten places very complete ones.' (Buys Ballot 1851, 2)

For receiving observations as much and as dispersed as possible, Buys Ballot was well aware that he had to deploy 'friends of meteorology' who were not always experts using the best instruments available and who would be able to make their observations under optimal conditions. To study meteorology, quantity was preferred above quality.

The readings of the thermometers and barometers were published in yearbooks. The first was published in 1851 and contained tables with the daily deviations form the normals for the years 1849 and 1850. The first yearbook published only Dutch data, but in the subsequent yearbooks gradually more and more

measurements and deviations from normals of foreign places were presented. These data received from other European places were used to calculate the normals of these places.

Buys Ballot's aim was to have ultimately a worldwide network of observatories,

> 'the whole of the globe must be covered with a network of observatories, where observers placed at equal distances are able to watch all phenomena of the weather as accurately as possible. […] which is preferable, observations at a thousand well disposed observatories for two or three years in addition to more years at some of them, or a series of hundred of years at a dozen stations? We prefer the former alternative.' (Buys Ballot 1872, 18-9)

Such a network, of course, asked for an international coordination and standardisation of the observations. For that reason the Dutch Meteorological Institute published Buys Ballot's pamphlet *Prière à tous ceux qui veulent bien de la Météorologie* (Plea to those who mean good to meteorology). The first international meeting had, however, to wait till 1872. This meeting in Leipzig was considered as preparatory to the official conference at Vienna the next year. As a preparation for the Leipizg meeting, Buys Ballot wrote a booklet *Suggestions on a uniform system of meteorological observations*. The aim for a global system had aggravated the problem that one has to deal with observations from all over the world of which one could never have any certitude about their validity. Therefore, Buys Ballot insisted, again, on publishing the departures from the normals: 'Departures are perfectly independent of the daily and annual range, and of the local disposition and correction of the instruments, because the normals are likewise computed for these circumstances' (18).

The problem however was that these calculated normals needed correction because some barometers have been moved or replaced or the hour of observation has changed, 'if we change an hour of observation, as it is, if the

climate of a place is changed'(Buys Ballot, 1872, 14-15; 1882, 51). For this reason, Buys Ballot persistently choose to publish departures from the normals: when comparing an instrument's reading with the normal reading of that instrument one is independent from the errors of the instrument and from the influence of the latitude (Buys Ballot 1872, 16; 1882, 52). So, Buys Ballot in fact suggested to the international community that they should import the Dutch system of presenting observations. This Dutch system was described in paper published in an engineering journal:

> 'It is obvious, but not generally acknowledged, that no absolute reading of the barometer has any significance, but only the difference (called departure) of an actual reading with the average reading of *that* instrument at the *same* place, at that latitude, longitude and height above the sea on the same day. The departure is the true and accurate measure of the perturbance [sic][7], and intimately connected, but, as we shall see, not identical, with the force that tries to restore the equilibrium. The single reading of the barometer, on the contrary, is an arbitrary number of no signification at all, unless you substitute an accurate approximation of the average height of the readings. I therefore call the determination of the average or normal height *the* characteristic of the Dutch system, and this base is adopted in the Dutch meteorological Annuaires, which have appeared regularly since 1852, and wherein for every day of the year a tabular view is given, representing the departure of temperature and atmospheric pressure for the whole of Europe.' (Buys Ballot 1865, 246)

Buys Ballot's biography written by E. van Everdingen, the fourth director of the KNMI from 1905 to 1938, discusses this so-called Buys Ballot's principle (Van Everdingen 1953, 86-7). This principle was that one should prefer to work with deviations instead absolute observations. Deviations have the advantage that they eliminate the errors of the instrument and position (see also KNMI 1954,

---

[7] He probably meant perturbation, in the sense of causing deviation of motion or of other behavior.

19). If a barometer has the wrong zero point or positioned on another altitude than assumed, the difference between an observation and a normal would nullify these instrumental errors. 'Every place self should take care of the virtue of their instruments. The comparison let us completely know their flaws' (Buys Ballot quoted in KNMI 1954, 18).

Buys Ballot's method can be reconstructed as follows. The facts to be found are the daily variations of, for example, pressure. They have to be inferred from daily readings of unreliable instruments. To deal with the errors, the method of arithmetical means - normals - is used, not to erase them but, on the contrary, to capture them. Let y_t be a reading of an instrument supposed to measure, say, pressure of a specific moment, t, on a specific day, e.g. January 10 at 12 am. Each particular moment on a particular day, t, is characterized by a specific value, x_t, indicating the 'normal' pressure typically for that particular moment on that particular day of the year. The instrument's error ε is caused by unknown defects of the instrument itself (not calibrated or calibrated with the wrong standards. For example the barometer is put on the wrong altitude). Buys Ballot was interested in the daily variations z_t. As a result:

$$y_t = x_t + \varepsilon + z_t$$

To uncover the deviations z_t, Buys Ballot first calculated the normal of that particular moment. A normal N_t was an arithmetical mean of all measurement readings of that particular instrument on the same moment across an as long period as available, e.g. ten years. Buys Ballot assumed that this normal would consist of the unknown true value and the unknown instrument's error:

$$N_t = x_t + \varepsilon$$

Then the deviation z_t, obtained by subtracting the normal from the observation,

$$z_t = y_t - N_t$$

does not contain the instrument's error anymore.

Brief Biography of Corneille Louis Landré

Born August 31, 1838 at Utrecht. Studied mathematics at the University of Utrecht, taking courses from Van Rees and Buys Ballot. As a student Landré, for a while, worked at the Meteorological Institute headed by Buys Ballot, where he calculated normals of the barometer of different observatories. In 1876, he was appointed as mathematician at the life-insurance company Dordrecht. In 1881 appointed as lecturer in mathematics at the normal school for teachers at Dordrecht. In 1895 appointed as actuary of the General Company of Life-insurance and Life-annuity at Amsterdam. Member of the Dutch mathematical society in 1875, from 1898 till 1905 member of the board and in 1902 and 1903 chairman. 1895 editor of the *Archief voor de Verzekerings-Wetenschap*. Cofounder and later board member of the society of mathematical consultants. Member of the Conseil de Direction of the Comité Permanent des Congrès Internationaux d'Actuaires. Died February 10, 1905.

Landré was in his days the actuary with the highest international reputation, mainly due his textbook that beside a Dutch edition *Wiskundige Hoofdstukken voor Levensverzekering* also was translated and published in a German edition *Mathematisch-Technische Kapitel zur Lebensversicherung* (Van Haaften 1923, 137).

Method of Least Squares

Landré had always been very critical towards the application of the method of least squares to actuarial science, it was based on 'hazardous suppositions' and would not in every case lead to the best solution (Mounier 1906b, 309). Landré never published his doubts about the applicability of the method of least squares to actuarial science (though he has written two articles on this method (Landré

1881, 1884)), but his fellow editor of *Archief voor de Verzekerings-Wetenschap en Aanverwante Vakken* (archive of actuarial science and related subjects), Guillaume Jacques Daniel Mounier (1854-1917) wrote two articles on this method on the suggestion by Landré (Mounier 1903, 1906).

Mounier mentioned four suppositions of the method of least squares:
1. $\varphi(\varepsilon) = \varphi(-\varepsilon)$;
2. the best method and the best instruments are used as well as possible;
3. the number of observations is very large, preferably as large as possible;
4. the probability of the error is an infinite small function of the error itself (Mounier 1903).

According to Mounier these suppositions reveal the original application of this method, namely inferring the most probable value from measurements giving different results. To apply this method to statistical research the suppositions should be revised by replacing the term 'error' by 'deviation' and by removing any reference to instruments. He, then, continues with discussing the four suppositions. Supposition 1 follows from supposition 2 and the latter means that the observations should be as precise as possible. The third supposition is related to the law of large numbers and aims at eliminating all random influences. It's meaning, according to Mounier, is that the method of least squares does not apply when the number of observations is not large.

Graduation

One of Landré's involvements in actuarial science was the construction of mortality (or life) tables. One important aspect of the construction of such tables is the graduation – smoothing - of the mortality data. The irregularity of a time series is an indication of errors of observations which should be removed. But now each moment of time there is only one observation/measurement of the phenomenon possible. Measurements taken at other moments (past and future) are used to remove errors but cannot simply be averaged. The method being

applied is the so-called moving weighted average (MWA) methods. The problem is to choose the appropriate weighting system. In several publications Landré discussed and compared various kinds of graduation methods (Landré 1889, 1900a, 1900b).

A graduated value v_t is produced as a weighted average of a certain number of consecutive observed values. The basic formula is

$$v_t = \sum_{i=-n}^{n} a_i y_{t+i}$$

where the a_i's form the weight system. In most cases the MWA formulae are symmetric: $a_i = a_{-i}$. The problem is to find a_i so that v_t is the best estimate of x_t. To find these values, it is assumed that the sequence x_t is closely represented by a low degree polynomial, such as a cubic. As a result of this approach there are still an infinite number of graduation formulae possible. Landré discusses several options, those of Finlaison, Woolhouse, Higham and Karup, but also combinations of them. To evaluate these different formulae, Landré used the following criterion: the revised (graduated) value v_t of y_t should be a *correction* of y_t and not a *rejection*:

$$v_t = y_t + k_t$$

That is, k_t should be such that it does not remove the original observation, y_t, from the formula: a_0 should not be equal to zero. Moreover, Landré explicitly preferred formulae for which the observation, y_t, has the greatest weight, in other words the weight a_0 is larger than any other weight a_i (Landré 1990a, 326).

However, this criterion is not sufficient to choose among the still many possible graduation formulae. The second criterion he proposed is to choose the one with the smallest mean squared error – a criterion on which modern graduation

methods are based. Landré's above correction criterion doesn't play any role today.

Conclusion

When observing or measuring phenomena, errors are inevitable, one can only aspire to reduce these errors as much as possible. An obvious strategy to achieve this reduction is by using more precise instruments. Another strategy was to develop a theory of these errors that could indicate how to take them into account. One of the greatest achievements of statistics in the beginning of the 19th century was such a theory of error. This theory told the practitioners that the best thing they could do is taking the arithmetical mean of their observations. This average would give them the most accurate estimate of the value they were searching for. Soon after its invention, this method made a triumphal march across various sciences. However, not in all sciences one stood waving aside. This method, namely, only worked well when the various observations were made under similar circumstances and when there were very many of them. And this was not the case for e.g. meteorology and actuarial science, the two sciences discussed in this paper.

In meteorology, each measurement came from a different instrument of which its reliability was not clear. Buys Ballot – the leading Dutch meteorologist of his days - had to develop a calculus of observations not based on the method of means but on the method of residues to turn unreliable observations into accurate estimates.

In actuarial science, measurements are not produced by an instrument but are the result of counting. Still, errors are inevitable. Here, we don't have many observations of a phenomenon at a certain point of time made under various circumstances (similar or not) but a time series that gives just one observation for each different point of time. To find the most accurate estimate of the variable at a certain point of time, the different observations should not be valued equally. A

simple arithmetical average would neglect the kind of useful information provided by the observation of that variable (though erroneous) at the same time compared with observations made at other moments. Landré – the leading Dutch actuary at that time – designed a requirement for a weighted average that most emphasised the concurrent observation of the variable one is measuring.

References

Beek, L. (2004) *De geschiedenis van de Nederlandse natuurwetenschap* (History of Dutch natural science), Kampen.

Buys Ballot, C.H.D. (1848) 'Iets over de meteorologische waarnemingen aan het observatorium te Utrecht' (On meteorological observations at the observatory at Utrecht), *Algemeene Konst- en Letterbode*, 379-384.

Buys Ballot, C.H.D. (1850) 'On the great importance of *deviations* from the mean state of the atmosphere for the science of meteorology', *The London, Edinburgh and Dublin Philosophical Magazine* 37.247, 42-49.

Buys Ballot, C.H.D. (1851) *Uitkomsten der meteorologische waarnemingen gedaan in 1849 en 1850 te Utrecht en op eenige andere plaatsen in Nederland* (Results of the meteorological observations done in 1849 and 1850 at Utrecht and on some other locations in The Netherlands), Utrecht.

Buys Ballot, C.H.D. (1865) 'On meteorological observations as made in Holland', *Civil Engineer and Architect's Journal* 28, 245-6.

Buys Ballot, C.H.D. (1872) *Suggestions on a uniform system of meteorological observations*, Utrecht.

Buys Ballot, C.H.D. (1882) *Beredeneerd register op de werken van het Koninklijk Nederlandsch Meteorologisch Instituut tot 1882* (Motivated register of the works of the Royal Dutch Meteorological Institute till 1882), Utrecht.

Cannon, W.F. (1961) 'John Herschel and the idea of science', *Journal of the History of Ideas* 22, 215-9.

Everitt, B.S. ed. (1998) *The Cambridge Dictionary of Statistics*, Cambridge.

Galileo, G. (1967) *Dialogue concerning the two chief world systems –Ptolemaic & Copernican*, trans. by Stillman Drake, 2nd ed., Berkeley, Los Angeles, London.

Gauss, C.F. (1809) *Theoria motus corporum coelestium*, Hamburg.

GUM (1993) *Guide to the Expression of Uncertainty in Measurement*, ISO.

Herschel, J.F.W. (1830) *Preliminary discourse on the study of natural philosophy*, new edition, London.

Herschel, J.F.W. (1836) 'Instructions for making and registering meteorological observations at various stations in Southern Africa and other countries in the south seas, as also at sea', *Edinburgh New Philosophical Journal* 21, 135-149.

IVM (1993) *International Vocabulary of Basic and General Terms in Metrology*, 2nd edition, ISO.

IVM (2004) *International Vocabulary of Basic and General Terms in Metrology*, 3rd edition, draft April 2004, ISO.

KNMI (1954) *Koninklijk Nederlands Meteorologisch Instituut 1854-1954* (Royal Dutch Meteorological Institute 1854-1954), Den Haag.

Klein, J.L. (1997) *Statistical visions in time. A history of time series analysis 1662-1938*, Cambridge.

Landré, C.L. (1881) 'Over de functie ϕ van de methode der kleinste kwadraten' (On the function ϕ of the method of least squares), *Nieuw Archief voor Wiskunde* 7, 214-219.

Landré, C.L. (1884) 'De middelbare fout bij waarneming ter bepaling van meer dan een onbekende' (The mean error of observation to determine more than one unknown), *Nieuw Archief voor Wiskunde* 10, 1-17.

Landré, C.L. (1889) 'Over correctie van getalreeksen door middel van tweede verschillen' (On the correction of time series by second differences), *Nieuw Archief voor Wiskunde* 15, 15-22.

Landré, C.L. (1893) *Wiskundige hoofdstukken voor levensverzekering*, Utrecht.

Landré, C.L. (1895) *Mathematisch-technische Kapitel zur Lebensversicherung*, 1901 2nd ed., 1905 3rd ed., 1911 4th ed., Jena.

Landré, C.L. (1900a) 'Afronding van sterftecijfers zonder een wet van afsterving ten grondslag te leggen' (Graduation of mortality rates without founding it on a law of mortality), *Archief voor de Verzekerings-Wetenschap en Aanverwante Vakken* 4, 314-358.

Landré, C.L. (1900b) 'Afronding naar een nieuw beginsel' (Graduation to a new principle), *Archief voor de Verzekerings-Wetenschap en Aanverwante Vakken* 4, 471-502.

Landré, H.F. (1906) 'Corneille L. Landré door zijne dochter' (Landré by his daughter), in *Jaarboekje voor 1906 uitgegeven door de Vereeniging voor Levensverzekering*, 194-208.

Laplace, P.S., (1809) 'Mémoire sur les approximations des formules qui sont functions de très grands nombres et sur leur application aux probabilités', in: *Mémoires de l'Académie des Sciences de Paris*, 353-415, 559-565.

Legendre, A.M. (1805) *Novelles méthodes pour la détermination des orbites des comètes*, Paris.

Maistrov, L.E. (1974) *Probability theory. A historical sketch*, transl. and ed. by S. Kotz, New York and London.

Mounier, G.J.D. (1903) 'Iets over de grondslagen van de methode der kleinste kwadraten' (On the foundations of the method of least squares), *Archief voor de Verzekerings-Wetenschap en Aanverwante Vakken* 6, 1-43.

Mounier, G.J.D. (1906a) 'In memoriam. Corneille Louis Landré', *Archief voor de Verzekerings-Wetenschap en Aanverwante Vakken* 8, 227-247.

Mounier, G.J.D., (1906b) 'Veelvuldig voorkomende toepassingen van de methode der kleinste kwadraten' (Frequently occurring applications of the method of least squares), *Archief voor de Verzekerings-Wetenschap en Aanverwante Vakken* 8, 309-348.

Paraira, M.C. (1905-1907) 'Corneille Louis Landré (1838-1905)', *Nieuw Archief voor Wiskunde* 7 second series, 1-6.

Simpson, T. (1755) 'A letter to the right honourable George Earl of Macclesfield, president of the Royal Society, on the advantage of taking the mean of a

number of observations, in practical astronomy', *Philosophical Transactions* 49, 82-93.

Stigler, S.M. (1986) *The history of statistics: the measurement of uncertainty before 1900*, Cambridge Mass. and London.

Van Everdingen, E. (1953) *C.H.D. Buys Ballot 1817-1890*, The Hague.

Van Haaften, M. (1923) *Het wiskundig genootschap, zijn oudste geschiedenis, zijn werkzaamheden en zijn beteekenis voor het verzekeringswezen* (The mathematical society, its oldest history, its activities and its meaning for insurance matters), Groningen.

Van Lunteren, F. (1998) 'De oprichting van het Koninklijk Nederlands Meteorologisch Instituut: Humboldtiaanse wetenschap, internationale samewerking en praktisch nut' (The foundation of the Royal Dutch Meteoroelogical Institute: Humboldtian science, international cooperation and practical use), *Gewina* 21, 216-43.

Whewell, W. [1847] (1967) *The philosophy of inductive sciences, part two*, reprint of 2nd edition, London.

Whittaker, E. and G. Robinson [1924] (1944) *The calculus of observations. A treatise on numerical mathematics*, 4th edition, London and Glasgow.