Variability of monthly time fraction of excess of atmospheric propagation parameters

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Agus Mawira

Variability of Monthly
Time Fraction of Excess of
Atmospheric Propagation Parameters
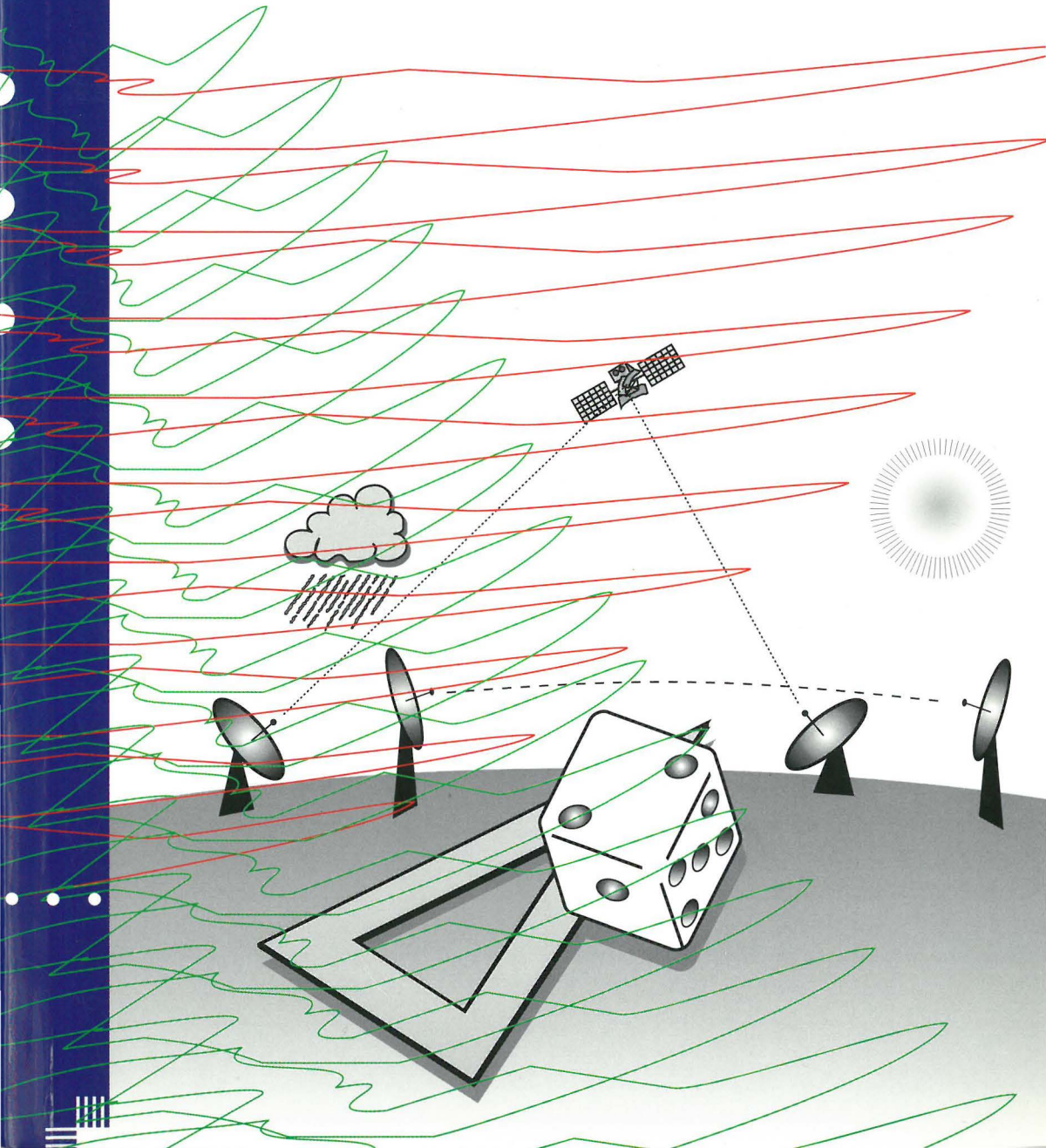
Variability of Monthly

Time Fraction of Excess of

Atmospheric Propagation Parameters

Variability of Monthly

Time Fraction of Excess of

Atmospheric Propagation Parameters

Agus Mawira

geboren te Jakarta

Dit proefschrift is goedgekeurd door de promotoren:

prof.dr.ir. G. Brussaard
en
prof. ir. A. C. van Bochove

To my mother and father

Summary

An important requirement of tropospheric propagation prediction models is their ability to predict the statistics of propagation parameters across a large probability range. Equally important to system planners is the ability of the model to quantify the variability of the statistical distribution of the propagation parameters when observed during shorter observation periods. The nature of such variability of radiowave propagation associated with atmospheric effects is the subject of study of this thesis.

In this thesis we consider the modelling of the variability of an important class of propagation parameters, namely the so-called monthly time fraction of excess (TFE). Such TFE is of great importance in the Grade-of-Service specifications of the ITU (International Telecommunication Union). The TFE of interest in this thesis are those associated with rain-attenuation on earth-satellite paths as well as those associated with ducting effects on terrestrial transhorizon paths. The work in this thesis is based on theoretical modelling as well as analysis of a very large amount of data on the monthly TFE. The data used are those collected by the European COST205 and COST210 projects and comprise a total of 92 site-years of observations obtained in some 9 different climatic zones of Europe.

The thesis starts by studying and analysing the general properties of the random process of the monthly time fraction of excess. In the first place a general theoretical framework is placed whereby the monthly TFE is defined as a random process. The ensemble associated with this random process is taken to be all the hypothetical radio links situated in a climatically homogeneous vicinity. All the links have equal parameter settings and configuration (frequency, path-length etc.). Subsequent analysis using the measurement

data shows that the random process of the TFE may, as a working assumption, be taken to be annually cyclo-stationary. Furthermore, the results of correlation analysis lead to the conclusion that for the most part the process may be assumed to be statistically independent. The usually complex seasonal dependence of propagation phenomena is approximated by the introduction of a dual-population model.

In the second part of the thesis the focus is on obtaining good (approximate) modelling of the statistical distribution of the monthly TFE. For this four alternative candidate models were studied as to their suitability. These models are the conditional exponential model (CEX), the log-normal model (CLN), the gamma model (GM) and the shifted gamma model (SGM). Various classical statistical tests were carried out with inconclusive results. An alternative novel method for statistical testing has therefore been devised. This method, which is based on normalisation and pooling of the data, produces conclusive result on the superiority of SGM and CEX models.

In the next step we investigate and show how the behaviour of the yearly TFE as well as the annual worst-month TFE can be modelled in a consistent and fundamental way using the found statistical model for the monthly TFE. For the worst-month analysis we show that the simplifying assumptions of dual-population would not lead to unduly large errors in the prediction of worst-month statistics. Further, we investigate the range of the expected variability and the dependence of this range on various values of the return periods. A climatic classification across Europe of the variability properties has also been performed (for slant-path rain-attenuation). The classifications result in four distinct regions, Scandinavia, North-West Europe, Central Europe, Alpine and Mediterranean.

In the last section we discuss various possible innovations to the current way of quantifying performance of radio links. We discuss and compare the methods of reliability engineering (e.g. the mean time between failure concept) and meteorology (e.g. return periods). Finally, as an example, we apply the found statistical model to an insurance type of analysis for a telecommunication system.

Contents

1. Introduction

1.1 The need for propagation modelling

Information on the propagation of radio waves is of primary importance in the development/design and deployment of radio communication systems. The role of propagation in the development of radio systems can be appreciated if we look at the very different techniques used in the various radio systems. For example, digital mobile (cellular) radio modems employ very sophisticated signal processing systems, while the signal processing system of geostationary satellite communication systems is relatively simple. This difference is directly related to the very different propagation modes encountered in these two applications.

In the mobile radio situation the mobile user is most of the time 'shadowed' by various objects such as buildings, trees, hills etc.. Here the transport of radio signal power from transmitter to receiver happens by way of many different radio paths (formed by reflections and/or diffraction). Each of such paths has a different path delay and contributes only a fraction of the total receiving power. In such a situation a transmitted digital pulse signal would result in a distorted signal reception (the received signal is the sum of many echoes of the original signal). The signal-processing unit in a digital mobile receiver is specifically designed to combat such an effect. One way to do this is to let the transmitter send, at regular intervals, test pulses in a known sequence. By analysing the received signal the receiver would then 'know' the propagation situation regarding the delays and can then, using such information, restore the distorted signal to its original state.

In a geostationary satellite communication system the transmitter and receiver are always in a line of sight situation with respect to each other so that only one propagation path is dominant in the transport of power. The specific problem associated with geostationary satellite commu-

nications is that the propagation loss is very high due to the long distances and possibly large attenuation caused by rainfall. In the early days of satellite communications (when satellite transmit power was very low) the ground station were equipped with very large antennas and the receivers had to be cooled by liquid nitrogen to allow detection of the very weak signals received.

In the procurement, the planning and the deployment of radio communication networks propagation information also plays an important role. While it is obvious that good information regarding the important parts of a communication network is indispensable for successfully carrying out the aforementioned exercises, radio propagation here has its own specific characteristics and problems. In the first place the behaviour of radio links is generally much more complex than that of guided transmission media. This regards not only the specific mechanisms (such as those discussed above) but also regards the statistics of occurrences of particular propagation phenomena that may vary from one location of the world to another. In the second place, there is a complicated interdependence of the propagation and the radio link parameters (e.g. frequency, length of radio hops, and heights of transmitter and receiver antennas). This fact actually gives the telecommunication providers the possibility to tailor their system to the required performance of the system, provided they have sufficient knowledge of the propagation.

Finally we note that there is an important difference between guided transmission media on the one hand and radio transmission media on the other regarding the validation of their conformity to the required grade of service criteria. In the case of guided media their transmission properties are stable and known, and the producers of the components will supply the necessary information required for planning of the system such as e.g. the expected loss per km of a cable. Furthermore, once a system based on guided media has been installed, the compliance of the whole system with the grade of service requirement can be established through limited series of test measurements. In the case of radio transmission systems their conformity to grade of service requirements cannot be easily established through test measurements on real systems. This is due to the seasonal, annual and multiyear variations of propagation conditions in the case of microwave and satellite communication systems, or the very complex behaviour of radio waves

due to interactions of natural and man-made structures over a large area in the case of cellular radio and broadcast radio.

From the above considerations it can be appreciated that good propagation models are vital when we want to build in a cost-effective and efficient manner good communication networks that include radio transmission systems. An important aspect of such models is their ability to predict the statistics across a large probability range. The nature of the statistics of propagation associated with atmospheric effects is the subject of study of this thesis. The scope of the study will be discussed in more details in Section 1.4. Before that, a brief history of the role of radio propagation is presented in Section 1.2. Section 1.3 discusses briefly the grade of service requirements of modern radio systems.

1.2 A short history of radio propagation

The early periods

In 1864 *Maxwell* formulated the theory of electromagnetic waves. With this theory the many properties of electricity and magnetism which had been discovered previously and explained by separate theories can be explained in a simple and unified way. Just as importantly *Maxwell* showed that the theory predicts the existence of a hitherto unknown entity, that of radio waves. The confirmation of this theory was provided by *Herz* in 1887 through a series of laboratory experiments [*Collins*, 1997].

Practical use of radio waves had to await the invention of radio by *Marconi* in 1895. With his equipment Marconi demonstrated that radio waves can be used as medium for transmitting and receiving information. In the early periods of his experiments Marconi used very short-wave radio (centimetre waves) but found that their range was restricted and subsequently switched to using long-wave radio. In 1901 he successfully transmitted Morse code across the Atlantic Ocean from Cornwall, U.K., to St. John's, Newfoundland, Canada, a distance of some 5500 km. The wavelength used was 366 meters (corresponding to a frequency of some 0.84 MHz) while

the transmit power was 15kW [*Delogne*, 1997]. Soon after this spectacular demonstration of the feasibility of radio wave communication, commercial services of radio telegraphy became commonplace using, in general, much longer wavelengths, longer than 4000 meters (frequencies lower than 100kHz), as it had been found that they were more efficient in long-distance transmissions.

To the designers and users of these early radio communication systems it quickly became apparent that the reliability of the communication was less than satisfactory. The main problem encountered was the strong variation in time and space of the quality of the radio links. Although some general patterns were noticed (at some period of the day signal transmission would, on average, be better than other periods of the day), some of the observed behaviour baffled the designers of the radio systems. For example, at times signals could successfully be transmitted across a very large distance but not to nearby receiving stations, while at other times the reverse would occur. This posed a problem to the service providers since on one hand the customers were demanding better reliability and on the other no one was able to pinpoint the cause of the problems. Knowledge concerning the propagation of radio waves and the structure of the atmosphere was too limited to allow the finding of a solution to this problem. For a while the providers concentrated on the expensive approach of developing larger and better antennas as well as developing more powerful transmitters.

At that time the radio community had no idea how such radio waves could travel so far beyond the visual horizon. By 1911 only an empirical formula, the so-called *Austin-Cohen* formula, was available by which the average dependence of signal strength with distance, wavelength, antenna heights and transmit power could be predicted with some accuracy [*Kerr*, 1951]. But there were no models available by which the above-mentioned variations in space and time could be understood. As we now know, at these frequencies two propagation modes are of importance. These are the ground-wave propagation mode and the ionospheric propagation mode.

In the ground-wave propagation mode the radio waves are able to follow the Earth curvature through a process called spherical diffraction, thereby allowing communications beyond the visual horizon. This effect is strongest in the frequency bands below 100kHz and is a stable

effect. The range allowed by this model, however, falls far short of the observed ranges. The second propagation mode is that related to the effects of the ionosphere, the ionised part of the atmosphere located at altitudes between 70km and 400km. Within this altitude range various layers with different density of ionised particles can be found (the layers are usually indicated as the E layer, the F layer, the G layers etc.). Due to such layering a bending of the propagation trajectory of the radio signals back to earth can occur. This "reflection" effect which occurs mostly at frequencies below 40 MHz can be utilised to allow very long distance propagation. However the propagation mechanism of the ionosphere is unstable. In the first place this is due to the fact that the electrical properties of the ionosphere are strongly time dependent. The main sources for the variability of the ionosphere are the variation of the sun's radiation (which is responsible for the ionisation process), and the rotation of the earth. Furthermore the ion density shows very complex variability which is related to atmospheric flow etc.. In the second place multiple propagation paths may be produced each with quite different propagation path lengths (or phases); the summation of the signal components associated with each of the individual paths may then lead to destructive and constructive interference. Small variations in path length in space and time would then produce very complex patterns of signal distribution.

A breakthrough was realised in 1913 when *Lee de Forest* (the famous inventor of the triode) published the results of his analysis done for the Federal Telegraph Company [*Delogne*, 1997]. In this work he analysed the observed behaviour of many radio transmission signals. He noted that two signals from the same transmitter but with slightly different wavelength may show completely opposing behaviour. At one moment one signal would offer good reception while the other would be totally extinguished while at some other moment the reverse would be true. He concluded that this behaviour could only arise if the propagation should consist of at least two propagation paths with sufficiently large path length difference leading to constructive and destructive interference (multipath effect). He postulated that the first path is that corresponding to the ground wave propagation and that the second path must correspond to a path produced when an atmospheric layer at high altitude reflects the radiowaves. His further analysis led him to the (right) conclusion that this reflection may be due to partially ionised masses of air at great heights. Although ionisation of the upper atmosphere and its capacity of bending radio waves had at that time already been postulated or speculated upon by scientists such as *Heaviside*,

Eccles and *Pierce* it was *de Forest* who gave convincing empirical evidence of this phenomenon [*Delogne*, 1997].

On the theoretical side, noteworthy work was done by *Sommerfeld* in 1911 concerning the propagation above a plane structure. In 1908 *Mie* derived a series expression regarding the influence of a sphere on the radio wave propagation; his expression however can only be evaluated numerically for spheres of small radius. The real theoretical breakthrough was achieved by *Watson* in 1918 [*Bremmer*, 1949]. *Watson* was able to solve the difficult mathematics involved in evaluating the propagation of radio wave on a very large smooth sphere (with and without a concentric reflecting layer). The analysis of the smooth sphere without a concentric reflecting layer gave convincing proof that ground-wave propagation cannot account for the long propagation ranges that had been observed. The analysis of the model with a concentric reflecting layer leads to the *Austin-Cohen* empirical formula [*Kerr*, 1951]. Noteworthy are also the theoretical works of Bremmer and van der Pol of Philips Research Laboratory in Eindhoven [*Bremmer*, 1949].

With this new knowledge, service providers were able to improve the reliability of their transmission. One method was to make use of the above-mentioned multipath phenomenon, that is to say to pick a suitable frequency for the transmission. Needless to say that in order to successfully use such a strategy, a good propagation model as well as accurate data concerning the ionosphere is required. A lesson learned here by designers of radio links is that knowledge of the propagation of the radio waves and their underlying causes is essential for successful design and deployment of radio links. In the following years much work was then performed to characterise the ionospheric propagation. Major measurement campaigns were set up to obtain information regarding the structure of the ionosphere; atmospheric physicists as well as radio engineers were involved in these studies. These measurement campaigns were carried out in international co-operation under guidance of e.g. the International Union of Radio Science (URSI) and the International Telecommunication Union (ITU). These measurements have resulted in major advances in the understanding of atmospheric physics as well as in providing important data and models for the radio communications providers. Models and data for predicting ionospheric propagation can be found in the ITU Recommendations [*ITU*, 1997a] and

Handbooks of HF propagation of the ITU [*ITU*, 1997b].

Since the introduction of long distance communication by trans-oceanic cables in the 1950's, and by satellite in the 1960's, communication through ionospheric propagation has lost much of its importance to the major communication systems. Nevertheless it has remained important in applications in the realm of the defence agencies and in some other specialised applications such as that of navigation.Furthermore, better and more detailed propagation information and modelling of ionospheric propagation are still wanted for improving the accuracy of the Global Positioning Satellite (GPS) navigation systems.

The period around the Second World War

During the Second World War and thereafter, rapid development in electronics allowed the utilisation of the higher frequency bands, which provided for more traffic capacity. At first the VHF (30MHz-300MHz) and UHF (300MHz - 2000MHz) bands were utilised, primarily for broadcasting, land mobile radio and for point-to-point communications, while some radar systems used in the second world war already utilised microwaves (wavelength of some 3cm or frequency about 10GHz) [*Kerr*, 1951]. At these frequencies the ionosphere does not play any significant role in the propagation and the main causes of impairments of the quality of radio signal are the troposphere or various objects such as hills, trees and buildings.

In broadcasting and mobile radio applications the propagation problems encountered are related to the blockage by objects and hills, although at the frequencies employed mostly (the lower VHF band) these effects are less serious. In 1947 Bullington presented a simple model to predict the field strength in mobile radio applications. This model was inspired by the theoretical plane earth model. The broadcasters developed their own empirical prediction method, which has been incorporated in Recommendation P.370 of the ITU. Curiously, although the problem at hand is similar for both applications (the same frequency band, the same geometrical configuration with one high station serving many users with low antenna heights), until recently hardly any co-ordination of activities took place.

The military encountered serious propagation problems in the use of their radar systems. These problems are related to anomalous vertical distributions of the refractive index of the atmosphere caused by anomalous humidity and/or anomalous temperature gradients. Under such meteorological conditions a stratified layer of atmosphere can exist in which the refractive index decreases very rapidly with height (the so called *surface ducts*) so that radiation of sufficiently short wavelength may be trapped in the layer and guided around the curved surface of the earth. The transmission problems encountered were akin to that encountered by early radio users: varying radio ranges, multipath effects etc.. Since these effects occur strongly in humid and hot areas (such as the Pacific basin where many important naval engagements took place) the U.S.A. invested much effort in developing the required propagation prediction models during the Second World War [Kerr, 1951].

Already during this period some pioneering work on rain-induced attenuation of microwave signals had been done. The attenuation effects of fog and rain on centimetre radio waves have already been treated as early as 1930 by *J.A. Stratton* [Ishimaru, 1972]. Other milestones of this period include the work of *Laws & Parsons* in 1947 who characterised the raindrop size distribution, thus allowing the calculation of the specific attenuation of rain of a given intensity, which was performed by *Ryde & Ryde* soon after that [Kerr, 1951] using the spherical model for the raindrops.

Finally we note that during the Second World War two famous statistical distributions were formulated describing the general behaviour of the amplitude of a radio wave composed of many varying independent terms. The first is the *Rice* distribution developed in the U.S.A., the second is the the *Nakagami* distribution developed in Japan. Both distributions represent generalisations of the *Rayleigh* statistical distribution, which was derived in the 19th century.

The modern era

From the 60's onwards, higher centimetre bands were utilised for so called microwave links and geostationary satellite links. The highest frequency used currently in satellite communications is 30 GHz, while microwave links operating in the 60 GHz band can now already be purchased. Mobile and terrestrial broadcast systems more and more are using the UHF frequency band.

The main propagation factors affecting terrestrial microwave and geostationary satellite systems are those related to the troposphere. Generally two categories of effects are discerned here according to the specific atmospheric process influencing the propagation. The first are the precipitation effects (effects of rain, snow or hail), the second are the effects associated with anomalous refractivity distributions of the atmosphere, the so-called clear-air effects. The main transmission effects associated with precipitation are attenuation (loss of signal power) and depolarisation (deformation of the polarisation state of the radio wave). Clear-air effects may produce (multipath) fast fading or distortion of digital signals, but may also cause excessive long-distance propagation leading to interference between systems utilising the same frequencies.

As mentioned above, some fundamental work on rain attenuation had already been done during the period around the Second World War. While such work produced important new insight in the propagation mechanisms, much work still had to be done before propagation models suitable for radio-planning purposes were obtained. In the first place, better modelling of the fine structure of rain had to be obtained. Some of the improvements are those with regard to the shape of the raindrops (which is not spherical as had been assumed before) the distribution of raindrop size (which as it turned out depends on the type of rain). The other improvement that had to be developed is with regard to new numerical calculation methods that allow the evaluation of propagation properties of non-spherical drops. Furthermore information is required on the horizontal and vertical extent of rain, the probability of occurrence of intense rainfall, their geographical dependence and distribution. To obtain such data, extensive long-term measurements on rainfall intensity and rain effects across a large frequency range had to be performed. Such activities required large-scale co-operation, since no single research institute

can bear the cost of maintaining experiments across many locations during a long period. During the 70's, 80's and early 90's many propagation campaigns were carried out, mainly in Europe, U.S.A. and Japan (in the late 80's INTELSAT and COMSAT started performing measurements in less accessible areas such as Africa, South America and South East Asia). In Western Europe, co-ordinated studies were carried out under guidance of the European Commission, the so called COST projects (specifically projects COST25/4 and COST205) and the European Space Agency (ESA). In the U.S.A. the research was co-ordinated by NASA. In Japan the Ministry of Post and Communications carried out national co-ordination. Satellites equipped with propagation beacons were put into orbit: SIRIO, OTS, OLYMPUS and ITALSAT for Europe, BSE and ETS for Japan, ATS, COMSTAR and ACTS for the U.S.A.. On a global scale the co-ordination of model developments is carried out by the ITU within Study Group 3 of the Radiocommunication Sector.

Other propagation effects which were investigated extensively in this period are those due to anomalies of the refractive index distribution of the atmosphere, such as the aforementioned *surface ducts* as well as *elevated ducts* (stratified anomalous refractive index gradient at higher altitudes which may reflect microwaves). Their occurrence may cause severe fast fading on narrowband microwave links and/or severe distortions in broadband microwave systems. They can also lead to anomalous long-distance propagation, which may lead to interference between systems, e.g. in the different countries that use the same frequency. This effect is especially troublesome to satellite communication systems since they operate in the same frequency band as the terrestrial microwave systems. Since the signals received by a satellite groundstation are very weak, they are especially sensitive to interference.

In some special cases the propagation due to particular stratification of the atmospheric refractive index can be solved mathematically. In the more complex cases ray-tracing techniques may be employed successfully. However, in real life situations these approaches were found to be non-productive. The problem is related to the fact that the distributions of refractive index profiles that produce the troublesome propagation do not allow simple description; many varying shapes of the profile can be encountered with a probability of occurrence that is strongly dependent on location and other environmental parameters. Therefore, the propagation researchers attacked the

problem on two distinct fronts. In the first place a primarily empirical modelling is pursued to obtain a means for predicting the statistical distribution of narrowband fading and the anomalously long distance propagation; in Europe this activity was co-ordinated within the COST230 project. In the second place separate models are developed theoretically and empirically for the proper understanding of dispersive properties of the broadband channel.

The attention to the problem of propagation in the field of mobile and terrestrial broadcasting was at first not so great. Rather restricted empirical models were produced by individual researchers, see [*Jakes*, 1974] (e.g. *Bullington* in 1947, *Egli* in 1957 and *Okumura* in 1965). *Okumura*'s model as modified by *Hata* is still currently used when high accuracy is not required. This model was derived from extensive measurements carried out in Japan in the frequency range from 150MHz to 2GHz. This model allowing the prediction of the average propagation effects in various types of regions (e.g. urban, suburban, rural) was presented by *Okumura* et.al. [1965]. The advent of digital systems and the very large expansion of cellular communications systems make, currently, the development of better models imperative. In Europe a successful co-ordinated effort, within the COST207 project, was made in the late 80's to model the broadband propagation aspects with sufficient accuracy to allow the specification of a new pan-European land-mobile cellular system: GSM. At the moment much work is carried out across the world for the development of a fundamental prediction model which would allow accurate assessment of e.g. coverage and interference from data bases giving detailed information regarding the environment (buildings, trees etc.).

The previous examples highlight the fact that the influences of the environment on radio communication systems are extremely complex. The magnitude of this complexity is such that the models for predicting the propagation effects will usually be statistical in nature. Furthermore, since high quality radio links are designed with stringent outage allowances (e.g. 0.01% of the time) the propagation models should perform well across a large probability range.

1.3. Grade of services

In the final analysis the goal of all service providers is to make profit and expand their market. To do this the service provider has to obtain customer satisfaction while at the same time minimising his cost. There are many aspects to obtaining customer satisfaction (e.g. novel services, quality, and knowledge of customer need, cost aspects); here we shall restrict our attention to those related to transmission quality (or also the "grade of service"). In the ITU domain many specifications regarding required grade of service have been defined. There is a broad range of such requirements, indicated as performance and availability criteria, covering the different transmission media (e.g. cable, radio-relay) and services (e.g. voice, data). The main goal behind the formulation of these international grade of service standards is to ensure some kind of minimum quality of international telecommunication links (which might be composed of many different transmission segments).

The ITU set of grade of service requirements is very extensive and caters for many different applications. Separate standards are given for high, medium and low-grade systems, for analog and digital etc.). For point to point communication links the structure of these standards is roughly the same, they are composed of:

1. Quality criterion: this is a measure of the status of the link. Two broad categories are used namely 'Unavailability' and 'Performance'. Unavailability is used to indicate the case where transmission quality has become so bad that the link is considered useless for communications. Performance indicates the quality of an available link. For digital links Performance is measured using several criteria: 'Degraded Minutes', 'Severely Errored Seconds', 'Errored Seconds'.

2. Reference Period: this is the period within which the statistics of the above quality criterion is evaluated. Two different reference periods are used: the 'any month' reference period and the 'long term' reference period.

3. Allowance Objective: this is the fractional probability that is allocated to the various quality criteria within the reference period.

4. Length of a Reference Link: different reference distances have been defined varying from several hundred kilometres to tens of thousands of kilometres.

The definition of the Performance and Unavailability parameters for analogue systems is rather complicated. Here we give as example the, somewhat more simple, definitions related to digital systems:

Degraded Minutes (DM)	a period of 1 minute is classified as degraded when the Bit Error Ratio (BER) is > 10^{-6}.
Severely Errored Seconds (SES)	a period of 1 second is classified as severely errored if the BER is larger than 10^{-3}.
Errored Seconds (ES)	a period of 1 second is classified as errored if at least 1 bit is wrongly received.
Unavailability	the period of unavailable time starts if during a 10 second period all the 10 x 1 second periods are SES (or that the link is totally interrupted, e.g. loss of timing). The unavailable period terminates when a 10-second block is encountered with each 10 x 1 second periods having a BER better than 10^{-3} (and obviously timing etc. has been restored).

The table below gives an example of the grade of service requirement for a terrestrial high-grade Hypothetical Reference Digital Path (HRDP) of 2500km taken from Recommendation 594 of ITU-R.

Quality Criterion	**Allowance Objective**
Degraded Minutes	0.4% of Any month
Severely Errored Seconds	0.054% of Any Month
Errored Seconds	0.32% of Any month
Unavailability	0.3% of the Long-term period

Note: when planning a real link composed of many short paths, the Allowance Objective is distributed among the individual paths. Since the length of a single microwave link is usually much shorter than 50km we can see that the Allowance Objective of a single link can be very stringent (e.g. < 0.006% of the long term period).

At first sight these grade of service objectives place a very stringent requirement on the degree of details that must be given by the propagation models. Fortunately, for the characteristics of interest in microwave and satellite communications some simplification can be made. In the case of rain and clear-air induced signal variations, a separation can be made of fine-scale effects and large-scale effects. It has been observed that fine-scale effects (rapid signal variations) follow some generic stochastic process laws (e.g. Rayleigh, Rice, $1/f$ law etc.). The large scale effects (relatively slow signal variations representing usually a loss of power due to e.g. absorption) do not in general obey any simple generic stochastic process laws and have to be modelled empirically. In the calculation of grade of service parameters, two steps are usually employed. First a model of fine-scale phenomena in conjunction with a model of the radio modulation system employed is used to determine (mathematically or using simulations) at which power loss level a SES, DM etc. situation is reached. In the second step the aforementioned power loss level is then translated to the relevant value of the (large-scale) propagation parameter. Subsequently the (statistical) propagation model for this parameter gives the predicted time percentage of excess in the required reference period. We note here that for a given radio modulation system the first step needs only to be done once for each propagation phenomenon and needs not to be redone for each location and link configuration. This step is usually done by the producer of the radio modem, the required 'slow' fading margin etc. forms then part of the equipment specification. The second step requires local meteorological information and the link configuration information. The planner of the radio network is usually responsible for finding the required information.

An aspect of the ITU grade of service requirement, which has strongly conditioned propagation research, is the definition of the reference periods ("Long-term" and "Any-month"). Most prediction models have been tuned to produce the statistical distributions for these two reference periods. More discussion on these statistics is given in Section 1.4.

Finally we note that many of the ITU requirements have undergone a long historical evolution and most certainly the developer of these requirements had the best interest of the user in mind when formulating these requirements and also took into account some techno-economical considerations. However, the environment within which these requirements were formulated was a quite a different one than the current environment. In the past, the national PTT's were the sole provider of telecommunication services. In such a situation the service providers do not have to deal with the large differentiation regarding services to be delivered that would take place in a free market place.

The old situation had the advantage that it was much easier to obtain international standards. Furthermore, the PTT's had only to pursue a limited set of objectives and a sense of security is obtained: once the grade of service requirements had been established, the PTT's could build their system without worrying too much whether it would really satisfy the customer's needs. This situation also conditioned the propagation research. The focus was to produce models that perform satisfactorily in the range dictated by the grade of service requirements. For example, the rain attenuation prediction methods were optimised for the prediction of attenuation with average occurrence of 0.01% (99.99% being one of the grade of service targets in satellite communication).

The new situation is much more complex, it is more difficult to obtain international standards, and furthermore the service providers must have a better insight in what a chosen grade of service requirement would mean from the point of view of the users. In this new situation, the propagation models must have a validity range that is much larger than previously. This poses somewhat of a problem, since development of propagation models takes a long time. One of the purposes of our study is to present a method by which the statistical domain of current models can be extended.

1.4. Scope of study

As noted previously, satellite and microwave radio communication links are susceptible to propagation effects induced by changes in atmospheric/weather conditions. For example, during heavy thundershowers the received signals of such links may be attenuated considerably, sometimes even to the point where communication is completely cut off. Another example is that during the presence of severe atmospheric temperature inversions a microwave radio signal can travel much further than normally, even very far behind the horizon, and may cause unintended interference at another microwave link. A good overview of (atmospheric) propagation effects can be found in [*Hall*, 1979], [*Boithias*, 1987], [*Brussaard & Watson*, 1994].

In general, propagation effects associated with atmospheric processes show such a complicated behaviour that they must be treated as random processes for which only some of the statistical features can be determined with sufficient accuracy. Many efforts in the past were directed to the determination of the so-called 'long-term' statistical distribution P_a:

$$P_a(a) = probability(\boldsymbol{a} \geq a) \qquad (1.1)$$

where $\boldsymbol{a}$ denotes the propagation random variate (e.g. attenuation) while a is a threshold value of interest (throughout this thesis random variates are indicated by bold symbols). Usually P_a depends on some link parameters (such as frequency, path length) and some meteorological parameters. The International Telecommunication Union (ITU) gives methods and data for estimating P_a, for various propagation effects, [*ITU*, 1997].

As it turned out P_a alone is not a sufficient description of the propagation process for the purposes of the planners. This is due to the fact that the time variation of propagation parameters is very complicated (ranging from rapid variation to seasonal variation). During shorter periods the user of the radio link may experience much more detrimental propagation effects than what at first sight may be expected from P_a. This is certainly the case for the monthly observation period that is used by the ITU as one of the reference periods for defining grade of service. The

observed values of the monthly distributions, i.e. the monthly time fraction of excess $X_{i,j}$:

$$\begin{aligned} X_{i,j} &= \text{fraction of time that } \boldsymbol{a} \geq a \text{ within month nr. } i \text{ of year nr. } j \\ &(i = 1, 2, \ldots\ldots, 12; \quad j = \ldots, 1981, 1982, 1983, \ldots\ldots) \end{aligned} \qquad (1.2)$$

will in many cases deviate considerably from P_a. This is illustrated by Figure 1.1 which gives examples of the behaviour of $X_{i,j}$ as was observed in Europe.

Therefore, in addition to P_a the ITU has adopted the average annual worst-month distribution $P_{a|WM}$, first proposed by *Brussaard & Watson* [1979], as a basis for defining grade of service parameters, see also Recommendation ITU-R P581-2. $P_{a|WM}$ is basically a conditional statistical distribution:

$$P_{a|WM}(a) = probability(\boldsymbol{a} \geq a \mid \text{the annual worst months periods}) \qquad (1.3)$$

The annual-worst-month period in a given year is defined as the month having the highest value for the monthly time fraction of excess. This selection is dependent on the threshold value a.
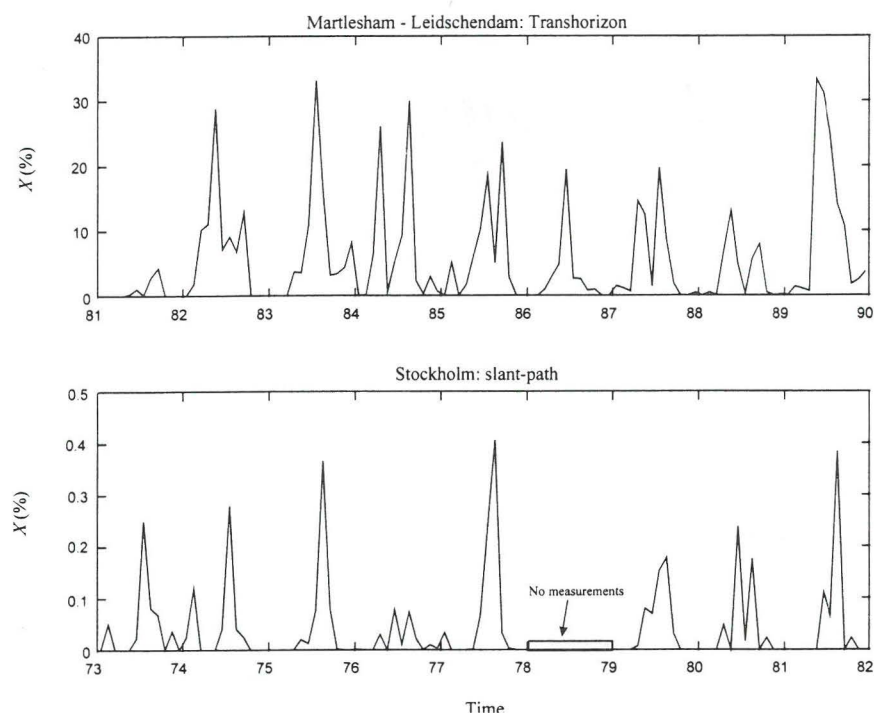
Figure 1.1 Examples of observed behaviour of the monthly time fraction of excess X_{ij} The threshold for the Trans-horizon link is -35 dB with respect to free space level. The threshold for the slant-path is 2 dB excess attenuation.

The usage of $P_{a|WM}$ in the last decennium has served the radio communication system planners well (see e.g. ITU-R Recommendation P.841 [*ITU*, 1997] for methods and data for estimating $P_{a|WM}$). However, the recent changes in the telecommunication market environment have created the need for a broader range of novel grade of service concepts (e.g. the concept of 'risk' discussed by *Fukuchi & Watson* [1989], *Watson* [1990], *Crane* [1991], *Crane* [1993], *Brussaard & Mawira* [1993]). This in turn will create a demand for a large range of new statistical information. The problem facing the propagation community now is how to efficiently produce such new information. Ideally, a complete stochastic model for the propagation random variate should be developed. From such a model the required statistical information can then be derived, either analytically or through simulation. Unfortunately, the development of such a stochastic model will take a lot of time and money.

The more modest approach considered in this thesis is to model the ensemble of $X_{i,j}$'s as a random process and develop good statistical models describing this random process. Such a good model, in combination with available models and data for determining P_a and $P_{a|WM}$ can then be used to derive various types of statistical information, provided that the reference period is not taken to be shorter than one month.

The idea of considering the ensemble of monthly time fractions of excess as a random process was first proposed by *Crane & DeBrunner* [1978]. They also noted that the statistics of the X_{ij}'s obtained from 2 years of microwave rain-attenuation measurements in Switzerland can be modelled well by the so-called conditional exponential distribution. Subsequent work by *Brussaard & Watson* [1979], *Mawira* [1980], *Dintelmann* [1984], *Mawira* [1984], *Fukuchi et al.* [1985], *Dellagiacoma & Tarducci* [1987], *Poiares Baptista et.al.* [1989] seems to support the usage of the conditional exponential distribution for rain-intensity and rain induced propagation effects. Alternatively *Crane* [1991] has suggested that the conditional log-normal model is the most suitable one.

A shortcoming of all of these studies is that, with the exception of rainfall-intensity data [*Mawira*, 1980] and [*Dellagiacomma & Tarducci*, 1987], the data used in each of the analysis were not numerous (at most 4 site-years of observations). Moreover, with the exception of the

study by *Crane* [1991], no alternative models were considered and the tests performed were more graphical and intuitive than precise.

In this thesis we therefore consider again the question of modelling of the statistics of the monthly time fraction of excess X_{ij}. We also consider the applications of this model. The matters to be addressed and solved in this thesis are the following:

1. A proper formal conceptual framework for modelling needs to be produced. This includes the general characterisation of the random process especially with regard to stationarity, independence, seasonal dependence and homogeneity of its population. Until now this matter has only been dealt with in a less formal and more intuitive manner.

2. Determination of the most suitable model for the statistical distribution of the $X_{i,j}$s must be done. In addition to the conditional exponential and the conditional log-normal models gamma alike models shall also be considered. Careful consideration has to be given with regard to statistical testing and inference methods.

3. The resulting model and theoretical framework will be applied to derive models for prediction of the variability of time fraction of excess related to other base observation times (such as worst-month and yearly times). Formerly the characteristics of the variability of these other TFEs were determined in a non-unified and purely empirical manner.

4. With respect to worst-month we need to investigate what the effect of the inevitable simplifying assumptions on the models for X_{ij}. will have on the accuracy of worst-month prediction. Another worst-month aspect looked into is that with respect to the effects of limited observation time on the accuracy of the ITU prediction method.

5. As an extension to current set grade of service criteria we shall consider various concepts used in reliability engineering. We shall demonstrate the possible usage of the above mentioned models in so-called risk or reliability analysis.

The analysis in this thesis is makes use of a very large amount of data on the X_{ij}'s with respect to diverse propagation processes (slant-path rain attenuation effects and ducting effects on trans-horizon paths). The data used are those collected by the European COST205 and COST210 projects and compose in total of 92 site years of observations obtained in some 9 different climatic zones of Europe (see [*COST205*, 1985], [*Fedi*, 1985] and [*COST210*, 1991]). See Figures 1.2 and 1.3. Other more limited sources of data have also been included. These are 5 year ducting data from KPN Research and processed results of 7 years of rain-rate data (also from KPN Research).

1.5 Outline of the thesis

In Section 2 the thesis studies the properties of the random process of the monthly time fraction of excess (TFE): the X_{ij}'s discussed previously in Section 1.4. The statistical characterisation of the monthly time fraction of such TFE's, is treated whereby various aspects of this random process are considered (independence, ergodicity, seasonality etc.). Particularly the focus is on obtaining good (approximate) modelling of the statistical distribution of the X_{ij}'s. For this purpose the merits of four alternative candidates are be studied and compared to each other. Attention is also given to the development of a novel method for statistical testing.

In Section 3 we analyse the behaviour of the yearly time fraction of excess (such time fraction of excess is often used as the minimum observation period in propagation experiments). Here we investigate the year to year range of variability. In particular a model is developed for predicting the year to year variability based on the results of the modelling performed in Section 2. In Section 4 we analyse the behaviour of the annual worst-month time fraction of excess (this time fraction of excess forms the basis of many of ITU's Grade of Service criteria). Again we shall endeavour to apply the model found in Section 2 for the analysis of the worst-month statistics. The behaviour is discussed of an important related parameter, the "worst month quotient", used extensively by ITU in prediction of the worst-month statistics. In Section 5, the application of variability analysis to the definition of risk and return period is discussed. Also examples are

given of a more practical nature of application of the models developed (in a so-called 'risk' and 'reliability' analysis of communication systems). Finally, in Section 6 the main conclusions are drawn.
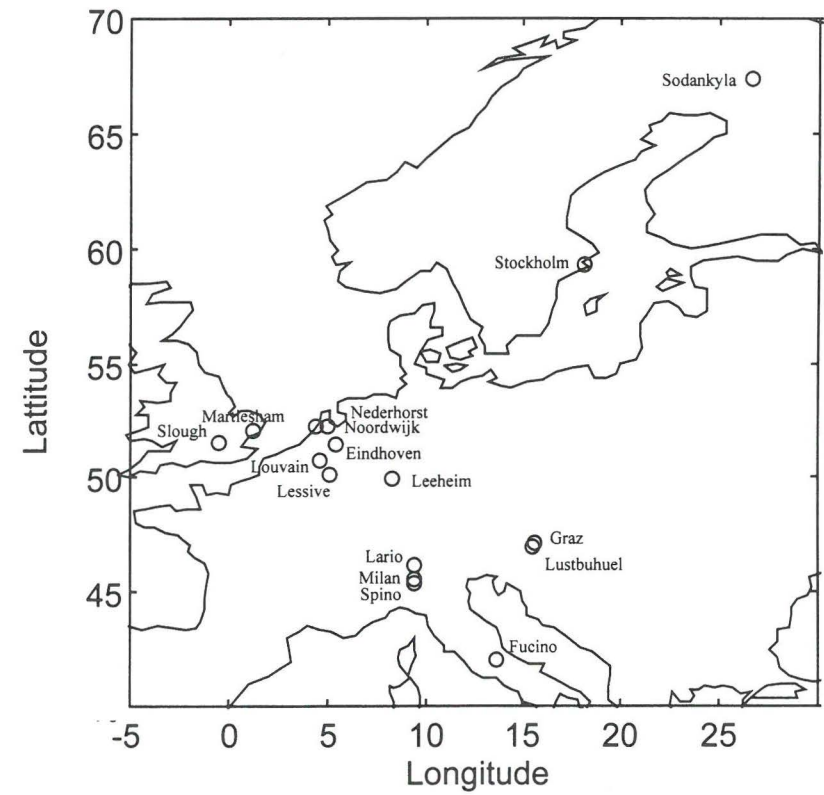

Figure 1.2 Map of those COST205 measurement sites whose data is used for analysis in this thesis.
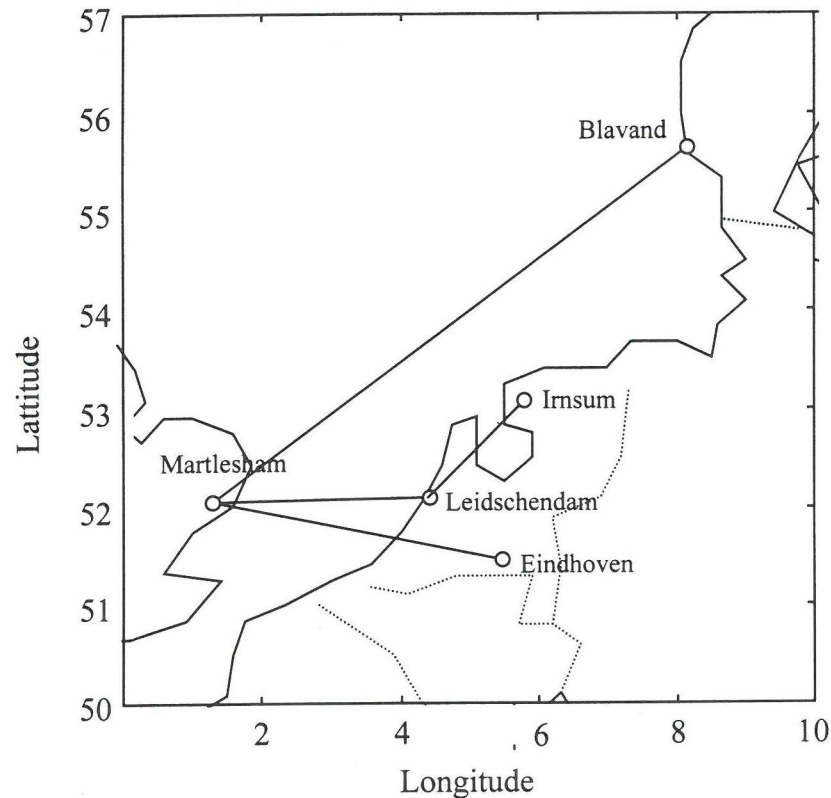
Figure 1.3 Map of the trans-horizon links where ducting measurements have been carried out. All data are from the COST210 project.

2. The random process X of the monthly time fraction of excess

2.1 General properties

As noted in the introduction a simplified but useful way to quantify the performance of a radio link is to evaluate the behaviour of the fractional duration in each month that a (propagation) parameter a, relevant to performance degradation, exceeds a given threshold value a. The discrete time series of such fractional durations, i.e. the monthly time fraction of excess (TFE) $X_{m,j}$:

$$X_{m,j}(a) = \text{\textit{fraction of time that } } a \geq a \text{ \textit{within month nr. m of year nr. j}}$$
$$(m = 1, 2, \ldots., 12; j = \ldots., 1981, 1982, \ldots.) \tag{2.1}$$

usually exhibits a very complex behaviour and can be looked at as a random process, viz.:

"a process running in time and controlled by probabilistic laws" [*Doob*, 1953].

Basically the goal of our study is to find these probabilistic laws or at least find suitable approximations to them.

In mathematical statistics, probability is defined with respect to an *ensemble* of possible outcomes; in the cases of a random process the ensemble is an (infinite) family of time series. This is a rather simplified description of the mathematical-statistical approach; for a more exact description see for example [*Doob*, 1953], [*Papoulis*, 1984]. To most people this approach seems very artificial; nevertheless, over time this approach has proven its usefulness; it gives a precise

framework for a statistical analysis of complex processes. Following this approach the (discrete) time series of the monthly TFEs, $X_{m,j}$, associated with a particular radio link, can be considered to be one member (or one '*realisation*') from a discrete-time random process X. Note that a separate random process X is defined for each threshold level a (the dependence on a is implied in the following analysis).

From a physical point of view we might consider this random process X as representing the family (ensemble) of time dependent outcomes of all possible (hypothetical) radio links with equal link parameters (e.g. frequency, polarisation, pathlength) located in a specific geographical zone. The area of such a geographical zone is assumed to be large enough to accommodate many radio links but small enough to ensure homogeneous climatic properties. There may be mathematical-statistical objections to this interpretation, but in any case such an interpretation is in accordance with the more operational aspect of the usage of statistical data in empirical science. Here we assume that ergodicity applies, i.e. that if we observe a radio-link long enough and derive its statistical properties, then these properties are valid as well for 'similar' links in the 'same' climatic zone.

To identify what constitutes a 'similar' link is usually reasonably simple once we know some of the fundamental aspects of the propagation, e.g. once we know the frequency dependence of the relationship between specific attenuation with rain intensity we can estimate which frequency bands must have equal propagation characteristics. We can even use this information to scale the statistics from one frequency band to another. From this example we can surmise that fundamental modelling of the propagation process, while not the object of our study, is an essential component even of statistical prediction models. The identification of the climatic zone is a more difficult operation, although as a first approach we can derive radiometeorological classification from the meteorological classification given by Meteorological Offices. As in the case with meteorological effects, many atmospheric propagation effects exhibit seasonal dependence. This is illustrated by Figures 2.1 and 2.2, which show examples of the observed dependence of the 84, 50 and 16 percentile values of X within each calendar month.
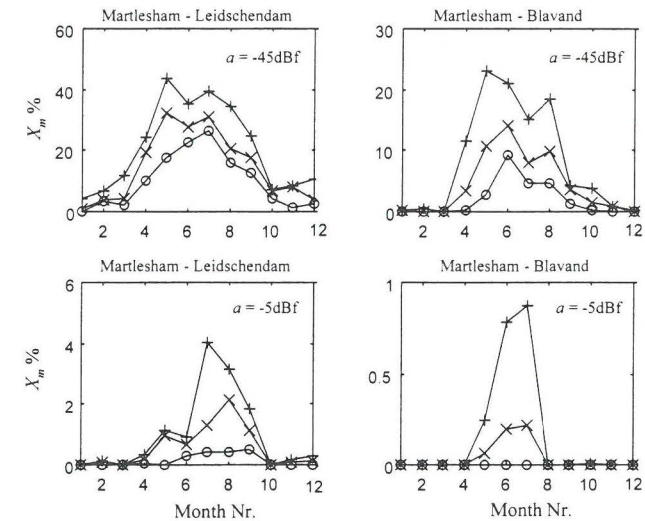
Figure 2.1 Examples of observed seasonal dependence of ducting effects. The x markers indicate the median value. The 84 & 16 percentile points are indicated by the o and the + markers. dBf = dB with respect to free space level.

That is to say that for each calendar month the statistical distribution $P_{X|m}$ is determined, subsequently from this distribution the values of X associated for $P_{X|m} = 0.84$, 0.5 and 0.16 were determined. Figure 2.1 shows ducting effects and was derived from trans-horizon measurements on two separate links (Martlesham - Leidschendam and Martlesham - Blavand). Figure 2.2 shows effects of attenuation due to rain on slant paths as was observed on two different European Locations (Stockholm and Northern Italy). These figures show that, at least at the higher threshold levels, a definite concentration of propagation activities can be observed around the summer period. This is of course related to the underlying meteorological processes. In Western Europe most of the intense rainfall occurs around the summer, and also the stable high temperatures and high pressures needed to produce atmospheric temperature inversion layers occur mostly in the summer. The seasonal behaviour in other countries of these and other propagation parameters may exhibit different patterns, e.g. fading due to snow in winters.
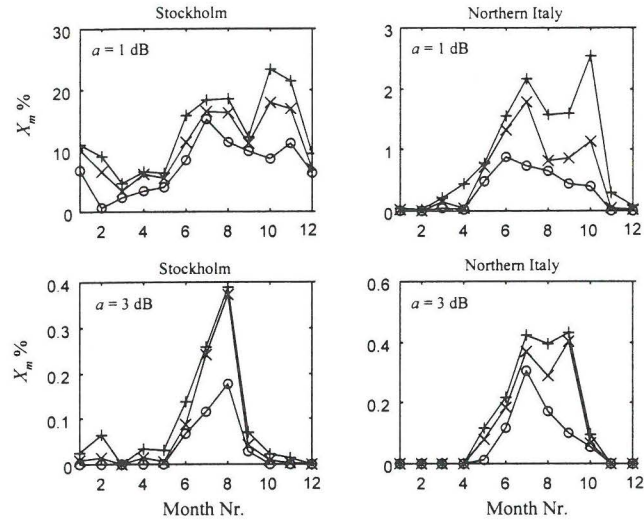
Figure 2.2 Examples of seasonal dependence of rain-attenuation

effects. The x markers indicate the median values of *X*. The 84

and 16 percentile points are indicated by the o and the + markers.

Another way of showing such repeated patterns of behaviour is through the normalised

autocorrelation function ρ_X. For of a given time-series of observation of *X*:

$$X(t_1), X(t_2), \ldots\ldots\ldots\ldots, X(t_N) \tag{2.2}$$

where t_1 = the 1st month of the 1st year of observation, t_2 = the second month in the 1st year of

observation, $\ldots\ldots$, t_N = the 12th month in the last year of observation. The estimator r_X for this

autocorrelation function is usually determined as follows:

$$r_X(\tau) = \frac{1}{s_1 s_2} \frac{1}{N-\tau-1} \sum_{n=1}^{N-\tau} (X(t_n) - \mu_1)(X(t_n + \tau) - \mu_2) \tag{2.3}$$

where the μ's represent the sample averages:

$$\mu_1 = \frac{1}{N-\tau} \sum_{n=1}^{N-\tau} X(t_n), \qquad \mu_2 = \frac{1}{N-\tau} \sum_{n=1}^{N-\tau} X(t_n + \tau), \tag{2.4}$$

and s_1 and s_2 represent the sample standard deviations:

$$s_1^2 = \frac{1}{N-\tau-1} \sum_{n=1}^{N-\tau} (X(t_n) - \mu_1)^2, \qquad s_2^2 = \frac{1}{N-\tau-1} \sum_{n=1}^{N-\tau} (X(t_n + \tau) - \mu_2)^2 \tag{2.5}$$

Some authors use an alternative expression for calculating r_X; see e.g. section 2 of [*Box &*

Jenkins, 1970]. The main difference being the usage of N-1 instead of N-τ-1 in the denominator

in Equation (2.3) which in fact artificially suppresses, at large lag times, the measured values.

This is sometimes done because the data at long lead times are expected to be unstable since the

number of data samples there becomes less and less as the lead times becomes longer.

Examples of the observed values r_X of ρ_X for the Martlesham-Leidschendam trans-horizon

microwave link and the Stockholm slant-path link, discussed in the introduction, are given in

Figure 2.3. This figure suggests clearly the *periodic* nature of the process *X*.
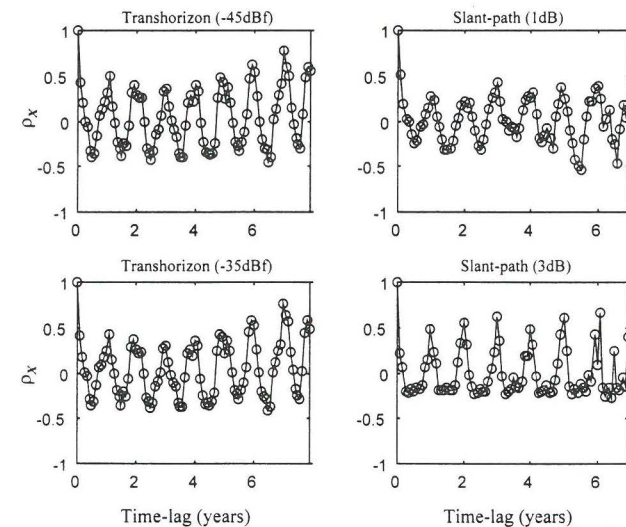

Figure 2.3 Examples of observed autocorrelation functions.

Such seasonal dependence observed of many propagation processes suggests that these processes are *cyclo-stationary* in nature with a time period $T = 12$ months (*annual cyclo-stationary*). Formally speaking a random process $X(t)$ is called cyclostationary if all of its statistics are invariant to a shift of the time origin by integral multiples of a constant period T [*Papoulis*, 1985]:

$$Probability(X(t_1 + kT) \geq X_1, \ldots\ldots, X(t_N + kT) > X_N) =$$
$$Probability(X(t_1) > X_1, \ldots\ldots, X(t_N) > X_N) \tag{2.6}$$

with k any integers value. It is clear that formally proving or disproving such a hypothesis is quite beyond our current abilities. The physics of weather phenomena is of such complexity that it is practically impossible to formally prove or reject the hypothesis, although we can note the yearly period of the earth's rotation around the sun as supporting some kind of cyclostationarity.

Theoretically we need not restrict ourselves to cyclic periods of 12 months; other cycles may also be present, one of which is the 11-year 'sunspot' cycle which can be observed in ionospheric radio propagation [*Picquenard*, 1974]. A study by [*Poires Baptista et al.*] 1986] seems to suggest that the annual rainfall intensity statistics in Europe exhibits also an 11 year cyclical behaviour. Examples of diurnal effects are the diurnal variation of propagation phase-length due to diurnal variation of the total electron content (TEC) of the ionospheric path [*Mawira*, 1990] (see Figure 2.4) and the diurnal variation in rain intensity distribution in Indonesia [*Brussaard et al.*, 1993]. Another cyclic phenomena is the well-known occurrence of weather 'spells', long periods of deviating weather conditions which span a very large area even up to continental scales, such as the phenomenon known as El Nino. Finally we mention the very long duration cycles associated with major climatic changes of the earth [*Miller & Thompon*, 1970].
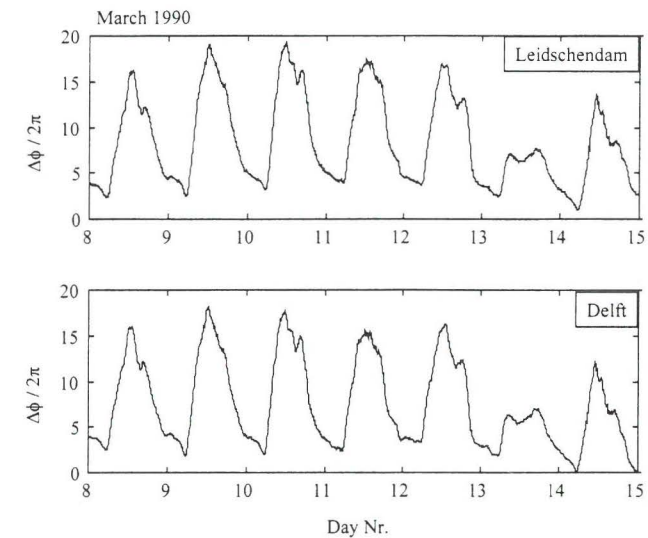
Figure 2.4. 12Ghz phase-length variations observed in Delft and Leidschendam, using the Olympus satellite beacon.

Effects shorter than 1 month, such as diurnal variation, will not be considered in this study. Furthermore, for the systems in which we are interested here (satellite and microwave) the main propagation processes of importance are those associated with rainfall and with ducting. For these, the existence of 1-year cycles in many parts of the world is well established. Although other longer-term cycles have certainly been noticed there is at the moment very little relevant data for us to work with. The effects of cycles associated with major climatic changes is, hopefully, not noticeable within our life-span.

We shall therefore, as a working hypothesis, formulate a cyclostationary model with period $T = 12$ months (annual cyclostationarity). In this case the first-order statistics of X may need 12 distinct statistical distributions:

$$P_{X|m}(X) = Probability(X \geq X \mid \text{within calendar month } m) \tag{2.7}$$

where $m = 1, 2, ..., 12$ indicates the twelve calendar months January, February, ..., December.

As a corollary of the annual cyclostationarity assumption we have the fact that the subprocess X_m of X associated with calendar months m is stationary [*Papoulis*, 1985]. We may thus view X as being composed of 12 distinct stationary random processes, X_m:

$$X = (X_1, X_2, ..., X_{12})$$ (2.8)

with each X_m, $m = 1, 2, ..., 12$, having its own distinct statistics $P_{X|m}$ (see Equation (2.7)):

$$\Pr(X_m \geq X) = P_{X|m}(X)$$ (2.9)

Note that the processes X_m are discrete-time processes at yearly time intervals.

A question of importance in mathematical statistics is that with regard to the *ergodicity* of the random process. In mathematical statistics ergodicity, in general, means that the statistical parameter which can be derived by observing one member of the ensemble over an infinitely long period is equal to the same type of parameter associated with the whole ensemble taken at a fixed point in time, [*Papoulis*, 1985]. Actually there are various forms of ergodicity. We are here interested in <u>distribution ergodicity</u>; this is the case where the distribution function of a random process can be derived from time observation of a single member of this random process. For a cyclo-stationary process this is mostly not the case. At most ergodicity may be valid for the sub-processes X_m associated with calendar months m (=1, 2, ..., 12). According to the theory of mathematical statistics, stationarity and independence comprise a sufficient condition for ergodicity, see Section 9.5 of [*Papoulis*, 1985]. To complete the framework for our model we have to consider the question of statistical *independence* of the process. Independence of the random processes in different months means that all their joint statistics can be decomposed to 'products' of the marginal distributions $P_{X|m}$ (or equivalently the probability distribution of the X_m's):

$$Probability(X_1 \geq X_1, X_2 \geq X_2, ..., X_{12} \geq X_{12}) =$$
$$Probability(X_1 \geq X_1)\,Probability(X_2 \geq X_2)......Probability(X_{12} \geq X_{12})$$ (2.10)

(see [*Papoulis*, 1985]). Therefore, in the case of independence the 12 statistics (the $P_{X|m}$'s) form a sufficient base to fully characterise the random process X.

Proving independence is a hard task, especially when the data are not so numerous. For example in our case when investigating pair-wise independence using the *contingency-table* analysis [*Cramer*, 1946], [*Kreysig*, 1970], with 2x2 tables the average number of points per table element per location would at most be 3 (12years/(2x2)) which would make the analysis practically meaningless. An alternative analysis is through the use of the *autocorrelation function* (or as is sometimes also called the *correlation coefficient*). A strong correlation value implies automatically a lack of independence, and although lack of correlation does not mathematically imply independence it is usually, in physical processes, a good indication of independence. In using the correlation analysis some careful consideration have to be made. In the first place it must be noted that the autocorrelation function only gives the average correlation properties between the months. E.g. for the correlation coefficient for one month time shift represents the average of the correlations of the months January vs. February, February vs. March, ... , December vs. January. A second aspect that need to be accounted for is the fact that since a cyclostationary process is non-ergodic we have to be careful in determining its statistical properties from a single time sample. At first sight Figure 2.3 would suggest that most of the months are strongly correlated and are therefore also strongly dependent. However, we need to note that Equation (2.3) by which the autocorrelation function has been determined is an equation developed for obtaining the estimator of the autocorrelation function of an ergodic process.

Dellagiacoma & Tarducci [1984] have applied such correlation analysis, the data used are composed of some 200 site years of observations from Italy. They reported very low correlation values for lag times ≥ 1 month, and concluded that the statistical processes in separate months are independent. For slant-path attenuation and ducting effects we also investigated the

independence hypothesis using data produced by the European COST205 and COST210 projects; the dataset is described in Appendix 2.1.

Here we consider the behaviour of the (normalised) autocorrelation function $\rho_{\Delta X}$ of ΔX which is the random process X corrected for the seasonal trend:

$$\Delta X(t_n) = X(t_n) - \mu_S(t_n) \tag{2.11}$$

where μ_S is a cyclic function of time constructed from the calendar monthly means $<X_m>$ according to:

$$\mu_S(12j + m) = <X_m> \tag{2.12}$$

where j is the year number and m is the calendar month index number. This removal of the seasonal trend is necessary, otherwise we would have in the autocorrelation the correlation related to the seasonal pattern. We could see this from the equation relating $\rho_{\Delta X}$ to ρ_X:

$$\rho_X(\tau) = \frac{\sigma_{\Delta X}^2 \rho_{\Delta X}(\tau) + \sigma_{\mu_S}^2 \rho_{\mu_S}(\tau)}{\sigma_{\Delta X}^2 + \sigma_{\mu_S}^2} \tag{2.13}$$

where ρ_{μ_S} and σ_{μ_S} are the autocorrelation function and the standard deviation of the 'deterministic' seasonal pattern μ_S, while $\sigma_{\Delta X}$ is the standard deviation of ΔX (the random component of X). We can see from this equation that even if the process ΔX is fully independent, in general ρ_X would not be = 0. The analysis for $\rho_{\Delta X}$ is first carried out using the two data set with the longest time period of measurement: that of ducting from the trans-horizon link Marthlesham-Leidschendam and that of rain-attenuation from the slant-path measurement in Stockholm.
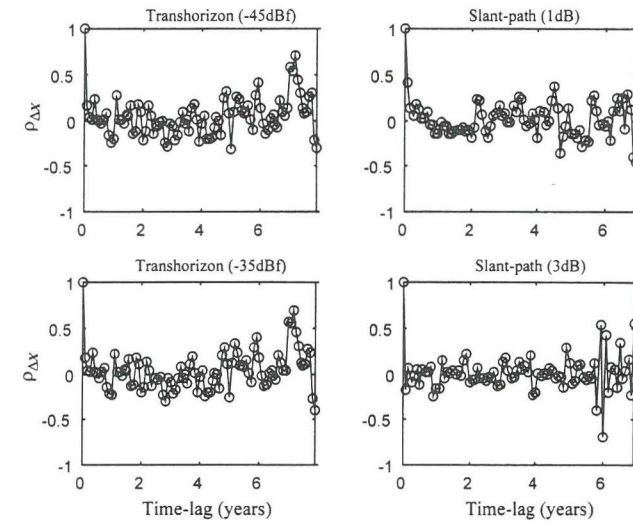
Figure 2.5 Examples of observed autocorrelation $\rho_{\Delta X}$ obtained after correction for the seasonal trend.

Figure 2.5 shows the obtained observed behaviour of $\rho_{\Delta X}$. In general it is found that at time lags larger than 1 month most of the observed values varies somewhat randomly between -0.2 and 0.2 with an average of 0, with a notable exception for time lags larger than 6 years. For these large time lags we can see large positive and negative excursions of the observed values; these large values may however be attributed to the instability of the estimation as the number of samples available becomes less and less as the time lag becomes larger. An indication of the instability of the estimated correlation coefficient can be obtained by using the results of *Fisher* concerning test and confidence intervals for the correlation coefficient for two normal random variates (see Section 18.4 of [*Kreysig*, 1970]).

This leads to the following table giving the 90% confidence interval range for various sample numbers for the uncorrelated case:

Number of Samples:	48	36	24	12
90% Interval range:	± 0.240	±0.279	± 0.359	± 0.499
95% Interval range:	± 0.284	± 0.329	± 0.403	± 0.574
99% Interval range	± 0.366	± 0.421	± 0.510	± 0.696

Table 2.1 Confidence intervals for the correlation coefficient

E.g. this table shows that when experiments were performed where each time we use 48 (independent) samples to estimate the correlation coefficient between two uncorrelated normally distributed random variables we may expect that in 90% of such experiments the values observed will be between the range -0.24 to +0.24.

From the above results we may conclude that at time lags longer than one month the months are uncorrelated which is indicative of independence. For a time lag of 1 month the values observed are between 0 and 0.4, with an average of around 0.2, which indicates some kind of dependence between the subsequent months.

By restricting ourselves to shorter time-lags we could analyse a larger set of data. The first of this consist of all the trans-horizon data, the second of this consists of the slant-path data obtained in NW-Europe. Figures 2.6 and 2.7 show the results when averaged over three ranges of thresholds corresponding to long-term average X. The ranges are (a) greater than 1%, (b) from 0.1% to 1%, (c) from 0.01% to 0.1%. Figure 2.6 gives the results for ducting effects on trans-horizon links, while Figure 2.7 gives the results for rain attenuation on slant-paths. These results also show very small correlation values at time lags larger than 1 month, both for ducting and for rain attenuation. For a time lag of 1 month we see a difference. For ducting the correlation value is about 0.3 irrespective of the probability range. For rain attenuation the correlation value is here also very small except for the probability range > 1%, here we have a value of also some 0.3. At

this range we have only data from one site (Stockholm) at one threshold level (1dB), however the measurement period is reasonably long: 8 years. A physical interpretation of these observations will be given in Section 5 on fundamental issues. Here we just note that the magnitude of correlation between the months depends strongly on the 'typical' duration of weather conditions conducive to the relevant propagation activities.

In conclusion, since the correlations are mostly very low, we feel justified in postulating indepen- dence in our model, the more so because, as will be discussed in Section 2.2, further simplification of the model must be implemented in order to make it useful to system planners. In Section 4 on worst-month aspects we make some estimate of the possible errors which may be produced by this simplifying assumption.
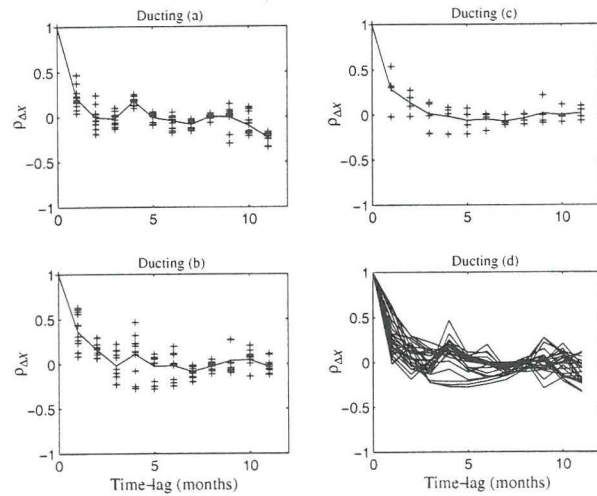
Figure 2.6. The autocorrelation function $\rho_{\Delta X}$ for pooled ducting data.

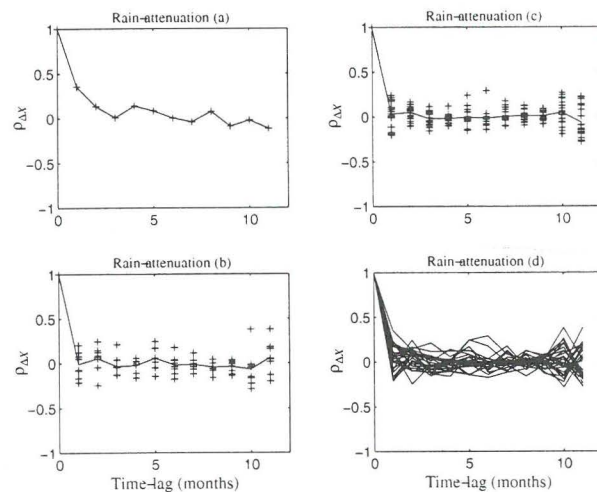(a) range > 1%, (b) 0.1 – 1%, (c) range < 0.1%, (d) all.

Figure 2.7. The autocorrelation function $\rho_{\Delta X}$ for pooled rain-attenuation data.

(a) range > 1%, (b) 0.1 – 1%, (c) range < 0.1%, (d) all.

2.2 A simplified model based on the aggregate distribution P_X

To obtain a working model we must apply some further simplification. This is because the available propagation data are usually not enough for characterising the 12 distributions $P_{X|m}$. The available data are at most only sufficient for quantifying the seasonal dependence of the average of X for some parts of Europe. Furthermore, seasonal dependence may be very different in different parts of the world, so that a generally applicable statistical model cannot be expected.

To circumvent this problem we shall use a simplified model which is based on the average, or aggregate, statistical distribution:

$$P_X(X) = \frac{1}{12}\sum_{m=1}^{12} P_{X|m}(X) = \text{probability}(X \geq X) \tag{2.14}$$

As will be shown in later sections this distribution can be determined with the aid of currently available and proven propagation methods for the prediction of P_a and supplemented with a one-parameter climatic map regarding variability/seasonality.

In this simplified model we assume that only the average distribution P_X is available to the user instead of the 12 statistical distributions $P_{X|m}$ associated with each of the 12 calendar months January through December. In this situation the only indication concerning any seasonal variation is to be found in the average number of months per year, M_{Co}, with $X > 0$:

$$M_{C_o} = \lim_{\varepsilon \to 0} 12 P_X(|\varepsilon|) \tag{2.15}$$

While P_X alone is sufficient for the determination of simple statistical parameters of X (e.g. its long-term average), it is reasonable to expect that the inhomogeneity of the real statistical process should have an impact on the values of more complicated parameters associated with X

(e.g. the worst-month, to be discussed in Section 4). *Brussaard & Watson* [1979] have proposed a simple way to model this inhomogeneity of the statistics:

> The monthly population is assumed to be composed of two groups ('seasons'): M months per year being considered 'active' the other 12 - M months being considered 'inactive'. The inactive months will always produce $X = 0$, the active months may produce $X > 0$. Furthermore, the active months have the same statistical distribution and are independent (they are *i.i.d.* = independent and identically distributed). The active months are equivalent to 'the months that could produce a worst-month' as defined by *Crane* [1991]. Finally, M is an integer number.

Within the ensemble of the active months the statistical distribution of X is given by:

$$P_X(X \mid X \in Active) = P_X(X) / \text{probability}(X \in Active) = \frac{12}{M} P_X(X) \qquad (2.16)$$

M is not directly observable (assuming that we only have P_X at our disposal), but a lower and upper bound can be established. The upper bound is obviously 12 while the lower bound is M_{Co}.

The aggregate distribution is sufficient to determine some of the statistical parameters of X. E.g. the time average of the *statistical moments* $\overline{<X^k>}$ of X can be determined by using only the aggregate distribution (instead of having to use all the 12 monthly distributions). This can be seen by noting that:

$$\overline{<X^k>} = \frac{1}{12} \sum_{m=1}^{12} \int_0^\infty dX \, X^k \left(-\frac{d}{dX} P_{X|m}(X)\right) = \int_0^\infty dX \, X^k \left(-\frac{d}{dX} \frac{1}{12} \sum_{m=1}^{12} P_{X|m}(X)\right) = \int_0^\infty dX \, X^k \left(-\frac{d}{dX} P_X(X)\right) \quad (2.17)$$

The right hand side of this equation is obtained by using the definition for P_X given in Equation (2.14).

The time-averaged first moment, i.e. $\overline{<X>}$, is related to the ITU's long term statistics P_a through:

$$\overline{<X(a)>} \approx P_a(a) \qquad (2.18)$$

The approximation stems from the fact that the calendar months may have different number of days; the maximum discrepancy will be found if the propagation effects should only occur in the month February (this discrepancy is however very small, the possible values of $\overline{<X>}/P_a$ lies between 1 and 1.08). A fine point to be noted here is that the ITU definition, given in Recommendation ITU-R P581-2 [*ITU*, 1997], only implies a long-term averaging across a sample of X.

Finally, it should be noted that for any two threshold values a_1 and a_2 such that $\overline{X(a_1)} \geq \overline{X(a_2)}$ the following inequality applies:

$$P_{X(a_1)}(X) \geq P_{X(a_2)}(X) \qquad (2.19)$$

In a later section concerning the worst-month it is shown that P_X, also known as the parent distribution, is sufficient to predict the extreme tail of the annual worst-month TFE, which makes it attractive for use by planners of radio links. Furthermore, the planners of the radio links are often only interested in the occurences of extreme months and do not need to know the specific month of occurrence (e.g. January 1995), so that the usage of P_X instead of $P_{X,m}$ is not so restrictive as it might seem at first sight.

2.3. Candidate models for P_X

Figures 2.8 and 2.9 give examples of measured P_X, for slant-path and ducting effects. These figures illustrate the often observed fact that severe propagation effects only occur in a fraction of the months, the other months producing $X = 0$ (conditionality or seasonality aspect). The shapes of the distributions suggest that exponential, lognormal or gamma type of distributions can be used to model P_X provided suitable modifications are included to accommodate for this conditionality aspect. See Figure 2.10 for examples of the behaviour of these distributions. In this section such models for P_X will be investigated.

Slant-Path, Stockholm
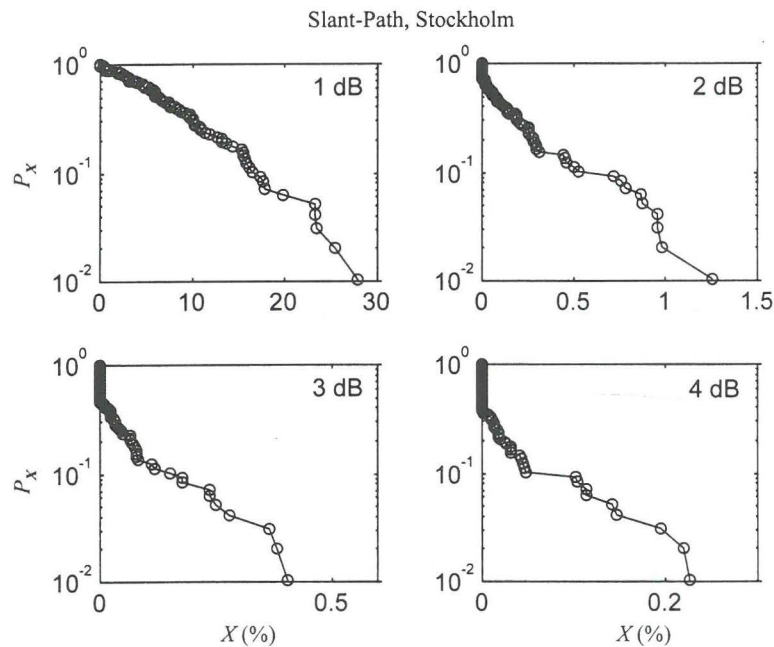
Figure 2.8 (a) Examples of P_X, from measured rain-attenuation

The first model is the Conditional Exponential (CEX) model, first proposed by *Crane and DeBrunner* [1978]:

$$P_X(X) = C_0 \, e^{-X/C_1} \tag{2.20}$$

for $X > 0$, $P_X = 1$ for $X < 0$; $0 < C_o < 1$, $C_1 > 0$.. In this study we extend the model by allowing C_0 values greater than 1: in such cases the above equation is valid when $X > C_1 \ln(C_o)$ otherwise $P_X = 1$; this extension was first introduced by the author in [*ITU*, 1992].

Slant Path, Northern Italy

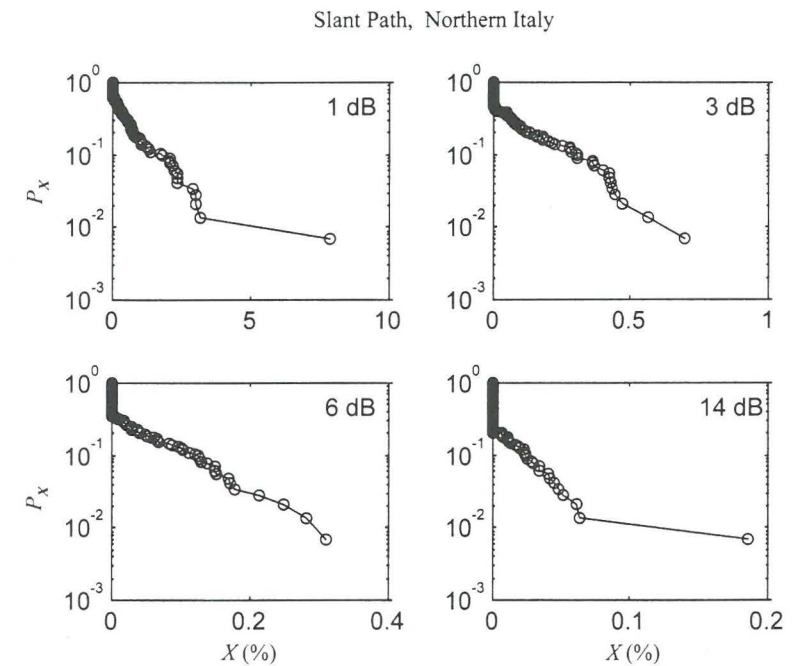Figure 2.8 (b) Examples of P_X, from rain-attenuation measurements

The second model is the <u>Conditional Lognormal</u> (CLN) model, more recently proposed by *Crane* [1991]:

$$P_X(X) = C_0 \int\limits_{\frac{1}{C_2}\ln(\frac{X}{C_1})}^{\infty} dt \, \frac{e^{-\frac{1}{2}t^2}}{\sqrt{2\pi}} \tag{2.21}$$

for $X > 0$, $P_X = 1$ for $X < 0$; $0 < C_0 < 1$, $C_1, C_2 > 0$..

Ducting, Martlesham - Leidschendam
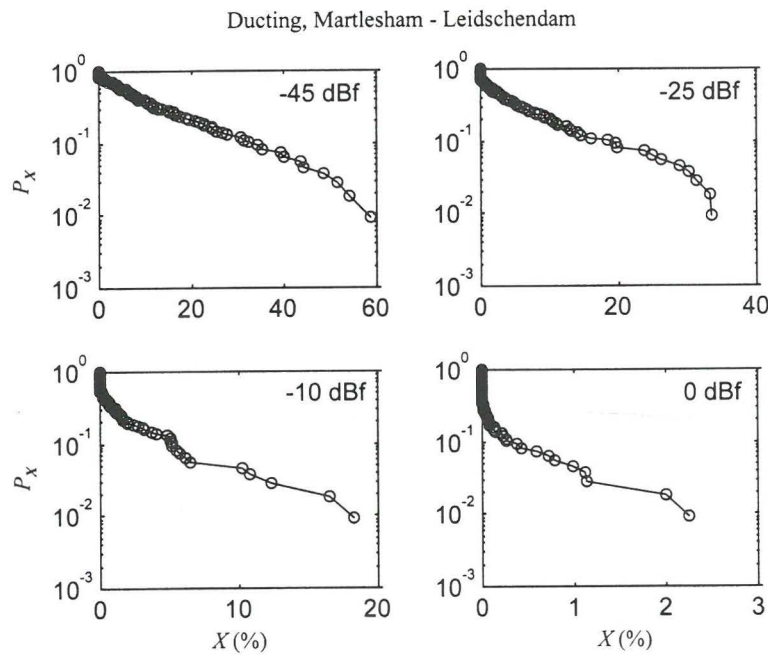
Figure 2.9 (a) Examples of P_X, from ducting data

The third model is the <u>Gamma</u> model:

$$P_X(X) = \int\limits_{\frac{X}{C_1}+\kappa}^{\infty} dt \, t^{C_2-1} \, \frac{e^{-t}}{\Gamma(C_2)} \tag{2.22}$$

for $X > 0$, $P_X = 1$ for $X < 0$; $C_1 > 0$, $C_2 > 0$. Γ is the well-known gamma function. κ is the shift parameter; for our study we consider two cases:

$$\kappa = 0 \tag{2.23}$$

and

$$\kappa = C_2(1 - \sqrt{C_2})H(1 - C_2) \tag{2.24}$$

where H is the unit-step function: $H(X) = 1$ for $X \geq 0$ and $H(X) = 0$ for $X < 0$. H is also known as the Heavyside function. In the first case we have a normal Gamma (GM) model. For the cases $\kappa \neq 0$ we have a shifted Gamma model (SGM), which is the distribution of a Gamma variate from which the value κC_2 is subtracted and afterwards the negative values are truncated to zero. A 'shifted' gamma distribution, with $\kappa = C_2$, was first proposed by *Fukuchi and Watson* [1989] for describing the variability of yearly time fraction of excess. However, our initial tests shows that the SGM distribution with $\kappa = C_2$ does not model the observed distribution of the monthly time fraction of excess very well. After some experimentation, Equation (2.24) was found to produce a much better model. The standard Gamma model (i.e. Equation (2.22) with $\kappa = 0$) was tested and rejected by *Crane* [1991]; this was to be expected, because the standard Gamma distribution is an unconditional model and is obviously unsuitable for modelling a conditional phenomenon. It always fails for very small values of X. In our study we shall consider again this Gamma model, here we investigate in how far the effects of very small values of X can be ignored.

Ducting, Irnsum - Leidschendam
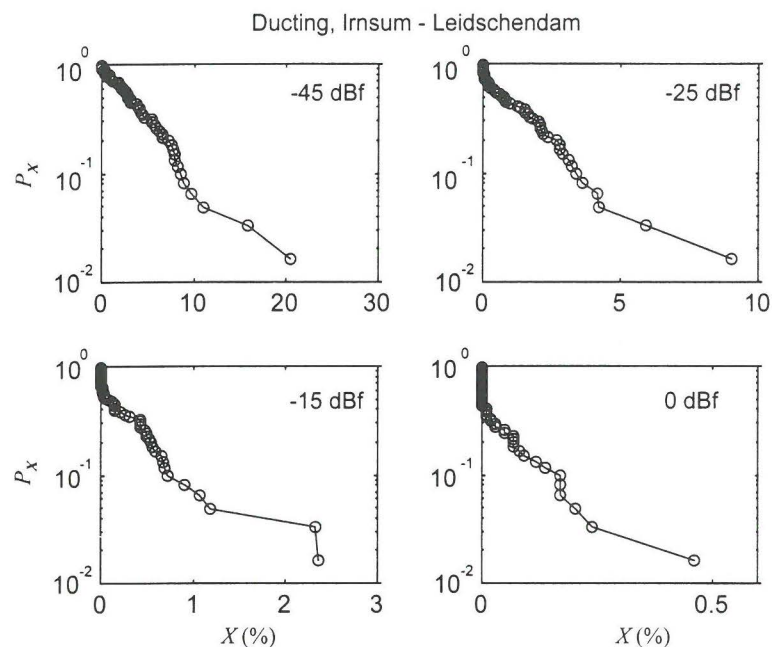
Figure 2.9 (b) Examples of P_X, from ducting data

Figure 2.10 Examples of behaviour of the 4 model distributions.

Figure 2.10 gives examples of the behaviour of these distributions. This figure is obtained by simulating a CEX process, with $C_0 = 0.5$ and $C_1 = 0.1\%$, and afterwards applying a 'best-fit' on the four models; such 'best-fit' procedures will be discussed in Section 2.4.2. This figure shows that there is an intermediate range where the four models are difficult to distinguish from one another. For very small values of X the models behave differently, however, here the measured data is not very accurate. At the tail of the distribution the four models also behave differently, however to measure the tail accurately a long-term observation is needed. It is interesting to note that at the tail of the distribution the CEX model descents to most rapidly, while the CLN is the slowest to descent to 0.

We note here that, mathematically speaking, the above models can only be considered to be approximations. This is because the monthly time fraction of excess X is, by definition, never larger than 1, so that P_X must be equal to zero for all $X > 1$ (which is clearly not the case for the three models). In practise, however, the thresholds of interest are associated with X values that are well below 10% and the approximation should cause no problems.

C_0 is sometimes called a location (or conditionality) parameter, C_1 is sometimes called a scale parameter and C_2 is sometimes called a shape parameter. The conditional exponential model and the shifted gamma model are two-parameter distributions while the conditional lognormal model

is a three-parameter distribution Note that in cases where confusion may arise we shall append the appropriate model index to these parameters (e.g. $C_{0,CEX}$ or $C_{0,GM}$).

The conditionality of the phenomenon is explicitly accounted for in the conditional exponential model and the conditional lognormal model by the inclusion of the conditionality parameter C_0. In the shifted gamma model the conditionality is implied, and we can define an <u>effective</u> C_0 parameter for the SGM model:

$$C_0 = \int_{\kappa(C_2)}^{\infty} dt\, t^{C_2-1}\, \frac{e^{-t}}{\Gamma(C_2)} \qquad (2.25)$$

As noted in Section 2.2 this long-term average $\overline{X}$ is related to the ITU's long term statistics P_a through:

$$\overline{X}(a) \approx P_a(a) \qquad (2.26)$$

The mathematical expression for the dependence of $\overline{X}$ on the model parameters has been derived and is given in Appendix 2.2. An important point to note here is that if P_a is known, e.g. by using good prediction methods, the number of parameters required to determine P_X is reduced by one. For the CEX and SGM models, therefore, in addition to a prediction method for P_a we would only need to determine C_0 in order to fully predict P_X. The prediction aspects of the determination of the model parameters are discussed in Section 2.5.

Finally, it should be noted that for any two threshold values a_1 and a_2 such that $\overline{X}(a_1) \ge \overline{X}(a_2)$ the following inequality concerning the model parameters can be derived, through use of Equation (2.19):

$$C_i(a_1) \ge C_i(a_2) \qquad (2.27)$$

for $i = 0$ for all three models. This equality is also valid for $i = 2$ for the CLN, GM and SGM models. For $i = 1$, Equation (2.27) has only general validity for the CEX model. However, intuitively we would expect it also to be valid for the CLN, GM and SGM models.

2.4 Testing of the models for P_X

2.4.1 Data used in the testing

In this section we shall test the suitability of the candidate models for P_X using measured propagation data. The data used for this testing are derived from those produced in the European co-operative projects COST205 and COST210, refs. [*COST205*, 1985], [*Fedi*, 1985] and [*COST210*, 1990]. Within the COST205 project, which took place in the period 1975-1980, data from propagation experiments carried out across Europe using the OTS and SIRIO satellites were collected and analysed. The data here that is of interest to us is that with regard to rain-attenuation. In the COST210 project, which took place in the period 1984-1990, data from trans-horizon links in Europe were collected and analysed. The data here that is of interest to us is that with regard to ducting. These data have been preserved in the form of monthly cumulative distributions.

An important requirement for accurate testing of models for P_X is the availability of a sufficient number of sample points. Furthermore, since the model to be tested is conditional it is the number, N_o, of samples with $X_{i,j} > 0$ that is important for the test. To establish some minimum size of the data we follow the following reasoning:

Let us assume that to reasonably test a single distribution we would require at least $N_o = 10$ points with $X_{i,j} > 0$. Further, for rain-attenuation in Europe for thresholds associated with a long-term probability of 0.01% (an important Grade of Service target) the average value of C_o is around 0.23, see [*COST205*, 1985]. This results in a required minimum of $N_o/(12C_o)$ = 4 years of observation per distribution to be tested. Since in order to avoid effects of measurement errors we need to remove some data points which are to close to zero we may establish 5 years as the lower bound for the required observation time per distribution.

For our study we apply a scheme of data pooling in order to obtain large number of samples for various climatic regions of Europe. We have here some 69 site years of 11GHz slant-path attenuation data (from 8 zones across Western Europe) and 23 site-years of trans-horizon

(ducting) data (from 5 zones around the English channel). Table 2.2 and 2.3 summarise the data, Figure 1.2 and 1.3 of the introduction gives the geographical locations of the measurement sites; for more detailed description of the measurement data see Appendix 2.1.

Table 2.2: 11 GHz slant-path rain-attenuation data

LOCATION	ID	Years	Source	COMMENTS
Austria	AUS	9	COST205	Lustbuehel: 4 y. OTS, Graz: 5 y. RADIOMETER. RADIOMETER *data scaled to OTS configuration*.
Belgium & The Netherlands	BNL	9	COST205	Louvain La Neuve: 2y., Lessive: 2y., Eindhoven: 1y., Nederhorst: 3y., Noordwijk: 1y.. All OTS data.
Germany	GER	7	COST205	Leeheim: 3y. OTS. Leeheim: 4y. RADIOMETER. RADIOMETER *data scaled to OTS configuration*.
Finland	FIN	5	COST205	Sodankyla: 5 y. OTS.
Italy (northern)	ITA1	12	COST205	Lario: 5y. SIRIO. Spino d'Ada: 4y. SIRIO. Milan: 3 y. RADIOMETER. RADIOMETER *data scaled to SIRIO configuration*.
Italy (central)	ITA2	11	COST205	Fucino: 5y. SIRIO. Fucino: 3y. OTS. Fucino: 3y. RADIOMETER. SIRIO & RADIOMETER *data scaled to OTS configuration*.
Sweden	SWE	8	COST205	Stockholm: 8y. RADIOMETER.
UK (southern)	UK	8	COST205	Martlesham: 3y. OTS. Slough: 2y. OTS. Slough: 3y. SIRIO. SIRIO *data scaled to OTS configuration*.

PS. The (attenuation) threshold values used in this thesis are 1, 2, 3, 4, 6, 8, 10, 12, 14 dB.

Table 2.3: Sources of ducting data

LOCATION	ID	Years	Source	COMMENTS
Martlesham - Leidschendam	MALS	8	COST210	Frequency = 1.3 GHz. Path-lenth = 110km.
Martlesham - Eindhoven	MAEI	5	COST210	Frequency = 1.3GHz. Path-length = 150km
Martlesham - Blavand	MABL	5	COST210	Frequency = 1.3GHz. Path-length = 200km.
Insum - Leidschendam	INLS	5	KPN Research	Frequency = 6.4GHz. Path-length = 100km.

PS. The threshold values used in this thesis are: -45, -35, -25, -15, -10, -5, 0, +5 dB below free space level.

2.4.2 Discussion of statistical test methods

Methods for testing of suitability of model distributions (goodness of fit tests) have been a subject of research in the branch of mathematical statistics for nearly a century. In this period numerous methods have been developed; most well known of these are the Chi-squared (or χ^2) test [*Cramér*,1946], the Kolmogorov-Smirnov test [*Massey*, 1951], the Cramér-von-Mises test and the Anderson-Darling test [*Pearson & Hartley*, 1976].

---

(1) Take N observations $\{X\}$ from the population X (<u>measurements</u>). The larger the number of independent observations taken the better we would be able to test the model. Obviously the measurements should be accurately performed.

(2) Construct from that sample the observed distribution P_{obs} (the <u>sample cumulative (or frequency) distribution function</u>, s.c.d.f. resp. s.f.d.f.). A method for constructing the s.c.d.f. is that due to [*Gumbel*, 1967]. Here the test points for the s.c.d.f. are taken to be identical with the observation points. The probability assigned to the i-th (Rank i) largest sample point, $X_{Rank\text{-}i}$, is:

$$P_{obs}(X_{Rank\text{-}i}) = i \,/\, (N+1)$$

This method of probability assignment is called *Rank-ordered* statistics. Note that some authors use, somewhat different estimators, e.g. i/N or $(i\text{-}0.5)/N$ or $(i\text{-}1)/N$. The s.f.d.f. is usually constructed by the well known histogram method.

(3) Estimate the model parameters $\underline{C}_{est}$ (<u>parameter estimation</u>). This is done using a suitable procedure, *EstimateParameter*(), applied to the observation points: $\underline{C}_{est} = EstimateParameter(\{X\})$. In some cases the model parameters are *a priori* given, e.g. this is the case when we are testing whether an in the past (fully known) process has changed with time.

(4) Determine the probability values <u>predicted</u> by the model, e.g. $P_{model}(X_{Rank\text{-}i},\underline{C}_{est})$. This is a simple application of the mathematical equation of the model distribution.

(5) Quantify the deviation D between the predicted and observed values. The <u>test-variate</u> D is determined using a suitable procedure, *Deviation*(): $D = Deviation(P_{obs}, P_{model})$.

(6) Derive conclusions from D (<u>statistical inference</u>). The basis for this step is the fact that D is actually a single observation from an ensemble of possible values $\boldsymbol{D}$ (random variate). For a given model one can, at least in principle, determine theoretically the expected distribution P_D of $\boldsymbol{D}$. Now, by considering this distribution and the observed value D a statement can be made regarding the likelihood that the assumed model is correct (or incorrect).

1

Panel 1. Procedural steps for goodness of fit tests

Even though the specifics of these test methods may differ considerably from one another, most of such test methods include common procedural steps that must be performed, see Panel 1.

Although at first sight these steps seem straightforward, there are subtle aspects which need to be considered before choosing and applying the various statistical test methods. A crucial element in making statistical inference is the knowledge regarding the statistical properties of the test (random) variate $\boldsymbol{D}$, such as the knowledge of its distribution P_D . It is on the basis of this knowledge that from an observed value D the likelihood of the validity of the model being tested can be quantified. E.g. in <u>hypothesis testing</u> the hypothesis: "the model is correct" is rejected if $P_D(D)$ is smaller than a (small) agreed value called the <u>level of significance</u>. Note that the level of significance can be interpreted as the risk we are willing to take of erroneously rejecting a correct model/hypothesis.

It may be argued that when we are comparing the merits of two models it is sufficient to use an 'intuitive' approach, that is to say that the model producing the smaller value for the test variate is the correct model. However, this approach is only valid if the test random variates associated with the two alternative models have the same statistical distribution, otherwise the comparison will be biased in favour to one of the models. Test methods which produce a test variate $\boldsymbol{D}$ with statistical distribution P_D which is independent of the distribution of the random variate X which is to be tested may be used in this 'intuitive' way.

A test method for which P_D is independent of the statistical distribution of the random variate X being tested is called a <u>distribution-free</u> test-method. Such a test method is attractive since that P_D need only to be determined once (preferably in analytical form) and is then subsequently available for many test applications. The Kolmogorov-Smirnov test, the Anderson-Darling test and the Cramér-von-Mises test are distribution-free methods with known expressions (or at least tabulated values) for P_D provided that the model distribution is <u>fully specified</u>. That is to say the model parameters $\underline{C}$ must be a priori known/given and not derived from the samples being tested. The basis for all such model is the probability transform of the observed values:

$$U_k = P_X(X_k, \underline{C}) \tag{2.28}$$

which is essentially transforming the random variate X to a random variate U which is uniformly distributed in the interval 0 to 1. The test variate D measures the deviation of observed/sample values $\{U\}$ from that expected for a uniform distribution. Although the specifics of the measure D varies from method to method, e.g. root mean square measure, maximum deviation measure etc., the essential feature of these methods is the assumption of uniform distribution of U in the derivation of the expected statistical behaviour of D.

In the cases where the model parameters were derived from the data samples we may not assume anymore that U is uniformly distributed, since the estimated values $\underline{C}_{est}$ will in general be different from the true values $\underline{C}$. In such cases the distribution of D will be different than the distribution obtained in the case of fully specified models. In fact the distribution of D will now be dependent on the distribution of the random variate being tested. The test method is now not distribution-free anymore. Finally, we note here that for popular distributions such the normal distribution the modifications to the distribution-free methods, when $\underline{C}$ has to be derived from the data, has been calculated and tabulated, see e.g. [*Pearson & Hartley*, 1976].

The χ^2 test method, [*Cramér*, 1945], is based on comparing measured with expected histograms and is only distribution-free in the asymptotic sense (for sample size $N \rightarrow \infty$). There are actually two versions of the χ^2 test method. The first, which we shall call the *classical* χ^2 test, is to be used only for testing fully-specified model distributions, this version is the most well known and (mis)used. Since for our analysis we must determine the model parameters from the data samples the above test method cannot be used. A variation of the χ^2 test which allows parameter estimation is the χ^2-minimum test [*Cramér*, 1947]. In this test the estimated model parameters $\underline{C}_{est}$ are obtained by minimising the value of the test variate. Again, this test is only distribution-free in the asymptotic sense, so that its application to our situation (limited number of samples) must be done with some caution. We shall discuss this test further in Section 2.4.3.

Another test method that is popular in engineering community is the regression test. In this test the model parameters are determined through regression. That is to say the optimum values of the model parameter are those which minimise the deviation between the model probabilities and the observed probabilities. The popularity of this approach stems from the fact that for some

probability distributions a transformation has been found which reduces the regression to a linear form (allowing manual graphical determination of the model parameters). The problem with this method is that we have no indication at all of what consists of significant deviation of the test variate. This test is furthermore very much model dependent. Because of the popularity of this test we shall apply it also in Section 2.4.3.

Finally, since one of limitations of severe testing is the limited number of points comprising each distribution, we shall apply a novel method to pool all the available points using a normalisation procedure and analyse the resulting "grand-pool" of data, this is also done in Section 2.4.3.

2.4.3 Results of tests

χ^2 minimum method

The "χ^2 minimum" method is based on the χ^2 parameter. For the calculation of this parameter the whole of the variable space is first divided into r bins. The observed number of sample points in the bins is then compared to the expected number of points predicted by the model distribution using the following test variate:

$$\chi^2 = \sum_{i=1}^{N} (n_i - N\Delta P_i(\underline{C}))^2 / N\Delta P_i(\underline{C}) \tag{2.29}$$

N is here the total number of samples, n_i is the number of observed samples in bin i, ΔP_i is the predicted probability of an observed value falling in the range of bin nr. i (for given model parameters $\underline{C}$). It has been proven (see [*Cramér*, 1945]) that if the number of samples $N \rightarrow \infty$ and if the model distribution is correct the variate χ^2 will have a so-called Chi Square distribution with number of degrees of freedom (NDF):

$$NDF = r - s - 1 \tag{2.30}$$

r is the number of bins used, s depends on the method by which the model parameters $\underline{C}$ has been determined. If the parameter vector $\underline{C}$ is known a priori then $s = 0$. If the parameter $\underline{C}$ has been estimated from the data then s equals the number of model parameters. Actually the latter theorem was only proven for cases when $\underline{C}$ was estimated by minimising Equation (2.29) (the χ^2 minimum method), however according to *Cramér* [1945] this theorem is also valid for a large range of parameter estimation methods. Our own investigation, using simulations, have shown that this statement is overly optimistic; large errors are incurred if we use the χ^2 in conjunction with various standard parameter estimation methods e.g. regression. For this thesis we will therefore use the χ^2 minimum method to estimate the model parameters.

The χ^2 distribution is equivalent to the Gamma distribution with scale parameter $C_1 = 2$ and shape parameter $C_2 = NDF/2$:

$$P_{\chi^2}(\chi^2) = \int_{\chi^2/2}^{\infty} dt \, t^{\frac{NDF}{2}-1} \frac{e^{-t}}{\Gamma(NDF/2)} \qquad (2.31)$$

The way in which the bins are constructed is described in Appendix 2.3.

Application of the method to the measured data described in Section 2.4.1 resulted in some 62 samples of χ^2 for each of the three models tested. Instead of applying the hypothesis testing to each sample we follow here a modified test method which obviates the need to somewhat arbitrarily choose significant levels. In this method we transform the χ^2 variate to a new variate U using:

$$U = P_{\chi^2}(\chi^2, NDF) \qquad (2.32)$$

(see Equation (2.31)). If the model (hypothesis) is correct then χ^2 will have the distribution according to Equation (2.31) and subsequently U will be uniformly distributed. The uniformity of U can be graphically investigated, since the (complementary) cumulative distribution, F_U, should be a linear function of U, (*probability*($U < U$) := $F_U(U) = U$). The results are given in

Figure 2.11. This Figure shows that the CEX, CLN and the SGM models produce reasonably linear relationship and can therefore not be rejected by this test. The points associated with the GM model, however, deviate considerably from the linear so that the GM is rejected by this test.
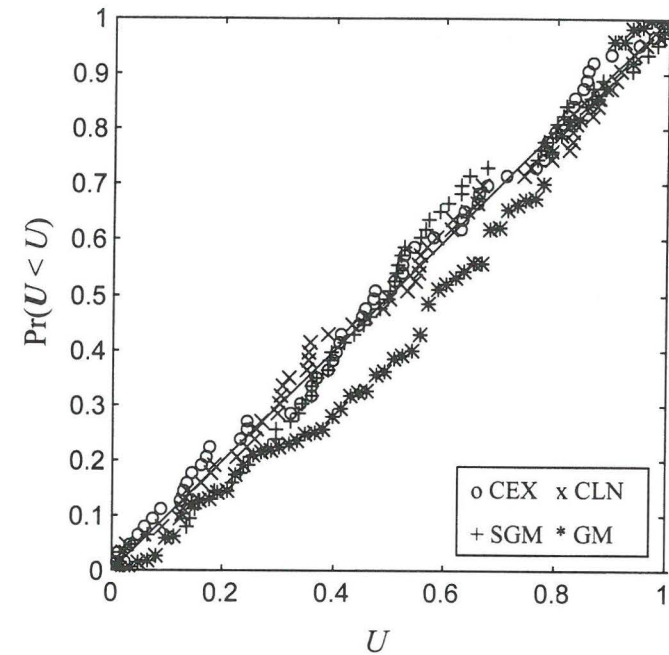

Figure 2.11 Results of the Chi-squared test

Regression method

In this method the best estimate for the values of the model parameters are taken to be those values that results in the 'best fit' between the model distribution and the sampling distribution (the observed distribution). Usually a probability transform is applied so that a linear relationship between the stochasic variate and the transformed probability is obtained. Subsequently a linear regression fit is applied which produces the estimated values for the model parameter. Then the quality of the fit is often quantified by the root mean squared (RMS) value of the logarithm of the ratio of modelled and the predicted probabilities, ε_{Px}, calculated using:

$$\varepsilon_{P_x}^2 = \frac{1}{N_0'} \sum_{n=1}^{N_0'} \ln^2 (P_X(X_n, \underline{C}) / p_n) \tag{2.33}$$

where X_n is n-th largest sample point, p_n is the probability value assigned to that sample (see *Panel* 1). The vector $\underline{C}$ represent the model parameters. N_o' ($\leq N_o$:= number of points with $X_{ij} >$ 0) is the number of points selected for regression. N_o' is introduced to avoid possible fitting problems associated with the conditionality of the models: e.g. small measurement errors may artificially considerably reduce or add samples with small values of X_{ij}'s which in turn can have a large effect on ε_{Px}. We note here that the regression approach has certain disadvantages, it emphasises in a somewhat arbitrary way the tail of the distribution. Furthermore, this method usually results in biased estimations of some of the model parameters (our simulations have shown that C_l is overestimated by this method). Nevertheless, since this method is popular in the engineering community we shall investigate it as well.

Further, it is important to consider the method by which N_0' can be determined (or in other words: how the 'active' population can be separated from the 'inactive' population). One method would be to just pick out the samples with $X > 0$, but this method was found to be less suitable due to sensitivity of measurements errors at low values of X (e.g. different filtering schemes, determination of reference levels). Furthermore for low values of X, secondary, less important seasons may have unduly large influence on the parameter estimation results. *Crane* [1991] has proposed a selection method. The selection method proposed, however, requires visual graphical inspection, which is of course not very suitable for processing large amount of data. Furthermore,

such a subjective selection method makes comparison of analyses by different researchers difficult.

We therefore introduced another (more objective) method for data selection. In this method, for each given N observation of X the active part is selected by taking the N'_o largest values of observed X that produce $z*100\%$ of the total sum of the observed X:

$$\sum_{i=1}^{N_{o'}} X_i = z \sum_{i=1}^{N} X_i \tag{2.34}$$

X_i is here rank ordered ($X_1 > X_2, \ldots, X_N$). Testing the consistency of the results leads to a recommended value of $z = 0.994$.

For the CEX model linear regression is applied to the $ln(P_X(X))$ vs. X relationship. For the CLN model firstly the parameter C_o is estimated using our implementation of the procedure devised by Crane, see Appendix 2.4, subsequently the inverse Gaussian transform is applied to $P_X(X)/C_o$ and a linear regression vs. $ln(X)$ then produces the parameter estimates. For the GM and SGM model no linear regression procedure is known. The parameter estimation here is done by the minimisation of ε_{Px}. This process is non linear and is done using MATLAB's FMINS function which is based on the Simplex search procedure.

Using the estimated model parameters, ε_{P_x} is calculated for the set of data described in Section 2.4.1. The values found are plotted against the number of points, N_0', with $X > 0$ in Figure 2.12. This figure shows that there is in general not much difference between the three models in the goodness of fit of P_X.
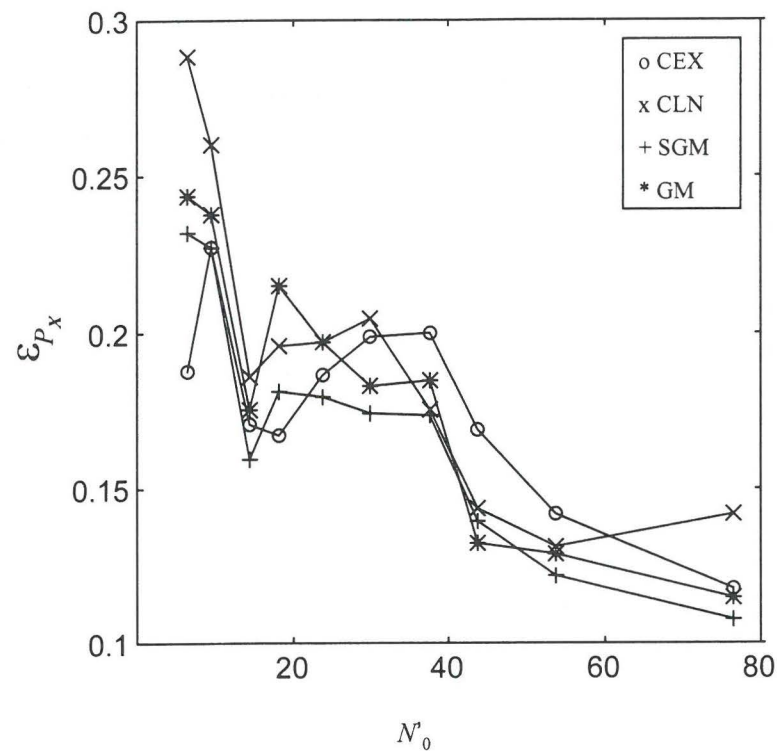
Figure 2.12 Results of the regression tests showing the dependence of the error variate ε_{P_x} .on the number of points N_0 with $X > 0$.

The normalisation method

The number of sample points of each s.c.d.f. (for a particular location and threshold) is rather small (at most, 120 for 10 years of measurements; the number of points with $X > 0$ is generally much smaller). This prohibits severe testing of the models. The total number of sample points with $X > 0$ is, however, very large (some 3500). If a suitable normalisation could be found then these can be considered as one large data pool.

A simple normalisation method which we will consider in this section is to use the fact that when a stochastic variate is transformed using the mathematical expression of its cumulative distribution function; the resulting new variate is uniformly distributed [*Gumbel*, 1967]. For a conditional distribution the variate with $X > 0$ is transformed to a uniform distribution in the range $0 - C_o$. In order to be able to pool all the data, we must use a transformation which results in the same range throughout. We can obtain this by using the unconditional distribution to perform the transformation:

$$V = 1 - P_X(X, \underline{C}) / C_0 \tag{2.35}$$

The variate V, associated with $X > 0$, obtained in this way is uniformly distributed in the range 0 - 1 (note that $X = 0$ will produce $V = 0$). Since all the data belong to the same type of process the pooling of all the transformed variates will lead to a very large set of a uniformly distributed population (actually, a conditionally uniformly distributed population due to the contribution of $X = 0$ respectively $V = 0$). The uniformity can be easily investigated through graphical inspection of the cumulative distribution of V.

Before going further we must first consider possible influence of the data selection procedure (during the parameter estimation phase). This selection procedure (also known as censorship) has been known to severely affect the shape of a distribution as well as the estimated parameters [*Hald*, 1952], [*Bury*, 1986]. Its possible effect should therefore be considered since in our new method we are going to consider the shape of the normalised distribution. The seriousness of

censorship depends on the degree of censorship and the nature of the parameter estimation used. In our case, censorship only affects a small portion of the data points with $X > 0$ (with the chosen value of 0.994 for the threshold z parameter only some 10% of the samples with the smallest values were censored). Changing the value of z from 0.9999 to 0.98 only changes the values of the estimated parameter by a few percent.

A second complication here is that we do not know in advance which parameter estimation method (in conjunction with the censorship) necessary to determine $\underline{C}$ in Equation (2.35) will guarantee a linear relationship. Since the mathematics of the problem seem intractable we tackle the problem by adding a simulation (Monte Carlo) procedure as a reference. The procedure is as follows:

(1) For each measured CDF we perform the parameter estimation (and the normalisation to produce V). The chosen estimation procedure used for the CEX, SGM and the CLN models are:

Model	Parameter estimation method
CEX	Gauss-Markov (BLUE)
CLN	Maximum Likelihood Method (modified)
SGM	Non-linear regression
GM	Non-linear regression

Table 2.3. Chosen parameter estimation methods for the normalisation tests.

Note we have investigated various parameter estimation methods, the above methods were the ones giving the best results. The Gauss-Markov method [*Scheffé*, 1959], is a so called Best Linear Unbiased Estimator (BLUE) method but requires the expression for the correlation matrix between the rank-ordered samples, we have only been able to derive the expressions for this matrix for the CEX model, see also [*Balakrishnan & Rao*, 1997]. Description of the Maximum Likelihood method can be found in most textbook on statistics, e.g. [*Kreysig*, 1970]. For our application to conditional distributions we have to slightly modify this method. The

modification is based on the separation of the data in two populations, see Equation (2.34). For the population with the higher values their contribution to the likelihood function is done in the usual way (using the probability density function). For the other population, that with the smallest value, the contribution to the likelihood function of each point is taken as the probability of observing a point with value smaller than the threshold value separating the two populations.

(2) Using the obtained model parameters, produce through simulation (appropriate to the assumed model) 10 CDF's for each measured CDF (here again we have used the MATLAB programming environment).

(3) Apply the same procedure as in step (1) above on the CDF's of the simulated data resulting in simulated V.

(4) Comparison of the cumulative distribution of the measured V and the simulated V indicates the merits of the model.

The advantage of this method is that we do not need the guarantee of linearity of the expected cumulative distribution of V. Furthermore, we are performing a valid statistical procedure: we are comparing the measured behaviour against the expected behaviour (even though we are using simulations instead of rigorous mathematical analysis). An objection that may be raised is that the chosen parameter estimation may be so biased that the 'ensemble' of the simulated data is very different from that which should be tested. On the other hand, a model with no corresponding reliable parameter estimation method is untestable and therefore quite useless.

The result is given in Figure 2.13. Here it can be seen that the SGM model gives the best performance with very good linearity in the range $0.05 < V < 1$. Next comes the CEX model the range with good linearity for $0.15 < V < 1$. The performance of the CLN model is not so good; a breakpoint can clearly be seen at $V = 0.5$. The GM model performs badly.
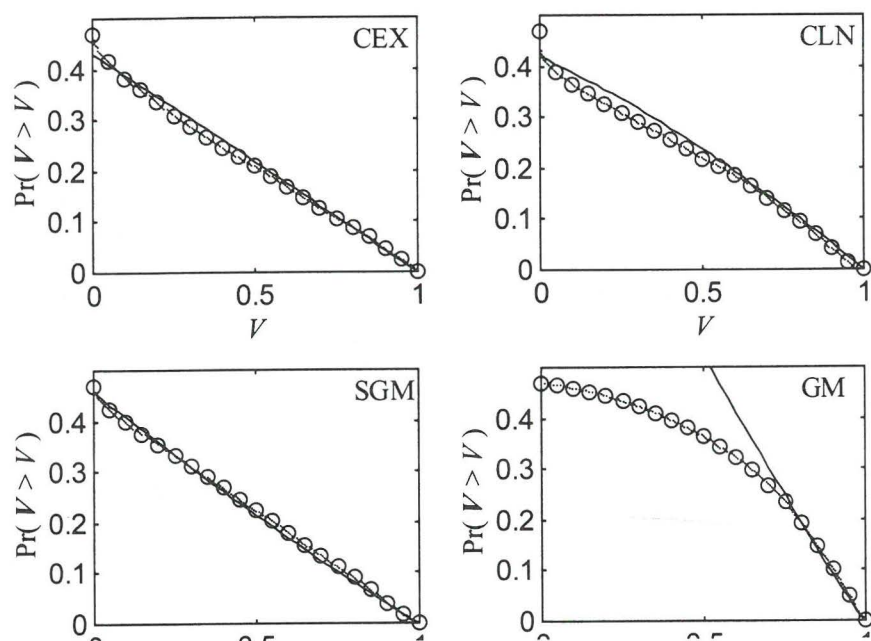

Figure 2.13 Results of the Normalisation tests (solid lines = simulated, circles = measured).

Table 2.4 summarises the performance of each model in terms of the mean and the root mean square (RMS) of natural logarithm of the ratio of observed to model (simulated) probability distribution (restricted to non-zero values of V):

	CEX	CLN	SGM	GM
Mean (%)	-0.6	-12.5	3.4	-33.7
RMS (%)	7.8	27.4	5.5	43.7

Table 2.4. Results of the normalisation tests

Finally, we note here that the CLN method using Crane's parameter estimation method produces somewhat poorer results, mean = -5.3% and RMS = 31.9%. Furthermore, the simulated results for the CLN model (for both Maximum Likelihood as well as *Crane*'s regression method) deviate markedly from the expected linear, which indicates that there may be a problem with the CLN in conjunction with censorship. To see this we have also tested a modified CLN procedure; here we take the full uncensored data and take C_o to be equal to the total number of points with $X > 0$ divided by the total number of samples. The result, given in Figure 2.13 (indicated as CLN-modified), shows a great degradation of performance, only now the simulated data are very linear (as they should be).
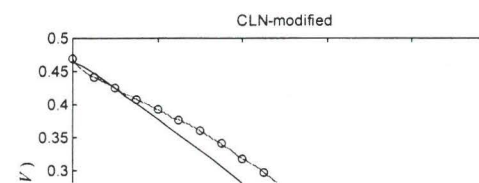

Figure 2.14 Results of normalisation test on the CLN model using "uncensored" data in the parameter estimations.

2.5 Interpretation of the model parameters

The mathematical expression for the dependence of the long-term average $\overline{X}$ of X on the model parameters has the general form of:

$$\overline{X} = C_0 \, C_1 \, k_{model} \tag{2.36}$$

where k_{model} is equal to 1 for the CEX model, while k_{model} is a function of C_2 for the SGM model and for the CLN model (furthermore for the SGM model C_o is an effective value which depends on C_2). Detailed expressions, have been derived and are given in Appendix 2.2.

An important point to note here is that if P_a is known, e.g. by using good prediction methods, the number of independent parameters required to determine P_X is reduced by one. For the CEX and SGM model we would therefore in addition to a prediction method for P_a only need to determine C_0 in order to fully predict P_X; for the CLN we need C_2 as well in addition to the two aforementioned parameters. The parameter C_1 can then be determined from:

$$C_1 = \frac{\overline{X}}{C_0 \, k_{model}} \approx \frac{P_a}{C_0 \, k_{model}} \tag{2.37}$$

In a study carried out by [*Mawira*, 1980] on the behaviour of C_o for rain-attenuation on a 12 km microwave link at 11 GHz and that of the C_o for simultaneously measured rainfall intensity it has been found that if the C_o is plotted against the measured long-term value $\overline{X}$ at the same threshold levels the plots C_o vs. $\overline{X}$ for both the rain attenuation and that for rainfall intensity falls very close to each other. Later studies carried out by the author for the COST205 project confirms such similarity between the C_o vs. $\overline{X}$ for slant path attenuation in various countries. For slant-path crosspolarisation such a similarity was also observed albeit with a tendency for the cross-polarisation to produce slightly lower values of C_o; this has also been observed by [*Fukuchi et.al.*, 1985]. For ducting effects, we also have evidence of such an invariance

relationship through comparison of the C_o vs. $\overline{X}$ relationships produced by different radio links [COST210, 1990]. Note that in all these studies the CEX model was assumed. We do not expect a different situation to emerge if the SGM or the CLN model has been used since these models has produced very similar values for C_o as the CEX model for the data used in the tests carried out in the previous sections.

This observation seems to warrant the interpretation of C_o as a meteorological quantity related to the occurences of meteorological conditions conducive to their associated propagation effects. From this point of view it makes sense in the future to develop climatic maps of C_o for the important time percentages. C_1 on the other hand is strongly related to the path link configuration (frequency, path length, elevation angle, polarisation etc.) so that mapping of C_1 would not be an efficient approach. As noted above, however, C_1 can be determined from C_o and P_a (which can be determined using current prediction models).

Since none of the proposed models can with certainty be excluded it may be expected that various models will stay in use. From this point of view it would be useful to determine the interrelationship between the parameters of the different models. The intention here is that if we are given the values of the model parameters (C_0, C_1 etc.) obtained from fitting a certain model to a particular given set of measurement data, we would then be able to use these parameter values to determine the parameter values that would have been obtained when assuming another model. Using the data described before, we have determined empirically such relationships, the results are shown in Figures 2.15, 2.16 and 2.17.

Figure 2.15 shows the observed relationship between the location parameter C_0 associated with CEX model and the shape parameter C_2 associated with the SGM model. The figure also show that the relationship is reasonably well approximated by the following (simple) equation:

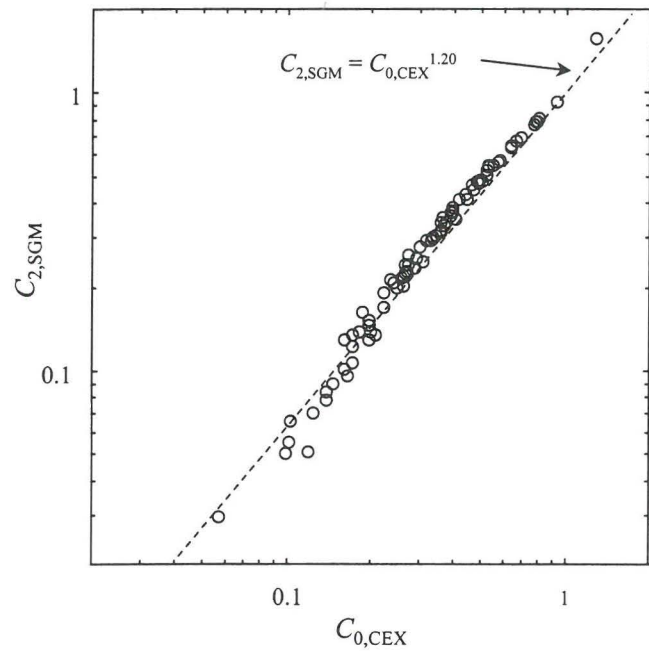$$C_{2,SGM} = C_{0,CEX}^{1.20} \tag{2.38}$$

Figure 2.15. C_2 (SGM model) vs. C_0 (CEX model),

circles = measured, dashed-line = approximation.

Figure 2.16 shows the observed relationship between the location parameter C_0 (associated with the CEX model) and the shape parameter C_2 (associated with the GM model). Again it is found that the relationship can be well approximated by a simple power law:

$$C_{2,GM} = C_{0,CEX}^{1.45} \tag{2.39}$$

For the CLN model only the location parameter gives a consistent relationship with respect to the location parameter of the CEX model. The results are shown in Figure 2.17. A reasonable approximation is given by the following equation:

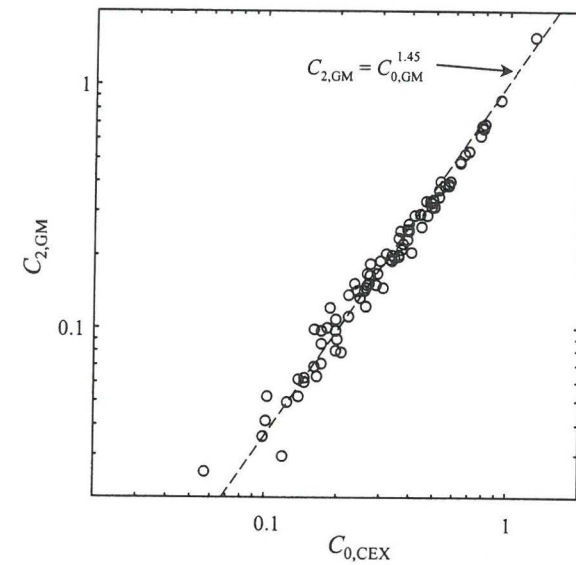$$C_{0,CLN} = 1 - \exp(-(C_{0,CEX}/0.6)^{1.35}) \tag{2.40}$$

Figure 2.16. C_2 (GM model) vs. C_0 (CEX model),
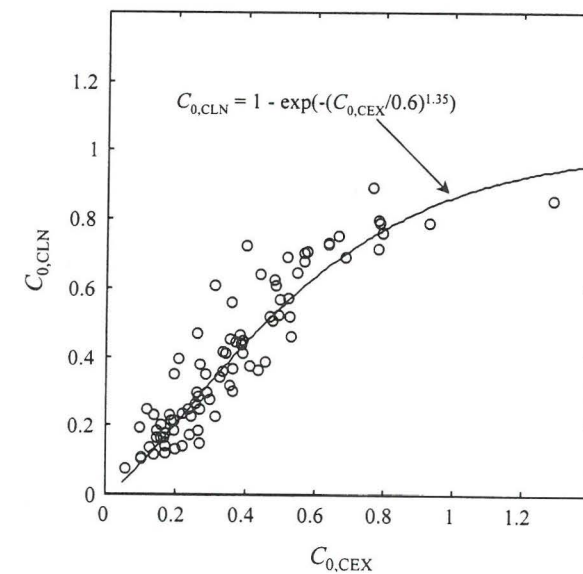circles = measured, dashed-line = approximation.

Figure 2.17. C_0 (CLN model) vs. C_0 (CEX model),
circles = measured, dashed-line = approximation.

2.6 Conclusions

The properties of the random process X have been investigated. The cyclostationary nature of X has been demonstrated. Furthermore the statistical independence of X has been demonstrated (at least for lag times larger than one month). A simplified model based on the aggregate parent distribution P_X and the feasibility of using the simplifying dual-population model has been demonstrated. Four models for P_X have been tested using some 92 site years of data from 12 locations in Europe (slant-path attenuation and trans-horizon signal levels). The models are the conditional exponential (CEX), the conditional lognormal (CLN), and the shifted gamma (SGM) models and the Gamma (GM) model. Classical test methods, for example the χ^2 minimum method and the regression method, do not allow definite statements regarding superiority/inferiority of these models. This is due to the limited number of sample points available for each sample distribution, and the similarity of the three distributions in the midrange. Application of a novel normalisation method (to create a "grand pool" of test data) has indicated the superiority of the SGM model, with the CEX model a close second. The CLN performs rather poorly, while the GM method performs badly.

3. The random process Y of the yearly time fraction of excess

3.1 Introduction

The yearly observation period is also considered to be an important reference period in telecommunication applications. This is because considerable year-to-year variations generally exists with regard to atmospheric propagation effects [COST205, 1985]. Information on the variability of single-year cumulative distributions with respect to the expected long-term distribution is therefore a useful supplement to other information (e.g. worst-month) used by the system designer. Moreover, by using the variability information the dependence between the accuracy of the measured distributions with the length of observation time can be established. This is of importance to system designers when deciding whether a given set of measured data may be used to reliably predict expected behaviour of the system. For propagation experimenters this information is useful when planning the minimum required duration of propagation experiments. In this section we shall study the modelling of the behaviour of the yearly time fraction of excess (TFE). Our approach is to base such modelling on the models for the random process X of the monthly TFE presented in the previous section. This approach is thought to be preferable over the purely empirical modelling of variability of yearly TFE, although it must be admitted that in some cases only the yearly TFEs are available.

In a similar way as with the monthly TFE we consider the discrete time series formed by the fractional durations Y_j in each year that the relevant propagation parameter $\boldsymbol{a}$ exceeds a threshold value a:

$$Y_j = \text{fraction of time that } \boldsymbol{a} \geq a \text{ within year nr. } j \tag{3.1}$$

as a realisation from a random process Y. In the same way as with the random process X this random process Y is considered to represent the family (ensemble) of time dependent outcomes of all possible (hypothetical) radio links with equal link parameters located in a specific (statistically homogeneous) geographic zone.

The relationship between the random process of the yearly TFE Y and the random process of the monthly TFE X is given by:

$$Y(j) = \sum_{m=1}^{12} \frac{D_m}{365} X(m + 12j) \tag{3.2}$$

where m denotes the month index, j denotes the year index, and D_m is the number of days in month m.

Ignoring the variations in the number of days in a month and the complication due to a leap year the following approximation is obtained:

$$Y(j) \approx \frac{1}{12} \sum_{m=1}^{12} X(m + 12j) \tag{3.3}$$

In this section we shall assume this approximative relationship throughout. The maximum error of 8% (which may happen if all the propagation events occur only in February) is, as we shall show later on, much smaller than the natural variation of Y.

Equation (3.3) can also be expressed in terms of the 12 sub-processes X_m of X associated with the calendar months m:

$$Y(j) = \frac{1}{12} \sum_{m=1}^{12} X_m(j) \tag{3.4}$$

In Section 2 we have postulated the independence of the random process X; this leads to the independence of the random process Y. Furthermore since we have assumed that X is (annually) cyclostationary Y is stationary, [*Papoulis*, 1984]. Finally we note that:

$$< Y > = P_a(a) \tag{3.5}$$

where $P_a(a)$ is the probability of $\boldsymbol{a}$ exceeding the threshold value a.

3.2 The coefficient of variation

For a first-order analysis or for getting a rough impression of the variability of Y it is often sufficient to consider the coefficient of variation Ω_Y of Y instead of its statistical distribution. The coefficient of variation of any random variate is defined as the ratio of its standard deviation to its mean, so that Ω_Y is defined as:

$$\Omega_Y = \frac{\sigma_Y}{< Y >} \tag{3.6}$$

where σ_Y is the standard deviation of Y. For calculating the coefficient of variation we need to express the first and second statistical moments of Y. in terms of the moments of X. The general expression for these moments, for the bi-population model discussed in Section 2, are as follows:

$$<Y> = \frac{M}{12} <X|\ X \in Active>$$ (3.7)

$$<Y^2> = \frac{M}{12^2} <X^2|\ X \in Active> + \frac{M^2 - M}{12^2} (<X|\ X \in Active>)^2$$ (3.8)

where M is the seasonality index (see Section 2.2). Using these two equations we obtain the following expression for the variance of Y:

$$\sigma_Y^2 = <Y^2> - <Y>^2 = \frac{M}{12^2}(<X^2|\ X \in Active> - <X|\ X \in Active>^2) = \frac{M}{12^2} \sigma_{X \in Active}^2$$ (3.9)

and subsequently by combining Equations (3.6), (3.7) and (3.9) we obtain the following expression for the relationship between the coefficient of variation of Y and the coefficient of variation of the active population of X:

$$\Omega_Y = \frac{1}{\sqrt{M}} \frac{\sigma_{X \in Active}}{<X|\ X \in Active>} = \frac{1}{\sqrt{M}} \Omega_{X \in Active}$$ (3.10)

With this expression and the statistical models for X we can easily derive the expression of the coefficient of variations of Y for the various models presented in Section 2.

For the CEX model we have previously derived the expression for Ω_Y and shown that it is only dependent on C_0 and M [Mawira, 1980]:

$$\Omega_Y = \sqrt{\frac{1}{6C_0} - \frac{1}{M}}$$ (3.11)

for $0 < C_0 \leq 1$ and $12C_0 < M \leq 12$. For the case of $C_0 > 1$, using the extended CEX model, introduced in Section 2, we can easily show that:

$$\Omega_Y = \frac{1}{\sqrt{12}} \frac{1}{1 + \ln(C_0)}$$ (3.12)

The upper bound for Equation 3.11 is obtained for $M = 12$, in this case:

$$\Omega_Y = \sqrt{\frac{1}{6C_0} - \frac{1}{12}}$$ (3.13)

The lower envelope of Equation 3.11 is obtained for the special cases with $C_0 = M/12$, in this case:

$$\Omega_Y = \sqrt{\frac{1}{12C_0}}$$ (3.14)

For the GM model the expression for the coefficient of variation is easily derived:

$$\Omega_Y = \frac{1}{\sqrt{C_{2Y}}} = \frac{1}{\sqrt{12C_2}}$$ (3.15)

(Note: for the GM model we restrict ourselves to the case $M = 12$; a single population model).

For the SGM model the exact expression for the coefficient of variation is rather cumbersome, see Appendix A3.1. However a simple and accurate approximation has been found:

$$\Omega_Y \approx \frac{1}{\sqrt{12C_2}} + 0.15(1 - C_2)H(1 - C_2)$$ (3.16)

with H the Heaviside unit step function. As in the case for the GM model we also restrict ourselves here to the case $M = 12$.

For the CLN model the expression for is easily derived. Here we have (with $M = 12$):

$$\Omega_Y = \frac{1}{\sqrt{12}} \sqrt{\frac{e^{C_i^2}}{C_0} - 1}$$

(3.17)

In the following we shall investigate the relationship between the theoretical values for the coefficient of variation and the observed values.

Figure 3.1. Observed relationship between Ω_Y and C_0 (for CEX model). The left graph is for rain-attenuation data, the right graph is for ducting data. Solid and dotted lines are the expected theoretical dependence.

Figure 3.1 show the observed relationship between Ω_Y and C_0 for the CEX model. In this figure are shown the two theoretical curves of Equations (3.13) and (3.14) associated with upper and lower bounds for Ω_Y. The upper bound is that associated with $M = 12$. The lower bound is that obtained for the special cases that $C_0 = M/12$ (with $M = 1, 2, \ldots, 12$). We can see from this figure that most of the slant-path data falls neatly within the two curves. On the contrary the ducting data exhibit in general a much larger value for Ω_Y than theoretically predicted. The discrepancy can be as large as 50%.
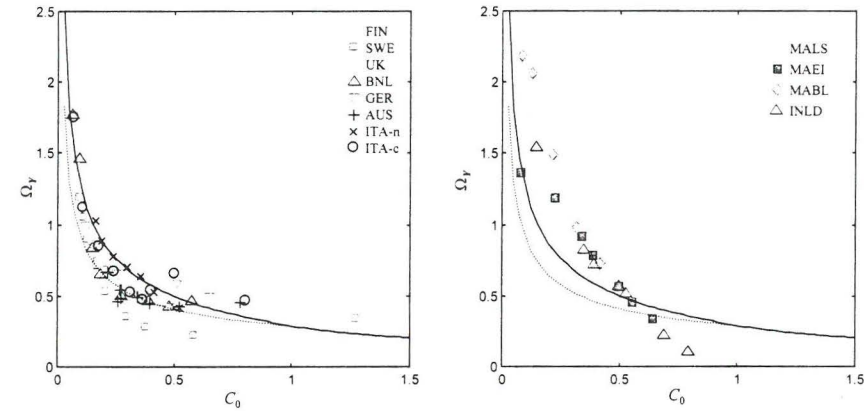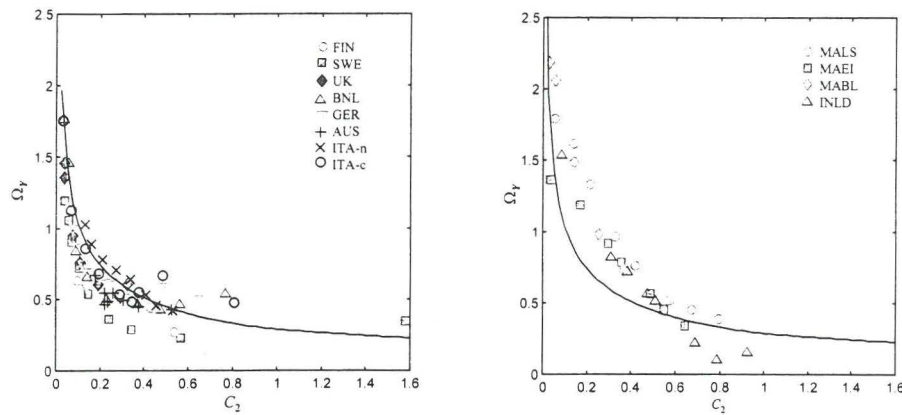
Figure 3.2. Observed relationship between Ω_Y and C_2 (for SGM model). The left graph is for rain-attenuation data, the right graph is for ducting data. Solid and line gives the expected theoretical dependence.

For the SGM model we investigate the relationship between Ω_Y and the shape parameter C_2. The results are shown in Figure 3.2. In this figure the theoretical curve for $M = 12$ is shown as well. As in the case for the CEX model we see a rather that the theory provides a good fit for the rain-attenuation data and a strong underestimation for the case of ducting. Further, comparing the results of rain-attenuation in Figure 3.2 with that of Figure 3.1 we may conjecture that here we may obtain a better fit by using another value of M instead of fixing it to 12. However, this slight complication of the model will not solve the discrepancy for the ducting data.

The result for the GM model is very similar to that of the SGM model and not shown here. For the CLN model Ω_Y depends on two parameters, the location parameter C_0 and the shape parameter C_2, so that a simple graphical dependence as in the case of the other cases cannot be given. Moreover the correspondence between the observed value of Ω_Y and that predicted using Equation 3.17 is extremely bad so that it is not worthwhile to go into too much details here. We suffice here with showing the plots of the predicted versus the measured Ω_Y, see Figure 3.3. This figure shows that often very extreme overestimation of Ω_Y occurs; in some cases the theory overpredicts by a factor of 10x. This overestimation is related to the very strong dependence of

Ω_Y on C_2 in Equation 3.17, where we have a term containing the exponent of the square of C_2. Admittedly errors in the parameter estimation method may also aggravate the situation (we have used here the method recommended by [*Crane,* 1992]).
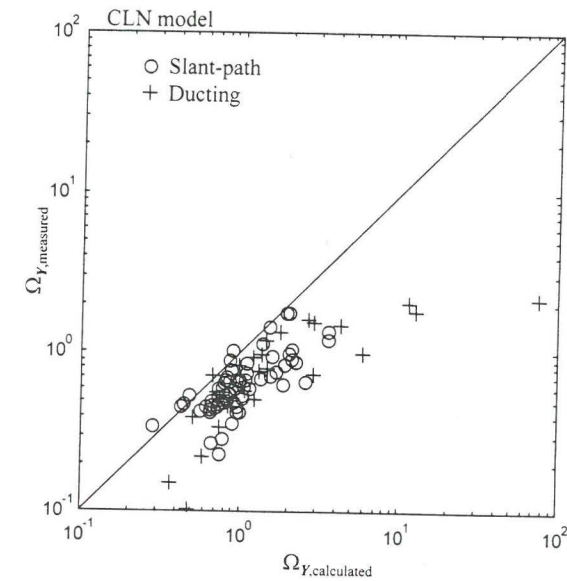
Figure 3.3 Calculated versus measured Ω_Y (CLN model).

Finally we note that the previous figures indicate that there is little difference between the CEX model and the SGM model in terms of the deviations between observed behaviour and predicted behaviour. More striking is the difference in the behaviour between the rain-attenuation data and the ducting data. While for the rain-attenuation data the model predicts the values of Ω_Y quite well, in the case of the ducting data the models rather strongly underestimate the values of Ω_Y. A speculative explanation may be that for the cases of ducting that the model of simple annual cyclostationarity of the processes is not fully correct.

3.3 The statistics P_Y of the annual TFE

3.3.1 Introduction

According to the theory of probability [*Papoulis*, 1984] the probability distribution P_Y of Y is related to the probability density functions f_{Xm} of X_m through the following integral relationship:

$$P_Y(Y) = \int .. \int_{X_1+...+X_{12}>12Y} dX_1 dX_2 .....dX_{12} \, f_{X_1}(X_1).........f_{X_{12}}(X_{12}) \qquad (3.18)$$

where we have made use of the independence of X_m's as discussed in Section 2. In general this integral is difficult to solve. However for the CEX and GM cases exact analytical expressions for P_Y can be derived by using the well known fact that the sum of N independent and identically distributed (*i.i.d.*) Gamma distributed random variates results again in a Gamma distributed random variate with however the shape parameter having a value N times the value of the shape of the N components (the scale parameter stays the same). For the SGM model a simple approximation can be derived. For the CLN model the situation is quite difficult, there is no closed form solution, and although approximation formulas has been developed [*Yeh & Schwartz*, 1984] they are quite cumbersome to use.

3.3.2 The CEX model

For the CEX model we have derived the analytical expression for P_Y before within the work done in project COST205, see pp 192-193 [*Fedi*, 1985]:

$$P_Y(Y) = \sum_{m=1}^{M} \binom{M}{m} (\frac{12}{M}C_0)^m (1-\frac{12}{M}C_0)^{M-m} P_{GM}(12C_0 \frac{Y}{<Y>},n) \qquad (3.19)$$

for $0 < C_o < 1$ and $12C_o < M < 12$. $P_{GM}(z,\nu)$ is the standard Gamma function with shape parameter ν:

$$P_{GM}(z,\nu) = \int_{z}^{\infty} dt \frac{t^{\nu-1}}{\Gamma(\nu)} e^{-t} \qquad (\text{for } z > 0) \qquad (3.20)$$

$P_{GM} = 1$ for $z < 0$.

For the case $C_o > 1$, i.e. the extended CEX model introduced in Section 2, the expression for P_Y can be easily derived by noting that in such cases X is a sum of a constant term and purely exponential stochast

$$X = C_1 \ln(C_0) + C_1 x \qquad (3.21)$$

Here x represents a standard exponential random variate. Further elaboration yields the following expression for P_Y:

$$P_Y(Y) = P_{GM}(12[(1+\ln(C_0))\frac{Y}{<Y>} - \ln(C_0)],12) \qquad (3.22)$$

for $C_0 > 1$.

3.3.3 The GM model

For the GM model Y is also Gamma distributed:

$$P_Y(Y) = P_{GM}(\frac{Y}{C_{1Y}}, C_{2Y}) \qquad (3.23)$$

with shape parameter C_{1Y} and scale parameter C_{2Y}:

$$C_{1Y} = C_1/12, \quad C_{2Y} = 12 C_2 \qquad (3.24)$$

where C_1 and C_2 are the scale and shape parameters associated with X. Since

$$<Y> = C_{1Y}C_{2Y} = C_1 C_2 \qquad (3.25)$$

we can also express Equation (3.23) as:

$$P_Y(Y) = P_{GM}(C_{2Y}\frac{Y}{<Y>}, C_{2Y}) = P_{GM}(C_{2Y}\frac{Y}{P_a(a)}, C_{2Y}) \qquad (3.26)$$

This formulation shows that once P_a is known only one additional parameter (C_2 or C_{2Y}) is needed in order to fully characterise the distribution P_Y.

3.3.4 The SGM model

For the SGM an analytical solution for P_Y is not found. However we have found that the following SGM type of formula gives a good approximation for P_Y:

$$P_Y(Y) \approx P_{GM}(\frac{Y}{C_{1Y}} + \kappa_Y, \ C_{2Y}) \qquad (3.27)$$

where C_{1Y} and C_{2Y} are the scale and shape parameters as given by Equation (3.24), κ_Y is the shift parameter:

$$\kappa_Y = C_{2Y}\sqrt{C_{2Y}/12} \ (1 - \sqrt{C_{2Y}/12}) \qquad (3.28)$$

This approximation has been obtained through analysis of simulation data, see Appendix A3.2.

For the SGM model the relationship of P_Y to P_a can also be determined by first noting that:

$$P_a(a) = <Y> = <X> = C_1[C_2 \ P_{GM}(\kappa, 1+C_2) - P_{GM}(\kappa, C_2)] \qquad (3.29)$$

Subsequently, by using Equation (3.24), and eliminating C_{1Y} we arrive at the following expression for P_Y:

$$P_Y(Y) \approx P_{GM}(\eta C_{2Y}\frac{Y}{P_a(a)} + \kappa_Y, \ C_{2Y}) \qquad (3.30)$$

with

$$\eta = P_{GM}(\kappa, 1 + C_2) - \frac{\kappa}{C_2}P_{GM}(\kappa, C_2) \qquad (3.31)$$

This formulation shows that also for the SGM model once P_a is known only one additional parameter (C_2 or C_{2Y}) is needed in order to fully characterise the distribution P_Y.

3.4 Accuracy in the characterisation of long-term statistics

One of the important questions in experimental radiowave propagation studies is that concerning the volume of data which needs to be collected to obtain accurate estimate of P_a (or $<Y>$). The models presented in the previous sections allow us to evaluate the dependence of the accuracy on the total length of measurement time. Another way is increasing the data size by measuring simultaneously at many locations. The impact of multiple spatial sampling on the accuracy of the estimate of the long-term statistics is not very well known. This aspect shall be discussed at the end of this section; first we consider the influence of length of measurement time on the accuracy.

An N-year measurement of the random variable Y results in the unbiased estimate Z_N:

$$Z_N = \frac{1}{N}\sum_{j=j_1}^{j_1+N-1} Y(j) \qquad (3.32)$$

for P_a (or equivalently $<Y>$). A natural measure of the accuracy of Z_N as estimate of $<Y>$ is the coefficient of variation of Z_N:

$$\Omega_{Z_N} = \sigma_{Z_N} / < Z_N > \qquad (3.33)$$

The dependence of Ω_{ZN} on the total duration N of the measurement can easily be derived by noting that subsequent samples of Y, in Equation (3.32), are independent so that:

$$\Omega_{Z_N} = \frac{1}{\sqrt{N}} \Omega_Y \qquad (3.34)$$

Figure 3.4, taken from [COST205, 1985], illustrates the effect of 5 year averaging on the coefficient of variation associated with rainfall intensity data in North-West Europe and Scandinavia.
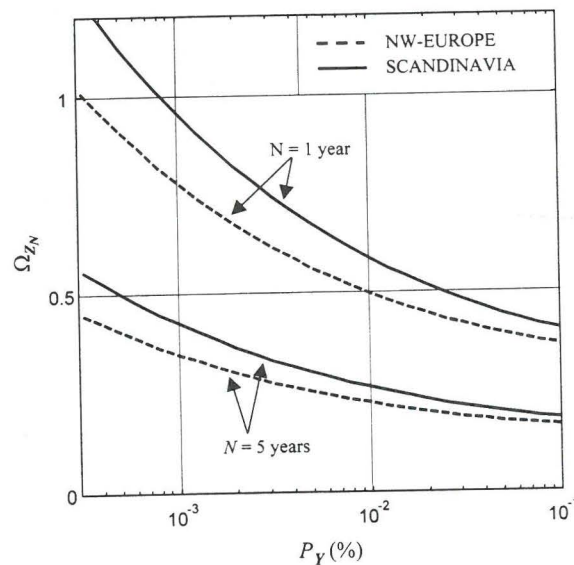

Figure 3.4. Influence of the observation time N on the relationship Ω_{Z_N} versus P_Y.

This figure is obtained using the CEX model and the empirical values for C_0 for the regions of Scandinavia and NW-Europe; this climatic classification will be discussed in Section 3.5. This figure illustrates the problem of characterising atmospheric propagation statistics through direct measurements only. Even with 5 years of measurements the tail of the distribution (< 0.01%) still exhibit large deviations.

Another way to improve the accuracy of estimation of Y is through simultaneous measurements at different locations (spatial sampling). We have analysed this effect previously using data from 7 years of measurements using a network of 6 raingauges located the vicinity of Leidschendam – Netherlands (the average spacing between the raingauges is about 3km), [COST205, 1985]. It was found that the variability of the spaced (network) averaged annual TFE is much less than the variability of individual annual TFE, see Figure 3.5. It is interesting to note that the reduction of the variability only reaches the expected factor of square root of 6 for the high rainfall-rates (respectively low values of P_Y). This is in accordance to the well-known fact that the higher rainfall intensities are concentrated in regions of small dimension (< 3km) so that for such intensities the network in effect produces independent samples.

Finally, we note here that comparison of Figures 3.4 and 3.5 suggests that a single year of measurement using a network of raingauges produces an accuracy comparable to a 5 year of measurements using only a single raingauge.
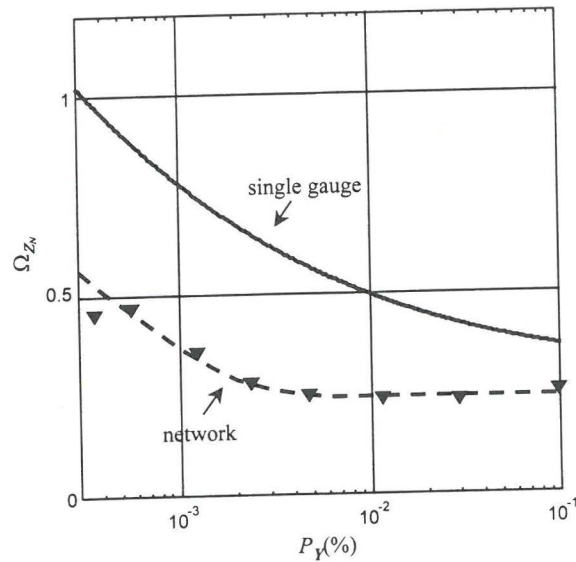
Figure 3.5. Comparison of Ω_{Z_N} obtained by using a single rain-gauge and that obtained through using a network of 6 raingauges.

3.5 Climatic zones for the C_0 versus $<Y>$ relationship

In the COST205 project the author has investigated the observed dependence between C_0 and $<Y>$ or equivalently P_a. The CEX model was assumed throughout in the determination of C_0. through regression on the measurement data. It was found that in general the dependence could be well described by the following equation:

$$C_0 = 1/(1 + \sqrt{\beta / P_a})$$
(3.35)

(Pa and β in the same units)

Furthermore the author identified four distinct zones with differing values of β:

Zone	Typical $\beta(\%)$
North-West Europe	0.03
Mediterranean	0.03
Scandinavia	0.09
Alpine	0.12

See Figure 3.6 (a) to (d), which shows the data and the curve of the above equation. Note that in the right vertical axis of these figures the values of the so-called worst-month quotient Q_1 are given, these values are those calculated from C_0 through the theoretical relationship described by Equation (4.51) in Section 4. We note further here that the chosen mathematical expression of the dependence of C_0 on P_a is in hindsight rather restrictive. Values of $C_0 > 1$ were at that time thought not to occur. We know now from [Poires Baptista et al, 1986] that $C_0 > 1$ can easily occur at the lower thresholds.

A formula, which does not have such a restriction, is the following:

$$C_0 = 0.4 \ln(1 + \sqrt{P_a / \gamma})$$
(3.36)

For small values of P_a both formula will have the same asymptotic expression. In such a case the old equation will converge to:

$$C_0 \rightarrow \sqrt{P_a / \beta} \qquad (3.37)$$

while the new equation will converge to:

$$C_0 \rightarrow 0.4\sqrt{P_a / \gamma} \qquad (3.38)$$

By equating these two expression we obtain the following relationship between β and γ:

$$\beta \approx 6\gamma \qquad (3.39)$$

Figure 3.7 (a) shows the curves associated with the new equation as well as the observed points obtained for the subset of the COST205 data selected for the tests in this thesis. The observed points fall mostly within the curves with γ values between 0.005% and 0.02%.(with the typical value $\gamma = 0.01\%$) The observations from Scandinavia falls closest to the lowest curve with $\gamma = 0.02\%$ while the Netherlands, Belgium and UK data corresponds closely to the upper curve with $\gamma = 0.005\%$. The data from Northern Italy and from Austria falls very close to the lowest curve with $\gamma = 0.02\%$ while the data from central Italy stays between the upper curve ($\gamma = 0.005\%$) and the mid-curve ($\gamma = 0.01\%$). While these results broadly confirm the classification of the previous COST205 studies there are some small differences. The results using the selected pooled data show that Germany warrants a separate classification (with $\gamma = 0.05\%$); in the COST205 studies Germany is put in the same zone as Netherlands, Belgium and UK.

Figure 3.7 (b) show the corresponding results for the ducting data (the COST210 data). The figure shows that data points falls within the curves with $\gamma = 0.02\%$ and $\gamma = 1.5\%$. The results suggest that there may be a dependence of γ with the land-sea distribution of the radio-link. The highest value of γ is obtained on the Martlesham – Blavand link which for 90% an oversea link, while the lowest value of γ is obtained on the Irnsum – Leidschendam link which only for 40%

traverses the (inland) sea.

Finally we note here that the ducting data exhibits values of γ which are much higher than those found for the slant-path data.
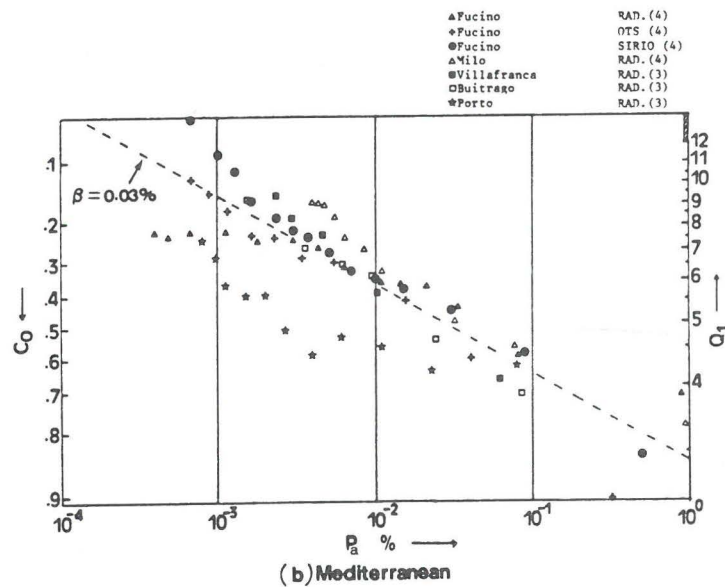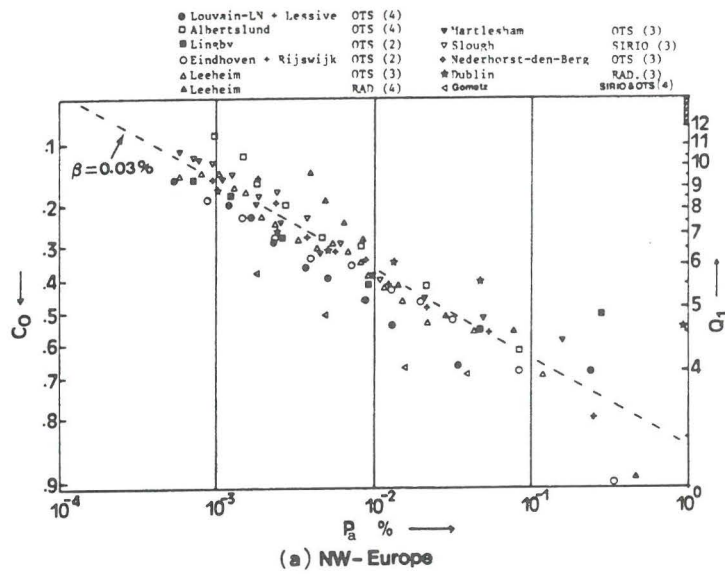
Figure 3.6 (a) (b) Observed dependence of C_0 vs P_a for slant-path rain-attenuation.
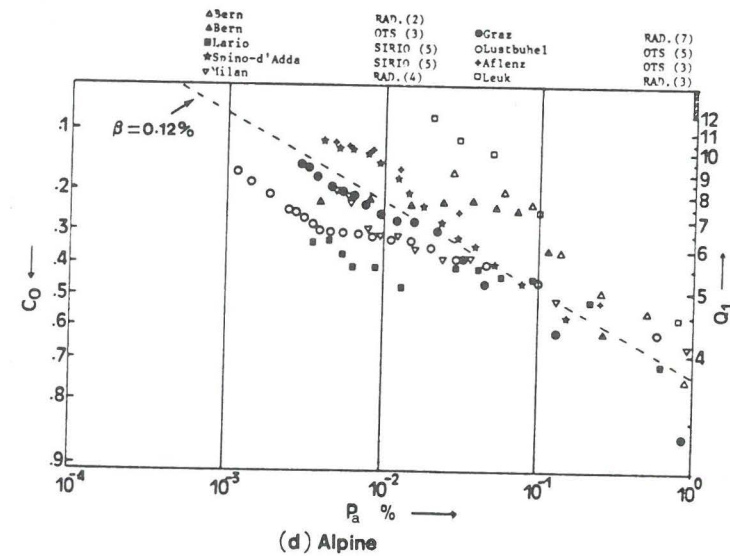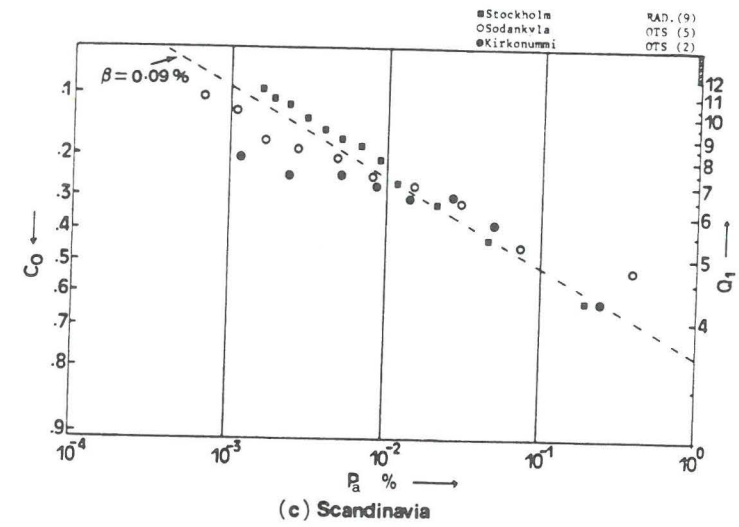
Figure 3.6 (c) (d) Observed dependence of C_0 vs P_a for slant-path rain attenuation.
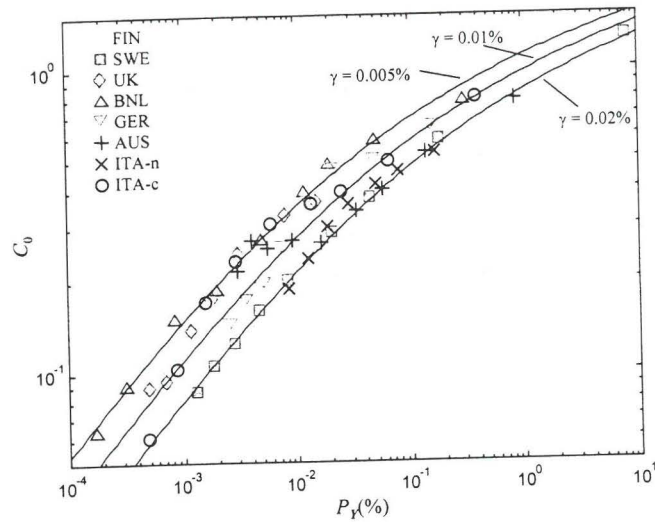
Figure 3.7 (a) C_0 vs P_Y for selected slant-path rain-attenuation (selected COST205 data)
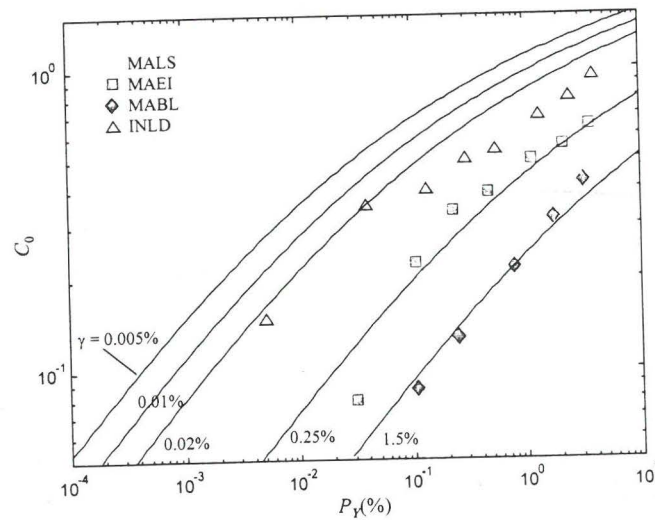

Figure 3.7 (b) C_0 vs P_Y for ducting (selected COST210 data)

3.6 Conclusions

In this section we have considered the feasibility of using the model(s) for the distribution of the monthly time fraction of excess (TFE), presented in Section 2, for the prediction of the variability of the annual TFE. It has been shown that the CEX, GM and SGM models allow reasonably good predictions of the magnitude of the variation of annual TFE around its expected mean (as expressed by the coefficient of variation Ω_Y). Specifically, the predictions performs good in relationship to rain-attenuation data and less well in relationship to ducting data. In this last case the observed values are systematically larger than predicted (the observed values of Ω_Y exceed those predicted by 0 to 50%). The reason for this large discrepancy is not known, it is conjectured that for ducting the assumption of simple annual cyclostationarity of the process is not correct.

The CLN performs here rather badly, resulting often in gross overestimation. This last result seems to be somewhat surprising since the CLN model was found to perform not too badly in the statistical tests of the monthly TFE done in Section 2. It may be that the parameter estimation method used overestimates the values of the shape parameter C_2, which in turn may lead to excessive predictions of the values of Ω_Y. In any way this illustrates the problem associated with a model using many parameters. Another surprising result is the reasonable performance of the GM model. This model was found to produce the worst performance in the tests performed in Section 2. The explanation for this may be that the bad performance of the GM model is strongly related to the deviation of the statistical distribution in the higher probability ranges (or the smaller values of X) but that the model performs well at the tail of the distribution.

The better performance of the two parameter models (CEX, SGM) over the three parameter model (CLN) simplifies life. It implies that in addition to P_a only a single additional parameter, e.g. C_0 for CEX or C_2 for GM and SGM, is needed to fully characterise the variability of the annual TFE.

4. The random process W of the worst-month time fraction of.excess

4.1 Introduction

As mentioned in the introduction, the <u>annual worst-month period</u> is an important reference period used by the ITU for defining a class of grade of service requirements for telecommunication systems. This period is defined as that month in each year having the highest value for the monthly t.f.e. (time fraction of excess). Associated with the annual-worst-month period in a given year j is the annual-worst-month time fraction of excess, W_j:

$$W_j(a) \;=\; \text{Maximum of } \{X_{1,j}(a), \; X_{2,j}(a), \; ..., \; X_{12,j}(a)\} \tag{4.1}$$

where a denotes the chosen threshold value at which the monthly t.f.e. $X_{m,j}$'s of a propagation variate a of interest is determined. This definition uses a data-driven selection procedure, the worst-month period can be different for different years and different threshold values.

Equivalently the time series $W_j(a)$ can be considered to be a particular realisation from a random process W defined as:

$$W(a) \;=\; \text{Ensemble of } \{W_j(a), \text{ for all } j's\} \tag{4.2}$$

that is to say that W is a family of time series, each of whom is obtained by applying the procedure described by Equation (4.1) to each time series of each of the member of the ensemble of monthly time fraction of excess X. Note that a separate random process W is obtained for each separate threshold value. In the following discussions we shall often dispense with explicit notation of a in order to simplify the expressions of the equations.

ITU's Recommendation 581-2 defines the annual worst-month statistics, $P_{a|WM}$, as the long-term average of the annual worst-month time fraction of excess:

$$P_{a|WM} = \lim_{N \to \infty} \frac{1}{2N+1} \sum_{j=-N}^{j=+N} W_j \tag{4.3}$$

Note, in Section 1 we introduced $P_{a|WM}$ in a more intuitive manner as a conditional statistic of a:

$$P_{a|WM}(a) = \text{probability } (a \geq a \mid \text{during the worst} - \text{month}) \tag{4.4}$$

The only difference with the formal definition of ITU, see Equation (4.3), comes from the fact that the number of days in a month is not constant (varying from 28 to 31).

In this section we shall investigate the random process W. We shall find the relationship between the statistical properties of W to the properties of the random process X. It will further be shown that the statistical distribution of W can be accurately determined by using only the aggregate statistical distribution P_X of X. Using these models we subsequently investigate the question of the accuracy of $P_{a|WM}$ when determined from a limited amount of data.

4.2 The statistics of the worst-month t.f.e. W

In Section 2.1 we have assumed annual cyclo-stationarity for the random process X. From this it follows that the statistical properties of W are time independent, i.e. its cumulative distribution function:

$$P_W(W) = \text{probability}(W \geq W) \tag{4.5}$$

is independent of time. Furthermore in Section 2.1 it is assumed that each of the subsets of the random process of X related to a particular calendar month is ergodic. Therefore W is also ergodic so that the long term average $P_{a|WM}$ (or $\overline{W}$) of W is equal to the ensemble average $<W>$:

$$P_{a|WM} = <W> = \int_0^\infty dW \, W \, f_W(W) = \int_0^\infty dW \, P_W(W) \tag{4.6}$$

where f_W is the probability density function of W; the last part of Equation (4.6) is obtained through application of partial integration.

Obviously P_W, the distribution of W, is related to the 12 distributions of X, P_{Xm} ($m = 1, 2, \dots 12$). The expression for P_W can be derived by noting that the probability of $W \geq W$ is equal to 1 minus the probability that all 12 samples of X in each year are smaller than the value W:

$$P_W(W) = 1 - \text{probability}(\Psi_1 \cap \Psi_2 \cap \dots \cap \Psi_{12}) \tag{4.7}$$

where Ψ_m, with $m = 1, 2, \dots 12$, symbolises the event:

$$\Psi_m = (X < W \mid \text{in month } m) \tag{4.8}$$

Assuming independence of the random process X, see Section 2.1, the above equation simplifies to:

$$P_W(W) = 1 - \prod_{m=1}^{12} \text{probability}(\Psi_m) \tag{4.9}$$

or

$$P_W(W) = 1 - \prod_{m=1}^{12}(1 - P_{X|m}(W)) \tag{4.10}$$

The expression for the associated probability density function is:

$$f_W(W) = \sum_{n=1}^{12} f_{X|n}(W) \prod_{\substack{m=1 \\ m \neq n}}^{12}(1 - P_{X|m}(W)) \tag{4.11}$$

Curiously the expression for P_W at the tail (i.e. for large values of W) reduces asymptotically to a rather simple form. We can see this by expanding Equation (4.10):

$$P_W = \sum_{m=1}^{12} P_{X|m} - \sum_{m=1}^{11}\sum_{n=m+1}^{12} P_{X|m}P_{X|n} + O_3 - O_4 \dots + O_{12} \tag{4.12}$$

(O_k represents summations containing k products: $P_{X|m1} P_{X|m2} \dots P_{X|mk}$)

and noting that for large values of W the first term will dominate (since $P_{X|m}(W) \to 0$ as $W \to \infty$):

$$\lim_{W \to \infty} \frac{P_W(W)}{P_X(W)} = \frac{\sum_{m=1}^{12} P_{X|m}(W)}{P_X(W)} = \frac{12P_X(W)}{P_X(W)} = 12 \tag{4.13}$$

In this derivation the defining Equation (2.14) for the aggregate distribution P_X given in Section 2.2 has been used.

An interesting feature to note here is that at the tail, the ratio of P_W over P_X approaches the constant value of 12. Furthermore, no detailed information regarding seasonality (e.g. the distributions $P_{X|m}$) is needed when predicting extreme values of the annual worst-month time

fraction of excess, since in the tail P_W only depends on P_X. Figure 4.1 illustrates such a behaviour, here we have assumed a CEX model with $C_0 = 0.25$ and $C_1 = 0.1\%$.
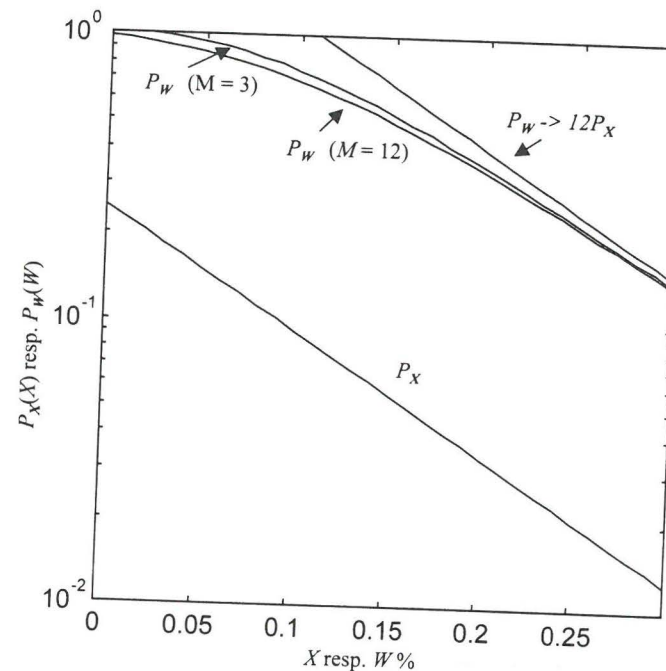

Figure 4.1 Example of convergence of P_W at the tail to $12\,P_X$. CEX model is assumed with $C_0 = 0.25$ and $C_1 = 0.25\%$. Two different population compositions are assumed respectively $M = 12$ and $M = 3$.

This figure gives plots for P_X and for P_W (for the two limiting cases of $M = 3$ and $M = 12$). This constant ratio of 12 can be explained intuitively as follows:

Consider how $P_X(W)$ and $P_W(W)$ are determined empirically from a given time series of observation of X_{ij}'s. For determining $P_X(W)$ the observer would simply count the number of times, N_X, that the value of W is exceeded, this number divided by the total number of months of observations would then result in the estimated value for $P_X(W)$. For determining $P_W(W)$ the observer would count the number of years, N_W, where the worst month t.f.e. exceeds the value W. The resulting number is then divided by the number years of observation. Now, if W is taken to be large enough then the probability of observing more than one month in each year with t.f.e. exceeding W would be negligible, so that in general we shall have $N_X = N_W$. Since the total number of months is 12 x the total number of years this then result in the ratio of 12 for the P_W/P_X.

4.3 Upper and lower bounds for P_W

While at the tail P_W is simply related to P_X, at the other parts of the distribution no simple equality can be expected to exist. However we now shall show that if P_X is given then narrow bounds for P_W can be determined. The lower bound for P_W can be easily found, it is given by the following equation:

$$P_W(W) \geq 1 - \prod_{m=1}^{12}(1 - P_X(W)) = 1 - (1 - P_X(W))^{12} \tag{4.14}$$

or

$$[\prod_{m=1}^{12}(1 - P_{X|m}(W))]^{1/12} \leq (1 - P_X(W)) \tag{4.15}$$

The lower bound is simply obtained by replacing $P_{X|m}$, in Equation (4.10), by the aggregate distribution P_X. To prove this inequality we first eliminate P_W, using Equation (4.10), which after some rearrangement of terms results in the following inequality:

Since $(1 - P_X(W))$ is simply the the arithmetic average of the twelve $1 - P_{X|m}(W)$'s this inequality states that the geometric mean is smaller than the arithmetic mean; this is a proven theorem of mathematics, see Section 3.2.1 of [*Abramovitch & Stegun*, 1975].

An upper bound can also be readily derived using the following well-known inequality:

$$\prod_{m=1}^{N}(1 - p_m) \geq 1 - \sum_{m=1}^{N} p_m \tag{4.16}$$

Taking $p_m = P_{X|m}$ and pplying this inequality with Equation (4.10) we obtain the following:

$$P_W(W) \leq \sum_{m=1}^{12} P_{X|m}(W) = 12P_X(W) \tag{4.17}$$

And since P_W cannot be larger than 1 the upper bound for P_W is given by:

$$P_W(W) \leq \text{Minimum\_of}\{12\,P_X(W), 1\} \tag{4.18}$$

The above bounds were derived under very general conditions, that is to say we have determined the mathematical bounds on $P_W(W)$ when $P_X(W)$ is given. In some cases a tighter bound can be obtained if we are willing to make assumptions regarding the number of months per year which are actually involved in the random processes X and W. This is the case for the CEX and CLN models where we have included the bi-population modelling, see Section 2; for the SGM and GM models we have restricted ourselves to assuming that all 12 months of each year belong to the same statistical process. In Section 2 we have seen that for higher values of the threshold a the average number of months, M_{Co}, having $X > 0$ is often smaller than 12. If we now assume that the propagation process is restricted to an integer number of months equal to:

$$\hat{M}_{C_0} = \text{the first integer} \geq 12\,C_0 \tag{4.19}$$

we can derive another lower bound for P_W using the same method of analysis as above, and obtain:

$$P_W(W) \geq 1 - (1 - \tfrac{12}{\hat{M}_{C_0}}\,P_X(W))^{\hat{M}_{C_0}} \tag{4.20}$$

Figure 4.1 gives an example of behaviour of the bounds for P_W; we have assumed a CEX model for P_X with $C_0 = 0.25$ and $C_1 = 0.1$.

Finally we note here that the upper bound is equivalent with assuming very artificial shapes for the monthly distributions $P_{X|m}(X)$. One of the distributions must have probability value of 1 in the range $X < X_o$ (where X_o is found from $P_X(X_o) = 1/12$), and probability values $12P_X(X)$ in the range $X > X_o$. The other 11 distribution must all have a value 0 for $X > X_o$. We have seemingly a rather 'unphysical' situation where in one of the months we have a 100% probability of observing the event $X > X_o$, while in the months just before and just after it the corresponding probability is 0%; we can consider such a case as an "absolute limit bound".

4.4 An approximation for P_W

In most cases the planners do not have the 12 distributions $P_{X|m}$, instead the planners have only the aggregate distribution P_X. As noted in Section 2.5 P_X can be predicted using a propagation model in combination with statistics of meteorological parameters such as e.g. that of rainfall-intensity and additional general information on seasonality e.g., the conditionality parameter C_o.

The analysis in the previous section on the bounds of P_W shows that P_X contains the necessary information to determine the upper and lower bound of P_W. Furthermore the bounds becomes very tight for large values of W. It seems therefore that the approximative model for the random process X discussed in Section 2, the so called dual-population model based on P_X, might provide a suitable framework for plannings purposes. Here we use the information on P_X to formulate a simplified model for the random process X, the statistical properties of W are then derived in the usual way. In this dual-population model introduced by *Brussaard & Watson* [1979], it is assumed that the propagation process occurs solely M months per year, the "*Active* months*". The exact value of M is not known but its bound can be derived from P_X resp. the conditionality parameter C_o associated with P_X:

$$\hat{M}_{C_0} \leq M \leq 12 \tag{4.21}$$

where $\hat{M}_{C_0}$ is given by Equation (4.19). Furthermore, within the active month the statistics of the propagation is assumed to be homogeneous, i.e.:

$$P_{X|m}(X) = \tfrac{12}{M}P_X(X) \quad (\text{for } m \in Active \text{ months}) \tag{4.22}$$

for all other months $P_{X|m} = 0$.

Application of this simplified model to Equation (4.10) leads to the following expression for P_W:

$$P_W(W) = 1 - (1 - \tfrac{12}{M} P_X(W))^M \qquad (4.23)$$

The expression for the associated probability density function is:

$$f_W(W) = 12 f_X(W)(1 - \tfrac{12}{M} P_X(W))^{M-1} \qquad (4.24)$$

Although these equations entail major simplifications, for the important cases of large values of W this approximation tends asymptotically to the exact value of the distribution for W. Expansion of the above equation for large values of W (respectively small values of $P_X(W)$) leads to:

$$P_W(W) \rightarrow 1 - (1 - M P_X(W)^{\tfrac{12}{M}} + ...) = 12 \ P_X(W) \qquad (4.25)$$

So we can see that the approximation form for P_W converges, in the tail, to that of the asymptotic form, Equation (4.13), obtained when using the exact expression given by Equation (4.10).

Another interesting feature of this approximation is that for $M = 12$ and respectively $M = \hat{M}_{Co}$ the approximation produces the lower respectively the upper-lower bound for P_W, see Section 4.3. Since this approximation converges in the tail to the true value and furthermore these two bounds are likely to be close to a physical realistic upper bound in the mid-range the simplified model can be used by the planner to evaluate worst-month behaviour with reasonable accuracy.

4.5 Prediction of $P_{a|WM}$ using ITU's worst-month quotient Q_1

In many of ITU's prediction methods the long-term average of W (which is equal to the ITU's $P_{a|WM}$) is calculated from the long-term statistics P_a (approximately equal to the long term average $\overline{X}$ of X) through the so called worst-month quotient Q_1:

$$<W> = P_{a|WM} = Q_1 P_a \approx Q_1 \overline{X} \qquad (4.26)$$

The reason for doing this is that most prediction methods in the past were developed for the prediction of P_a. This is necessary because most of the meteorological information is only available as long term statistics. For example global maps of statistics P_R of rainfall intensity **R** may be available but the meteorological institutes have never bothered to produce worst-month statistics of **R**. Furthermore, the empirical data suggest that the parameter Q_1 is not very sensitive to the radio link configuration parameter (path length, frequency, polarisation etc.) but seems related to more general climatic properties of various meteorological phenomena. The prediction methods for P_a are, therefore, supplemented by the ITU with tables/formulas of Q_1 for different propagation phenomena and for several climatic zones. This suffices to produce reasonably accurate predictions of $P_{a|WM}$.

For theoretical analysis it suffices to use the following expression for Q_1:

$$Q_1 = <W> / \overline{X} \qquad (4.27)$$

The dependence of Q_1 on the parameters of the three models considered in this thesis has been determined numerically. The most important part of the calculation is that which deals with the calculation of $<W>$ using Equation (4.6) and (4.29). An efficient expression for Q_1 for the CEX model has been derived before by *Brussaard & Watson* [1979] through expansion of Equation (4.29) and solving the integrals. No such simplification can be obtained for the CLN and SGM models, in these cases numerical integration is directly applied to Equation (4.29). In Appendix 4.1 the equations for Q_1 are given.
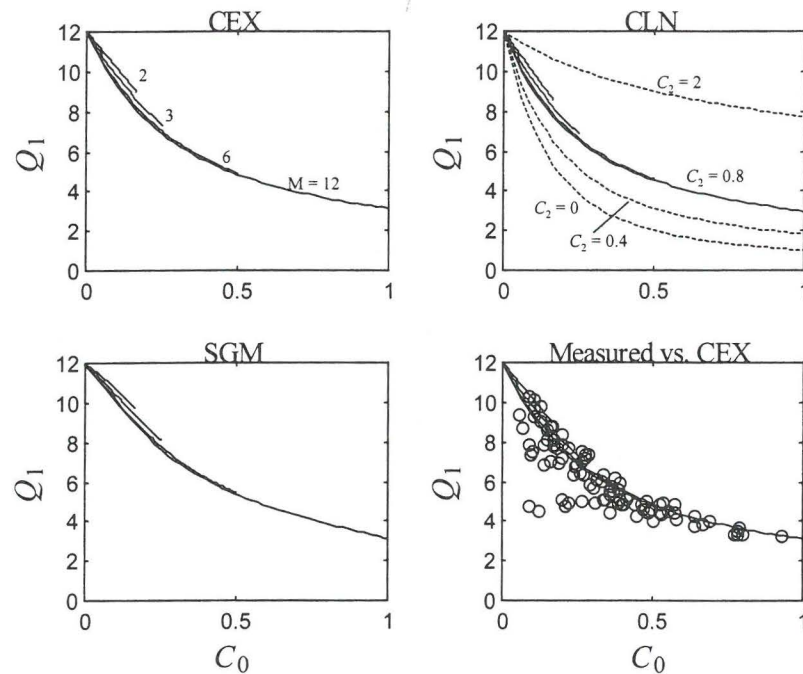
Figure 4.2 Q_1 versus. C_0 relationship. Lines give the theoretical relationships
according to various models. Markers (o) indicate measured values.

The results are plotted as a function of C_0 in Figure 4.2 (for the SGM model we use the effective C_0 defined in Section 2). Note that Q_1 does not depend on the scale parameter C_1. Also shown in this figure is the measured dependence of Q_1 on C_0, whereby the parameter C_o was determined using the *method of moments* and assuming the CEX model; see Section 2.4 for discussions on parameter estimation methods. As can be seen, except for a few points the three models predict the relationship reasonably well. Note that the observed spread of measured Q_1 vs. C_o, in Figure 4.2, also contains the uncertainty due to the estimation of C_0. This uncertainty becomes quite large for small values of C_0 since very few points with $X > 0$ are available there for reliable parameter estimation.

In general the CLN model permits the largest dispersion range, the upper bound is achieved when the shape parameter $C_2 = \infty$:

$$Q_{1,CLN}(C_0, \infty, M) = 12$$

(4.28)

the lower bound is achieved when the shape parameter $C_2 = 0$ and $M = 12$:

$$Q_{1,CLN} = (1 - (1 - C_0)^{12}) / C_0$$

(4.29)

Finally, we note that for all models the influence of the seasonality index M on Q_1 is, in general, not so large. On the whole the theoretical uncertainty in Q_1 due to M is within the range of $\pm 5\%$. In the next Section 4.6 we shall show that the uncertainty in experimental determination of Q_1 due to limited observation time is much larger than this.

4.6 Variability of the short term observed values of Q_1

In practical applications Q_1 values were determined from measurements spanning a limited observation period. For a particular sample of the random process X a limited observation spanning N years, from e.g. year $j-N+1$ to year j, the estimated value, $q_N(j)$, of Q_1 is given by:

$$q_N(j) = \left(\frac{1}{N} \sum_{j'=j-N+1}^{j} W_{j'}\right) \bigg/ \left(\frac{1}{N} \sum_{j'=j-N+1}^{j} \frac{1}{12}\sum_{m=1}^{12}(1+\delta_m) X_{m'j}\right) \tag{4.30}$$

where δ_m gives the fractional deviation of the number of days in a particular month to the average number of days in a month (e.g. for February we have $d_2 = 28/(365/12) - 1 = -0.08$). Since the influence of δ_m's is very small in all the following analysis set $\delta_m = 0$ throughout. To complete the above equation we must specify the action that must be taken in the cases that all the observed X_{mj} 's are zero; this specification is especially relevant when averaging a series of observations. One possible action is to just discard that particular observation, another possible action would be to assign q_N the value of 12 (which is the expected value for very rare events). The results of both courses of actions will be considered later on in this section, see Equation (4.34).

$q_N(j)$ can be considered to be a sample from a random process q_N which at each (yearly) time j is defined as:

$$q_N(j) = \frac{\mu(W; \ [j - N + 1 : j]\text{year})}{\mu(m_X; \ [j - N + 1 : j]\text{year})} \tag{4.31}$$

where $\mu(U;[t_1:t_2])$ symbolises the time-averaging performed on each member of the ensemble of the random-process U across a specified time interval $[t_1:t_2]$; m_X the random process obtained by time averaging of X across one year:

$$m_X(j) = \mu(X; \ [12j + 1 : 12j + 12]_{month}) \tag{4.32}$$

The assumption of annual cyclo-stationarity of the underlying propagation process implies that q_N is stationary (the statistical properties of q_N are independent of j).

q_N is an estimator for Q_1, it is therefore useful to obtain the statistical properties of q_N in order to be able to evaluate the accuracy of the measured Q_1 and also to establish the required observation time of measurements in order to meet the specified accuracy in measured Q_1. In general since q_N is a ratio of two correlated random processes it would be very difficult to obtain an analytical expression for its statistics. However in the case of the CEX model for P_X with the further assumption of the simplified dual-population season model [*Mawira*, 1985] has derived some interesting properties as well as an analytical expression for the moments of q_1 the 1st and 2nd moments of q_N for $N > 1$. Later [*Dellagiacoma & Targucci*, 1987] obtained the same results.

The exact derivations are somewhat involved. Here we give the main results, while the exact derivation is given in Appendix 4.2. The results are as follows:

(1)   If the process within the assumed M *Active* months in a year is purely exponential (i.e. $C_o = M/12$ so that $P_X(X|Active \text{ months}) = (12/M)C_0\exp(-X/C_1) = \exp(-X/C_1)$) then:

(1.1)   The estimator q_1 and the annual mean m_X are uncorrelated.

(1.2) q_1 is an unbiased estimator of Q_1, that is to say $<q_1> = Q_1$.

(1.3)   The k-th moment, $g_{k,M}$, of q_1 is given by:

$$g_{k,M} = 12^k \frac{M!}{(1+k)_{M-1}} \sum_{l=0}^{M-1} \binom{M-1}{l}(-1)^l \frac{1}{(1+l)^{k+1}} \tag{4.33}$$

here we have used the Pochammer notation: $(u)_i = u(u+1) \dots (u+i-1)$ and $(u)_0 = 1$; see Section 6.1.22 of [*Abramovitch & Stegun*, 1975]. Further, the above expression is for $M \geq 1$; also, we define $g_{k,0} = 12^k$.

(2) In other cases (i.e. $C_0 \neq M/12$):

(2.1) q_1 is in general a biased estimator of Q_1 with a non-negative bias: $\langle q_1 \rangle \geq Q_1$. The bias is largest ($\approx 1.0$) when $C_0 = 2/12$ and at the same time $M = 12$.

(2.2) The k-th moment, $\langle q_1^k \rangle$, of q_1 is given by:

$$\langle q_1^k \rangle = \frac{1}{r_x} \sum_{m=m_x}^{M} \binom{M}{m} \left(\frac{12}{M}C_0\right)^m \left(1 - \frac{12}{M}C_0\right)^{M-m} g_{k,m} \tag{4.34}$$

The value of m_x and r_x depends on whether we are looking at this equation from the point of view of the "naïve" experimentalist or the mathematician. Since the process here considered is conditional there will always be years where all the observed monthly t.f.e. will be equal to 0. The experimentalist will remove this year from the analysis since the worst month quotient cannot be determined (0/0). The mathematician, on the other hand, will consider this situation to be a limit case of a process occurring in one month only with $12C_0 \to 0$, with as limiting value for the moments 12^k. Intuitively we can say that in regions with very small C_0 the probability of having years with more than one month having $X > 0$ becomes remotely small in comparison of observing years with exactly one month having $X > 0$. Of course the total number of years with all 12 months having $X = 0$ will predominate. This however would not influence the estimation value of q_1 at all. We therefore have:

$$\begin{aligned} m_x &= 1, \quad r_x = 1 - (1 - \tfrac{12}{M}C_0)^M \quad \text{(the experimentalist)} \\ m_x &= 0, \quad r_x = 1 \quad \text{(the mathematician)} \end{aligned} \tag{4.35}$$

(3) For $N > 1$ we have, in all cases, the following expression for the 1st and 2nd moment of q_N:

$$\langle q_N^i \rangle = \sum_{m_1=0}^{M} \cdots \sum_{m_N=0}^{M} B_M(m_1, \tfrac{12}{M}C_0) x \ldots x B_M(m_N, \tfrac{12}{M}C_0) \langle q_N^i | m_1, \ldots, m_N \rangle \tag{4.36}$$

with $i = 1$ or 2; B_M represent the binomial probability function:

$$B_M(m, p) = \binom{M}{m} p^m (1 - p)^{M-m} \tag{4.37}$$

and:

$$\langle q_N | m_1, m_2, \ldots, m_N \rangle = \sum_{j=1}^{N} g_{1, m_j} \left(\frac{m_j}{\sum_{i=1}^{N} m_i} \right) \tag{4.38}$$

$$\langle q_N^2 | m_1, m_2, \ldots, m_N \rangle = \sum_{j=1}^{N} \sum_{k=1}^{N} [\delta_{j,k} g_{2, m_j} + (1 - \delta_{j,k}) g_{1, m_j} g_{1, m_k}] \frac{m_j (m_k + \delta_{j,k})}{\sum_{i=1}^{N} m_i (\sum_{i=1}^{N} m_i + 1)} \tag{4.39}$$

where the $g_{k,m}$'s are as given by Equation (4.33) and the δ's are the Kronecker symbol.

Using these equations the standard deviation σ_{q_N} of error in q_N can be determined. The resulting mathematical expression for the standard deviation is in general very cumbersome, except, somewhat surprisingly, for the cases when $C_0 = M/12$. In this case the expression for the standard deviation for multi-year observation depends on a simple way on the single-year standard deviation:

$$\sigma_{q_N} = \sigma_{q_1} \sqrt{\frac{1+M}{1+MN}} \tag{4.40}$$

(for the cases $C_0 = M/12$)

The standard deviation σ_{q_N} gives an indication in the uncertainty of estimating Q_1 through limited duration measurements. Figure 4.3 shows, for 1 year observation period, the dependence of this standard deviation on C_0. Two extreme cases of the seasonal population are shown, one with M = 12 the other with $C_0 = M/12$. In combination with the results found in Section 4.4 on

the dependence of Q_1 on C_0 we can see that the root-mean-square (RMS) uncertainty in predicting $P_{a|WM}$ from P_a (assuming perfect prediction) when using Q_1 determined from 1-year of measurements is throughout some 35%.

The dependence of σ_{q_v} on N has also been calculated, the results are shown in *Figure* 4.4 for the two extremes allowed in the 2-population model, the case $M = 12$ and the case $C_o = M/12$ ($M = 1, 2, ..., 12$). From these results it can be seen that to measure Q_1 accurately a long observation period is needed, e.g. 10 years of measurements are needed to obtain some 10% accuracy in Q_1. Previously in Section 4.2 we have found that in the theoretical relationship between Q_1 and C_o there is a variation of possible Q_1 values, for fixed C_0, which range some ±5% due to the possible spread of value of M which in the 2-population model is the 'unknown' parameter. We see now that this uncertainty is negligible compared to the empirical uncertainty (most propagation measurements up to date produce less than 4 years of measurements per site).
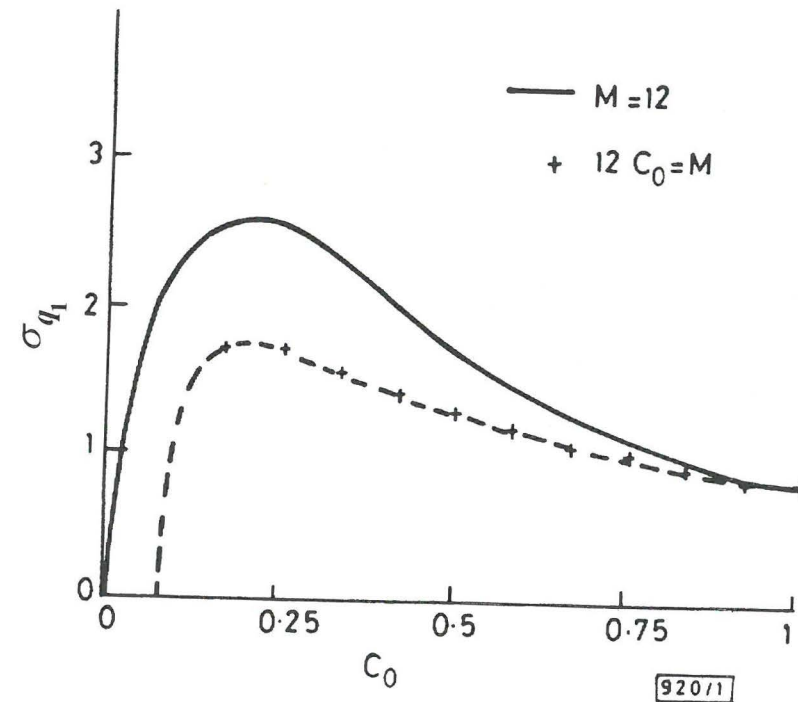
Figure 4.3 Dependence of σ_{q_v} on C_0, for $N = 1$ year. Two cases are shown: solid line = a uniform population ($M=12$), markers(+) for the special cases that $C_0 = M/12$.
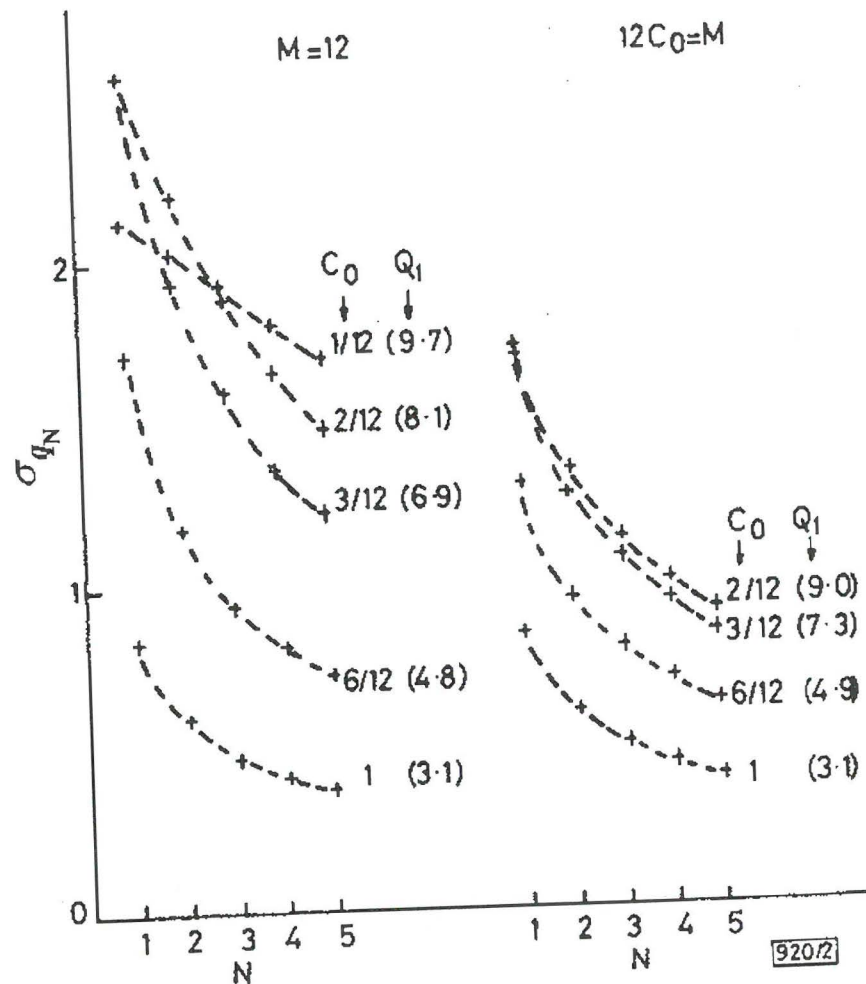
Figure 4.4 Dependence of σ_{q_N} on the number N years of observations. The left part gives the dependence for a uniform population ($M = 12$). The right gives the dependence for cases when $C_0 = M/12$.

4.7 Conclusions

A theoretical analysis shows that the simplified model for the monthly TFE X (using only the aggregate distribution and dual population assumptions), can be used to predict the statistics of annual worst-month TFE W within known tight bounds. A theoretical analysis with regard to another related parameter, the ITU worst month quotient Q_1, has been performed. The relationship between Q_1 and the location parameter C_0 as predicted by various models has been determined. The theoretical analysis has also produced formulas and graphs to determine the accuracy of worst-month statistics from time-limited measurement data. The results show that for all models the influence of the seasonality index M on Q_1 is, in general, not so large. On the whole the theoretical uncertainty in Q_1 due to M is within the range of $\pm 5\%$. Also, from the results it can be seen that to measure Q_1 accurately a long observation period is needed, e.g. 10 years of measurements are needed to obtain some 10% accuracy in Q_1.

5. Application to Reliability & Risk analysis

5.1 Introduction

Traditionally most radio networks have been designed so as to meet the grade of service objectives formulated by the ITU. These internationally agreed set of standards give the telecommunication operators a clear set of objectives to pursue. Furthermore these standards can be seen as representing a synthesis of the range of requirements of the various telecommunication operators as well as having incorporated their experiences regarding feasibility and particularities of various systems. Such international standards are still very important in telecommunication practices especially in ensuring system compatibility and inter-operability in international telecommunications.

However, such common standards take a very long time to develop within a multiple-party negotiation process where compromises have to be made. Moreover, the global deregulation of the telecommunication industry has now resulted in a much higher differentiation of the services which have to be delivered in a competitive free market environment. In this much more complex environment each telecommunication operator is individually responsible for making design choices and price-quality trade-offs with regard to novel value-added services/products he wants to deliver. This also means that techno-economical scenarios must be analysed in much more depth and sophistication than previously. In particular, the analyst must know how the short-term deviations of the performance of a system around the expected long-term performance could impact on the customer satisfaction. The significance of this information can be appreciated when we note that the way the clients perceive the quality of the system will be based on their direct and individual experience of the system and not on its ensemble & long-

term averaged quality. Clients who think that they are being provided with less than the expected quality may decide to give their business to the competitor.

It can be therefore stated that nowadays it is not sufficient anymore to characterise a system by just a single all encompassing figure of merit. The need for more extended analysis of the impact of propagation on the performance of the system has been pointed out already by several authors [*Karasawa & Matsudo*, 1988], [*Fukuchi & Watson*, 1989], [*Watson*, 1990], [*Brussaard & Mawira*, 1992], [*Crane*, 1993]. In these papers the concept of 'risk' is used to denote the magnitude of deviation of deviations of performance parameters around their ensemble means.

Obviously for such extended analysis more detailed propagation information is required than, say, the long-term statistical distributions of e.g. the attenuation. Ideally we would like to have for such analysis a model which can give realistic and detailed time-series information. What we have at the moment on time-series modelling is very restricted and based mostly on data from earth-satellite paths. Therefore, as an alternative approach, we shall consider in this section the possibility of using the monthly time fraction of excess (TFE) for such analysis. The monthly TFE is of course a much cruder type of information than the detailed time-series information; on the other hand the behaviour of the monthly TFE has successfully been modelled as discussed in the previous sections.

We note here that in a more recent development better modelling of time series behaviour of propagation parameters is being attempted with good progress. In particular analysis of fade duration on satellite links [*Brussaard*, 1995] has resulted in a (seemingly) generally applicable model for the fade duration referred to as the Simple Exponential Model (SEM). Another study proposes a bi-exponential model for the fade duration statistics with regard to slant path attenuation [*Lekla, et al.*, 1998].

5.2 Basic concepts

The continuous need to develop better and more reliable products, especially in the field of manufacturing, has led to the development of special engineering disciplines such as Quality Engineering and Reliability Engineering. Within these disciplines, tools and methodologies have been developed in an attempt to ensure the end-product quality at various phases of product development (e.g. identifying customer requirements, translating these requirements into engineering characteristics and assigning limits based on customer tolerances etc., see [*Taguchi*, 1986], [*Boza et al.*, 1994]). Reliability engineering is involved in various aspects concerning "reliability" of a product and includes reliability estimation, malfunction and failure analysis, cost aspects, risk analysis etc. [*Ireson*, 1966], [*Barlow & Proschan*, 1975], [*Suhir*, 1997].

Curiously in radio network design the formulations and practices of reliability engineering do not seem to be used extensively in the evaluation of performance aspects related to propagation conditions. Consider the three important classes or types of concepts/parameters from reliability engineering namely *Life-length*, *Availability* and *Risk*, defined as:

(1) the *Life-length* class of concepts/parameters quantifies the expected duration of the time that a system shall function until failure. A well known example is the reliability function $R(t)$ which gives the probability that a failure-free operation will be obtained of a duration at least exceeding a value of t; here the reliability is indicated in terms of a cumulative distribution function. Another parameter, the *MTBF* (Mean Time Between Failure), is a popular single figure of merit for indicating the reliability of an equipment.

(2) the *Availability* type of parameters measures the fraction of time that the system is operating (in a satisfactory way). There are various definitions (e.g. Intrinsic Availability, Operational Readiness) which depend on how the reference time is defined (e.g. whether or not the period for preventive maintenance is to be included in the reference time). In this section we restrict ourselves to systems which have only two possible states (operating satisfactorily or in failure). The fraction of time the system is functioning satisfactorily is indicated as $P_{Available}$. The complement of *Availibility* is the *Outage*, $P_{Outage} = 1 - P_{Available}$.

(3) *Risk* is used to characterise/quantify the uncertainties or the chance of incurring unwanted effects. There is a large range of associated parameters and concepts, e.g. risk functions, risk curves, consumer's risk, producer's risk.

Of these three classes of concepts only the *Availability* concept is readily recognised as being used extensively for the characterising propagation effects on radio systems and in the definitions of Grade of Services requirements of the ITU. In the ITU definition the term Availability is used to denote a special case of *Availability* as defined above; see discussion of ITU Grade of services in Section 1.2.

One of the advantage of using a *Life-length* type of parameters is that it provides the user with a measure on the temporal scale with which a system will perform satisfactorily. The basis of such time analysis is quite simple. The system always resides in one of two conditions, the systems is either operational or in failure. Two time durations are defined:

$$TBF = \text{time-between-failure}$$
$$TTR = \text{time-to-repair}$$

see Figure 5.1. Note we are here only interested in repairable systems, for a non-repairable system the relevant time parameter is the time to failure (*TTF*). We can in principle easily find the equivalent parameters to *TBF* and *TTR* for propagation effects by noting that the duration that a propagation parameter exceeds a critical limit (leading to outage of the system) is the propagation equivalent to *TTR*, while the period between the exceedances (outages) corresponds to *TBF*; see Figure 5.1.
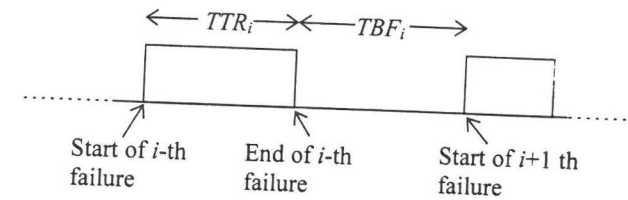
Figure 5.1 Definitions of TTR and TBF

In reliability engineering the *MTBF* and *MTTR* parameters defined as:

$$MTBF = \text{the mean time between failure} = \text{mean of } TBF$$
$$MTTR = \text{the mean time to repair} = \text{mean of } TTR$$

are very popular for characterising a system. These two parameters are related to *Availability*:

$$P_{Available} = \frac{MTBF}{MTBF + MTTR} \tag{5.1}$$

and to *Outage*:

$$P_{Outage} = 1 - P_{Available} = \frac{MTTR}{MTBF + MTTR} \tag{5.2}$$

These equations can be easily derived by considering the definition of *Availability*, or $P_{Available}$, as a limit:

$$P_{Available} = \lim_{N \to \infty} \frac{\sum_{n=-N}^{N} TBF_n}{\sum_{n=-N}^{N} TTR_n + TBF_n} \tag{5.3}$$

see Figure 5.1, which can further be expressed as:

$$P_{Available} = \lim_{N \to \infty} \frac{\frac{1}{N}\sum_{n=-N}^{N}TBF_n}{\frac{1}{N}\sum_{n=-N}^{N}TTR_n + \frac{1}{N}\sum_{n=-N}^{N}TBF_n} = \frac{MTBF}{MTTR + MTBF} \qquad (5.4)$$

In a similar way Equation (5.2) can be proved as well.

The above equations show that from the *Life-length* parameters $MTTR$ and $MTBF$ we can determine the *Availability* parameters $P_{Available}$ and P_{Outage}, but the converse is not true. This is due the fact that while the two Life-length parameters $MTTR$ and $MTBF$ can be independently specified the two Availability parameters $P_{Available}$ and P_{Outage} are complements of each other.

In meteorological and hydrological sciences a closely related parameter, the Return Period, T_{return}, is often used. This Return Period is intuitively defined as the expected duration between two successive occurence of a class of events [*Gumbel*, 1956], [*Barry & Chorley*, 1974], [*Buishand & Velds*, 1980], or more precisely, the Return period is the average of the durations between two subsequent events:

$$(5.5)$$

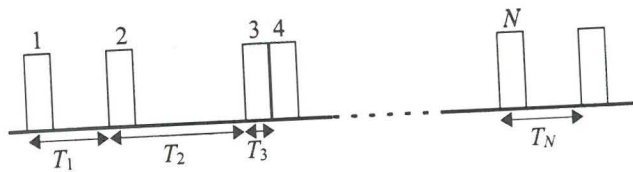$$T_{Return} = \lim_{N \to \infty} \frac{1}{N}\sum_{i=1}^{N}T_i$$

see Figure 5.2.

Figure 5.2 Definitions of durations between events

Note that in this scheme the definition of the event is based on discrete time periods (of equal length e.g. an hour); when, within a given time period certain physical conditions are met an event is considered to have occurred. Each event, in this scheme, has a duration equal to the chosen discrete time period.

The Return Period, T_{Return}, can be related to the probability by first noting that the probability of occurrence of an event can be described as:

$$P_{Event} = \lim_{N \to \infty} N\Delta T / \sum_{i=1}^{N}T_i \qquad (5.6)$$

where ΔT is the sample duration. By comparing this equation to Equation (5.5) we obtain:

$$T_{Return} = \frac{\Delta T}{P_{Event}} \qquad (5.7)$$

where P_{Event} is the probability of occurrence of the event of interest (e.g. in radio communications studies we often take outage as the event of interest, *Event = Outage*).

Note that in this definition the mathematical definition of the Return Period requires some care in its interpretation, in particular with regard to the definition of an "Event". Here it is assumed that "Events" are singular (defined at discrete time points). As an example of possible ambiguity in this definition, consider the case of a new time series constructed through oversampling a given time series by a factor of two (i.e. $(\Delta T)_{Oversample} = (\Delta T)_{Original}/2$). Application of Equation (5.77) results in the Return period value for the new time-series which is half the value of the Return period belonging to original time series (since P_{Event} stays the same). What has actually happened is that by the oversampling we have actually doubled the number of events while at the same time the sum of the time intervals, as in Equation (5.5), (within a fixed long period) does not change. This ambiguity in the interpretation of what consists of an 'event' is the cause that although T_{Return} on the one hand and $MTBF$ on the other hand each give a measure regarding the average duration between 'events' there is a subtle difference between the two which makes their inter-relationship less obvious.
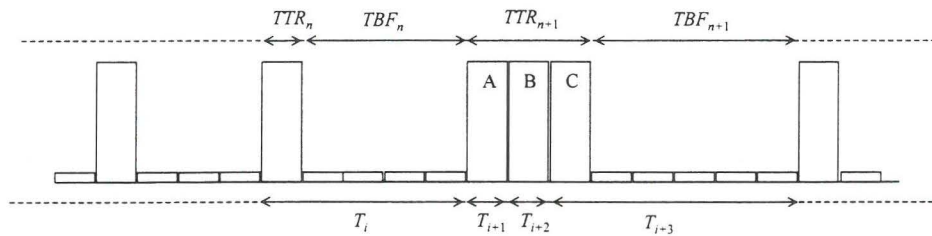
Figure 5.3 Interpretation of time-intervals

Figure 5.3 illustrates this difference. Here we can see that the most important difference between the two definitions regards the interpretation of time intervals with multiple successive failures. In the reliability method this period (e.g. A-B-C) is considered to be one long period failure, in the return period calculation each (failure) sample (A, B, C) is considered a separate event with each defining a return period with length of one sample distance ΔT. Actually, the problem at hand is not caused by the concept of return periods but is caused by an indiscriminate use of Equation (5.7). By comparing Equation (5.2) to Equation (5.3) (taking *Event = Outage*) we obtain the following relationship between T_{Return} and MTBF:

$$T_{Return} = (MTBF + MTTR)\frac{\Delta T}{MTTR} = MTBF\frac{\Delta T}{MTTR} + \Delta T \qquad (5.8)$$

Since $MTTR$ is always larger than ΔT, T_{Return} will in general be smaller than $MTBF$ (for the important cases that $MTBF$ is much larger than ΔT). However if P_{Event} is much smaller and the samples independent then it can be expected that the probability of obtaining two failures in succession becomes negligible. In such cases the value of $MTTR$ comes close to ΔT and $MTBF$ is much larger than ΔT so that we have:

$$T_{Return} \approx MTBF \qquad (5.9)$$

The reason why the *Life-length* concept is in general not used in characterising propagation effects on radiocommunication systems is not well documented. Probably one of the detriments in using it is the difficulty of obtaining the relevant duration statistics. Even nowadays we have only limited information on fade duration statistics (which is the equivalence in propagation to the *TTR* statistics). Another problem is related to the presence of very many but very short duration events in propagation which makes the resulting values of e.g. *MTBF* strongly dependent on the integration/sample time of the data (a highly undesirable feature). This sensitivity of 'duration' of propagation events on e.g. integration time is a known problem which occurs in studies on duration statistics [*Brussaard et al..*, 1985], [*McCormick*, 1998].

In order to make the *Life-length* concepts workable for application to propagation effects we extend the concept of failure from describing the instantaneous (malfunction) state of the radio link to a more global definition using more accessible time frame, e.g. the month. In fact this approach is akin to the method of ITU-R is separating the various propagation effects from the microscopic effects (e.g. Severely Errored Seconds criteria) to the more macroscopic effect (the availability criteria), see Section 1.2. In the following section we shall take the month as a unit reference time and the TFE as the quantity defining the failure (non-failure) mode of the system.

5.3 Return periods of Yearly and Worst-Month TFE

Most ITU Grade of Service requirements are defined with respect to the long-term average of worst-month and/or the yearly time fraction of excess (TFE). As a simple extension to the characterisation of the impact of propagation on radio systems we can consider the analysis of the variability of the annual worst-month TFE resp. the yearly TFE's and their Return periods (as their *Life-length* parameter). As noted in previous sections the annual worst-month TFE, *W*, and the yearly TFE, *Y*, can each be considered to be a random process. Both processes are discrete in time with sampling time $\Delta T = 1$ year. Furthermore the assumption, introduced in Section 2, of cyclo-stationarity and independence of *X* leads to a stochastic model of *Y* being stationary and independent.

As already discussed in Section 1 the ITU-R grade of services are mostly expressed in terms of the long-term probability, P_{LT}:

$$P_{LT} = <Y> \qquad\qquad\qquad (5.10)$$

and/or average-annual-worst-month probability, P_{WM}:

$$P_{WM} = <W> \qquad\qquad\qquad (5.11)$$

A properly designed and implemented system will produce average values of *Y* and *W* when observed across a very long period very close to the target probabilities P_{LT} and/or P_{WM}. On a yearly basis however deviations of the observed value of *Y* and *W* with respect to P_{LT} and P_{WM} respectively, must be expected. These deviations are expressed by the probability distributions P_W and P_Y discussed in the previous sections. The associated Return periods are given by:

$$T_{Return,WM} = \frac{1}{P_{WM}} \qquad \text{(months)}$$

$$T_{Return,Y} = \frac{1}{P_{LT}} \qquad \text{(years)} \qquad\qquad (5.12)$$

Using the CEX model we can easily calculate the dependence between the Return period, the TFE value and the model parameters. The results are given in Figure 5.4. This figure gives the dependence of the normalised TFE's (W/P_{WM} respectively Y/P_{LT}) on C_0 for several values of the Return periods. One of the interesting properties that can be seen from this figure is that there is very little difference in the worst-month curves as compared to the yearly curves.
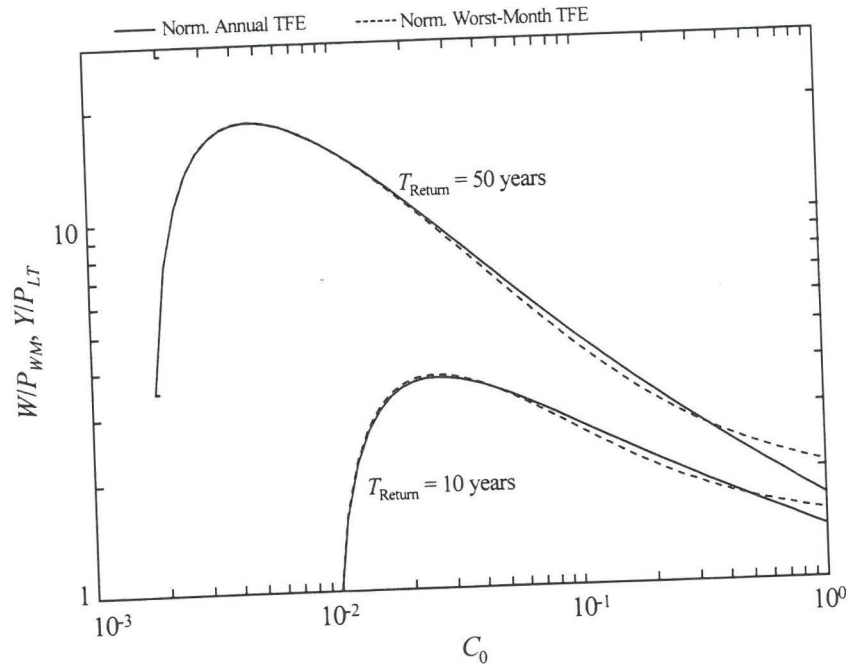
Figure 5.4 Normalised Worst-month and normalised Yearly TFE values.

This striking similarity between the two normalised TFE's form the basis for the curves developed by the author, in 1992, for Recommendation 620 of ITU-R. In this Recommendation the parameter C_0 is replaced by the parameter Q_1 using theoretical expression for CEX model given in Appendix 4.1 of Section 4. This replacement was performed since most users are more familiar with the parameter Q_1.

5.4 Example of risk assessment

There is currently a growing interest in the radiocommunication community for usage of risk analysis (in addition to the classical design methods with availability/unavailability criteria etc.), see for example the development within ITU SG3 [reference to WP 3J Chairman's Report DOC. 3J/32-E, 2/6/98]. While there is at the moment no single definition of what consists risk analysis we shall in this section discuss a simple example that we think is a good illustration of risk analysis. We consider the situation of a provider who wants to set up a (radio) communication system with many users and with each user having a separate link. The links have similar propagation properties but are statistically independent. The special service being provided is that for any monthly period in which the monthly TFE X exceeds the promised long-term worst-month $<W>$ value the victim is given a financial compensation:

$$payment = \begin{cases} z \bullet (X - <W>)/<W> & \text{(when } X > <W>) \\ 0 & \text{(when } X \leq <W>) \end{cases} \qquad (5.13)$$

Of course many alternative compensation schemes may be devised. However, the above scheme seems to be a reasonable proposal for a compensation scheme: it produces a compensation in proportion with the excess outage deviation from the promised value ($<W>$); obviously a necessary requirement for an honest compensation scheme. The normalisation to $<W>$ seems also to be reasonable, it furthermore simplifies our analysis (by eliminating the scale parameter C_1). Of course in the end the service provider might want to make the choice of the value of z dependent on many other factors (e.g. so as to avoid excessive loss in say unfavourable propagation regions).

The expected payment per user per month is:

$$cost = <payment> = z \int_{<W>}^{\infty} dX \, f_X(X)(X - <W>)/<W> \qquad (5.14)$$

where f_X is the probability density function of X. Equation (5.12) can also be expressed in terms of P_X:

$$cost = \frac{z}{<W>} \int_{<W>}^{\infty} dX \, P_X(X)$$

(5.15)

this is obtained by applying partial integration to Equation (5.12).

Using the extended CEX model we can readily obtain an analytical expression for the expected cost per user per month:

$$cost = \begin{cases} \dfrac{z}{Q_1(C_0)} \exp(-C_0 Q_1(C_0)) & \text{(for } C_0 \leq 1) \\ \dfrac{z}{Q_1(1) + \ln(C_0)} \exp(-Q_1(1)) & \text{(for } C_0 > 1) \end{cases}$$

(5.16)

where the dependence of Q_1 on C_0 is as given in Appendix 4.1 of Section 4.

Figure 5.5 shows the dependence of *cost* on C_0 (for several values of the seasonality index M). From this figure two interesting features can be seen. In the first place we can see that the largest cost value is obtained for the smallest value of C_0, here we have a cost of $z/12$ per user per month (or simply $z(\$)$ per user per year). In the second place we have a large range of C_0 (0.5 - 2) where the *cost* is practically constant = $0.0167\ z$ per user per month (or equivalently $0.2\ z(\$)$ per user per year). For C_0 below 0.5 the cost rapidly increases with decreasing value of C_0. This rapid increase of cost is not very obvious at the onset of this exercise This demonstrates the necessity of properly analysing the problem at hand with the aid of good models.
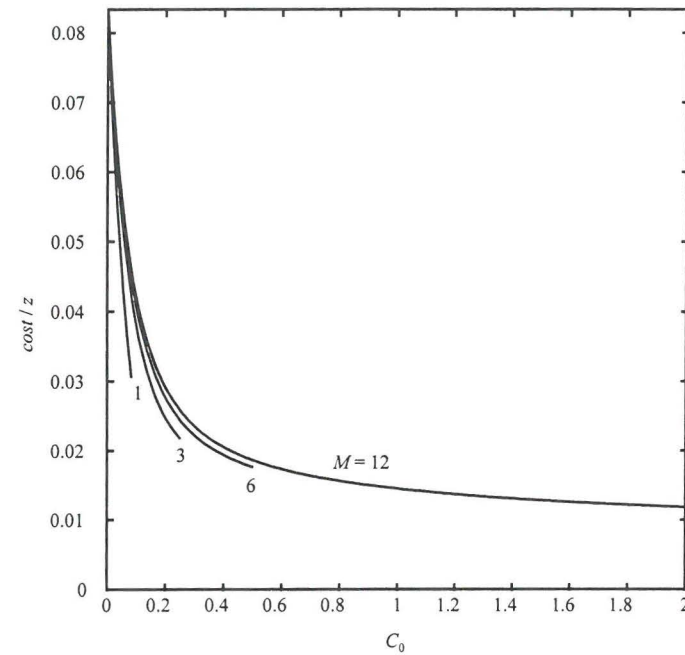
Figure 5.5 Expected cost per user versus C_0.

5.5 Conclusions

In this section the basic-concepts from reliability engineering (e.g. MTBF) , meteorology (e.g. Return Periods) and from radio engineering were looked into, in particular their inter-relationships. It is concluded that the return-period concept would be a useful addition for the analysis for radio link performances. Using the CEX model for X (and the consequent distributions for Y and W) the relationship of return periods with the conditionality parameter C_0 were determined for Y and W. An interesting fact that emerged from comparison of the curves for Y to that for W is their very close similarity when the values are normalised by their means ($Y/<Y>$ respectively $W/<W>$). Finally, an example is given where the model is used to estimate the financial risk of an insurance scheme on a satellite communication system.

6. Conclusions and recommendations

The primary purpose of the research of this thesis is to produce proper models for the statistics of the time fraction of excess (TFE) of parameters related to the propagation of microwave radiowaves through the troposphere. The TFE considered is that which is based on the monthly period although applications to TFE based on longer periods have also been considered. The work in this thesis is based on theoretical modelling as well as analysis on a very large amount of data on the monthly TFE with respect to diverse propagation processes (slant-path rain attenuation effects and ducting effects on transhorizon paths). The data used are those collected by the European COST205 and COST210 projects and compose in total of 92 site years of observations obtained in some 9 different climatic zones of Europe.

The main results obtained by our studies are the following:

(1) A proper formal conceptual framework has been obtained:
Herein the TFE, associated with a particular given link, is modelled to be a realisation from a random process X. This random process represents the ensemble of TFEs of all (hypothetical) links with equal link parameters located in an area with homogeneous climatic properties. Analysis using measured data leads to the conclusion of a random process that is cyclo-stationary and exhibits independence. The proof of these two aspects is, admittedly, more circumstantial than rigorous due to the limitations of the data. The longest observation time of the available data was 9 years so those cycles longer than 9 years cannot be detected in this study. The independence has been conjectured from the, in general, observed lack of correlation between the samples; for TFEs associated with lower thresholds values some correlation between subsequent months have been observed. Nevertheless these two

properties can be taken as simplifying approximations. The (seasonal) complexity of having to use the 12 statistical distributions (one for each month of the year) is reduced to a simplified model based on the aggregate distribution P_X and a dual-population model has been proposed.

(2) A best model for the aggregate distribution P_X has been given:

For the modelling of the aggregate distribution P_X four candidate models have been tested using some 92 site years of data from 12 locations in Europe (slant-path attenuation and transhorizon signal levels). The models are the conditional exponential (CEX), the conditional lognormal (CLN), the shifted gamma (SGM) models and the Gamma (GM) model. The CEX model has been extended to cases with $C_0 > 1$. The SGM is a new model introduced by the author. The analysis shows that classical test methods, for example the χ^2 minimum method and the regression method, do not allow definite statements regarding superiority/inferiority of these models. This is due to the limited number of sample points available for each sample distribution, and the similarity of the four distributions in the midrange. To circumvent these problems a novel normalisation method (to create a "grand pool" of test data) has been devised. The analysis using this method has indicated the superiority of the SGM model, with the CEX model a close second. The GM and the CLN model do not seem to offer any advantage over the other two models. The table below giving the RMS errors illustrates these results:

SGM	CEX	CLN	GM
6%	8%	27%	44%

We note here further that the CLN model has one more parameter than the other three models, which makes it even less attractive in practical usage.

Finally, since the difference in accuracy between the SGM and CEX model is not so large we may feel justified to use the CEX model whenever it is more convenient to do so. This may be the case in mathematical analysis where often using the CEX model is simpler than using the SGM model.

(3) Models for variability yearly TFE have been obtained:

In the next step we have considered the feasibility of using the model(s) for the statistics of X for the prediction of the variability of the annual time fraction of excess Y. It has been shown that the CEX, GM and SGM models allow good predictions of the magnitude of the variation of Y around its expected mean (as expressed by the coefficient of variation Ω_Y). The CLN performs here rather badly, resulting often in gross overestimation. This last result seems to be somewhat surprising since the CLN model was found to perform not too badly in the statistical modelling of X done in Section 2. It may be that the parameter estimation method used overestimates the values of the shape parameter C_2, which in turn may lead to excessive predictions of the values of Ω_Y. In any way this illustrates the problem associated with a model using many parameters. For the CEX, GM and SGM models the theoretical expressions for the probability distributions P_Y of Y have been derived.

(4) Models for worst-month predictions have been obtained:

An important part of the thesis is dedicated to the analysis of the "worst-month" time fraction excess W. A theoretical analysis shows that the simplified model (using only the aggregate distribution) presented before on the statistics of X, can be used to predict the statistics of W within known tight bounds. A theoretical analysis using another related parameter, the ITU worst month quotient Q_1, has produced formulas and graphs to determine the stability estimation of worst-month from time-limited data. Finally, from the measured data, a classification of Europe in several zones with distinct worst-month properties has been produced. The zones found are (1) North-West Europe, (2) Alpine, (3) Mediterranean, (4) Scandinavia.

(5) Extension of grade of service criteria to risk and reliability alike methods:

In the first place the basic-concepts from reliability engineering (e.g. MTBF) , meteorology (e.g. Return Periods) and from radio engineering were looked into, in particular their inter-relationships. It is concluded that the return-period concept would be a useful addition for the

analysis for radio link performances. Using the CEX model for X (and the consequent distributions for Y and W) the relationship of return periods with the conditionality parameter C_0 were determined for Y and W. An interesting fact that emerged from comparison of the curves for Y to that for W is their very close similarity when the values are normalised by their means ($Y/<Y>$ respectively $W/<W>$). Finally, an example is given where the model is used to estimate the financial risk of an insurance scheme on a satellite communication system.

To sum up we can state that in this thesis we have obtained a comprehensive model for the random process of the monthly TFE. This monthly TFE exhibits independence and annual cyclo-stationarity and has an aggregate distribution P_X which has a SGM distribution (the CEX model is slightly less accurate but is also good model for P_X). Seasonality can for many applications be simplified to a dual-population model. This model can be used to for predictions of properties of TFEs that can be constructed from the monthly TFE, e.g. worst-month TFE, annual TFE. The model is therefore very suitable for producing more extended analysis of the impact of the propagation on telecommunication systems, e.g. the risk analysis of the 'insurance' case given in this thesis.

Recommendation for further research

A fact that is observed from the analysis of the data is that there does not seem to be any fundamental difference between the X associated with slant-path attenuation and the X associated with ducting. This is somewhat surprising since these two phenomena are based on very different mechanisms: attenuation is caused by rain while ducting is caused by anomalous refractive index stratification/layering of the atmosphere (during non-rainy conditions). This is probably a matter worth investigating further. Another fact that should be investigated is the possible unification with duration statistics (e.g. how the models now developed in conjunction with duration statistics can explain the statistics of the monthly TFEs).

References

Abramovitz, M., Stegun, I.A., (1975), *Handbook of mathematical function*, Dover Publication.

Balakrishnan, N., Rao, C R, (1997), A note on the Best Linear Unbiased Estimation Based on Order Statistics, *American Statistical Association*, Vol. 51, No. 2, pp 181 – 185.

Barlow, R.E., Proschan, F., (1975), *Statistical Theory of Reliability and Life Testing: Probability Models*, Holt, Rinehart and Winston, Inc., New York, ISBN 0-03-085853-4.

Barry, R.G., Chorley, R.J., (1974), *Atmosphere, Weather and Climate*, Methuen & Co, London, pp 108-109.

Boithias, L, (1987) *Radiowave propagation*, North Oxford Academic, ISBN 0-946536-06-6.

Box, G.E., Jenkins, G.M., (1970), *Time Series Analysis: forecasting and control*, Holden-Day, Library of Congress Cat. 77-79534.

Boza, L.B., Ciaccia, T.J., Gatenby D.A., Muise R.W., Ng, K.K., Yanizeski, G.M., (1994), "Achieving Customer Satisfaction Through Robust Design", *AT&T Technical Journal*, Jan./Feb., 1994, pp 48 - 57.

Bremmer, H., (1949), *Terrestrial radio waves*, Elsevier.

Brussaard, G., Watson, P.A., (1979), Annual and annual-worst-month statistics of fading on earth satellite paths at 11.5 Ghz, *Electronic Letters*, Vol.14, No.9, pp278-280.

Brussaard, G., Mawira, A., (1993), Propagation statistical factors affecting the planning of radio networks, *presentation at session FC2-1 of XXIVth General Assembly of the International Union of Radio Science, Kyoto-Japan.*

Brussaard, G., Dijk, J., Wijdemans, L., (1993), 11 GHz satellite beacon data in the western pacific, *Final Report INTELSAT* 770-B, December.

Brussaard, G., (1995), Extreme-value analysis of outage durations due to rain in satellite communication systems, *URSI Comm. F Open Symposium, Ahmebabad, India.*

Buishand, T.A., Velds, C.A., 1980, *Neerslag en verdamping*, Royal Netherlands Meteorological Office, pp 104-105.

Bury, K.V., (1986), *Statistical Models in Applied Science*, Krieger, Melbourne, Fla.

Collin, R. E., An introduction to the classic paper by M.I. Pupin, *Proc. of the IEEE*, Vol. 85, NO. 2, 1997, pp 301 - 305.

COST205, (1985), Final Report of EURO COST Project Nr 205, Publication of the Commission of the European Communities, Luxembourg, 1985, ISBN 92-825-5412-0.

COST210, (1991), Final Report of EURO COST Project NR 210, Publication of the Commission of the European Communities, Luxembourg, 1991, ISBN 92-826-2400-5.

Cramér, H., (1946), *Mathematical methods of statistics*, Hugo Gebers Forlag, edition printed in the U.S.A..

Crane, R.K, DeBrunner, W.E., (1978), Worst-month statistic, *Electronic Letters*, Vol.14, No.2, pp38-40.

Crane, R.K., (1991), Worst month: a new approach, *Radio Science*, Vol.26, No.4, pp 801-820.

Dellagiacoma, G., D Tarducci, D., (1987), A different approach to worst month rain statistics: theory and experimental results, *Radio Science*, Vol.22, No.2, pp 266-274.

Delogne, P., (1997), Radio in 1913: An introduction to the classical paper by L. de Forest, *Proc. of the IEEE*, Vol. 85, No. 2, pp 323 -326.

Dintelmann, F., (1984), Worst-month statistics, *Electronic Letters*, Vol.20, pp. 890-892.

Doob, J.L., (1953), *Stochastic processes*, John Wiley & Sons Inc, New York.

Fedi, F. (ed.), (1985), Special issue on the COST 205 project on earth-satellite radio propagation above 10 GHz, *Alta Frequenca* 1980, Vol. LIV, No. 3.

Fukuchi, H., Kozu, T., Tsuchiya, S., (1985), Worst-month statistics of attenuation and XPD on earth-space path, *IEEE Trans. on Antennas and Propagation*, AP-33, No. 4, pp. 390-396.

Fukuchi, H., Watson, P.A., (1989), Statistical stability of cumulative distributions of rainfall rate in the UK, *Proceedings IEE*, Vol.136, Pt. H, No.2, pp 105-109.

Gumbel, E.J., (1967), *Statistics of extremes*, Columbia University Press.

Hall, M.P.M., (1979), *Effects of the troposphere on radiocommunication*, IEE Electromagnetic Wave Series, Peter Peregrinus Ltd, 1979.

Hewitt, M.T., Mehler, M.J., Bye, G.D., Larsen, U.O., Mawira, A., Dijk, J., (1989), Worst-month prediction model for transhorizon microwave interefence, *Electronic Letter*, Vol. 25, No. 24, pp 1671-1672.

ITU, (1955), Atlas of ground wave propagation curves for frequencies between 30Mc/s and 300 Mc/s, Geneva.

ITU, (1997), *P-series Recommendations of ITU Radio communication sector as approved by the plenary Assembly 1997,* Volume 1997 – P Series, Part 1&2. Order information can be found at World Wide Web: www.itu.int/publications/. See further the list of some ITU-R recommendations given at pp146 - 147.

Ireson, W.G. (ed.), (1966), *Reliability Handbook,* McGraw-Hill, New York.

Ishimaru, A., Lin, J.C., (1972), Multiple scattering effects on wave propagation through rain, *AGARD Conference Proceeedings No. 107, Kansas, U.S.A.,* pp 1.1 - 1.13.

Karasawa, Y., Matsudo, T., (1988), A concept of the factor of safety for prediction of rain attenuation, Trans. Inst. Electron. Inform. Comm. Engrs. (Japan), Vol. J71-B, No. 6, pp. 772 – 778.

Kerr, D.E. (ed.), (1951), *Propagation of short radio waves,* McGraw Hill.

Kreysig, E., (1970), *Introductory mathematical statistics: principles and methods,* John Wiley & Sons, ISBN 471 50730x.

Lefèbvre, J., Roussel, H., Walter, E., Locointe, D., Tabbara, W., (1996), Prediction from Wrong Models: The Kriging Approach, IEEE Antennas and Propagation Magazine, Vol.38, no.4, pp 35 - 45.

Lekka, R., McCormick, K.S., Rogers, D.V., (1998), 12-Ghz fade duration statistics on earth-space paths in South-East Asia, *Proceedings of the Climpara'98 conference, Ottawa, Canada,* 27 - 29 *April,* pp 167 - 170.

Massey, F.J., (1951), The Kolmogorov-Smirnov test for goodness of fit, *American Statistical Association Journal,* March, pp. 68-78.

Mawira, A., (1980), Statistics on rain-rates, some worst-month considerations, *Proceedings of URSI-F Symposium, Lennoxville-Canada,* pp 2.3.1-2.3.8, 1980. Also in *Ann. des Telecomm.,* Vol.35, No.11-12, pp423-428.

Mawira, A., Neessen, J.T.A., (1984), Propagation data for the dsign of 11/14 GHz satellite communication systems, *Proc. of the 10th AIAA Communication Sytems Conference,* pp 629-639, Orlando, Florida, USA, March 19-22.

Mawira, A., (1985), Variability of worst-month quotient Q, *Electronic Letters,* Vol.21, No.23, pp1073-1074.

Mawira, A., (1989), Prediction of the yearly and worst-month time fraction of excess using the conditional exponential model, *KPN Research Report* 789 RNL/89, November.

Mawira, A., (1990), Slant path 30 to 12.5 GHz copolar phase difference, first results from measurements using the OLYMPUS satellite, *Electronic Letters,* Vol. 26, No. 15, pp 1138-1139.

Mawira, A., (1997), Variability of monthly time fraction of excess of atmospheric propagation parameters, *Radio Science,* Vol. 32, Nr. 5, pp 1797-1811.

Miller, A.,Thompson, J.C., (1970), *Elements of meteorology,* pp 216-225, Charles E. Merril Publishing company, Ohio, Library of Congress C.C. Nr. 78-98476,.

Morita, K., (1970), Prediction of Rayleigh fading occurence probability of line-of-sight microwave links, *Rev. Elec. Comm. Labs.,* Vol. 18, 11-12, November-December.

Pearson, E.S., Hartley H.O., (1976), *Biometrica Tables for statisticians,* Biometrica Trust, ISBN 0 904653 10 2.

Papoulis, A., (1984), *Probability, random variables and stochastic processes*, McGraw-Hill, ISBN 0-07-048468-6.

Picquenard, A., (1974), *Radio wave propagation*, The MacMillan Press Ltd, SBN 333 13312 9, Library of Congress Catalog Nr: 72-2046.

Poires Baptista, J.P.V., Zhang, Z.W., McEwan, N.J., (1986), Stability of rain-rate cumulative distributions, *Electronic Letters*, Vol. 22, No. 7, pp 350-352.

Poires Baptista, J.P.V., Kubista, A., Witternigg, N., Randeu, W.L., (1989), Worst month statistics for high outage probabilities, *IEE Conf. Publ. No. 301*, part 2, 10-13, ICAP89, Univ. of Warwick, Coventry, UK.

van der Pol, B.,Bremmer, H., (1937), The diffraction of electromagnetic waves from an electrical point source round a finitely conducting sphere, with application to radio telegraphy and the theory of rainbow, *Philips Magazine*, Vol. 24, Suppl, November.

Sheffé, H., (1959), *The analysis of variance*, John Wiley & Sons.

Sherman, B., (1957), Percentiles of the ω_n statistics, *Ann. Math. Stats.*, 28:259, 1957.

Shorack, G.R., Wellner, J.A., (1986), *Empirical processes with application to statistics*, Section 8, John Wiley & Sons, 1986, ISBN 0-471-86725.

Sommerfeld, A., (1911), *Jahrb. drahtlosen Telegraphie*, Vol. 4, p 157.

Stratton, J.A., (1930), The effect of rain and fog on very short radio waves, *Proc. IRE*, Vol. 18, pp 1064-1075.

Suhir, E., (1997), *Applied Probability for Engineers and Scientists*, McGraw-Hill, New York, ISBN 0-07-061860-7, Section 10.

Taguchi, G., (1986), *Introduction to Quality engineering*, Asian Productivity Organization, American Supplier Institute, Dearborn, Michigan, 1986.

Watson, G.N., (1919), *Proc. Royal Society*, Vol. 95, 83, 1919.

Watson, P.A., (1990), Statistical methods for representing extreme radio propagation events for systems performance prediction, *XXIII URSI General Assembly*, September, Prague.

Yeh, Y.S., Schwartz, S.C., (1984), Outage probability in mobile telephony due to log-normal interferers, IEEE Transactions on Communications, COM-32, April, 388.

Recommendations of ITU-R: P Series, Volume 1997

The P-series (Propagation) Volume 1997 contains some 62 Recommendations, here we list only some of the Recommendations which are of interest for our thesis. Note the date after the Recommendation number indicates the last time that the Recommendation has been modified.

Order information can be found at World Wide Web: www.itu.int/publications/.

P.310-9 (11/93) Definitions of terms relating to propagation in non-ionized media

P.368-7 (03/92) Ground-wave propagation curves for frequencies between 10 kHz and 30 MHz.

P.370-7 (10/95) VHF and UHF propagation curves for the frequency range from 30 MHz to 1 000 MHz. Broadcasting services.

P.452-8 (08/97) Prediction procedure for the evaluation of microwave interference between stations on the surface of the Earth at frequencies above about 0.7 GHz.

P.529-2 (10/95) Prediction methods for the terrestrial land mobile service in the VHF and UHF bands.

P.530-7 (08/97) Propagation data and prediction methods required for the design of terrestrial line-of-sight systems.

P.581-2 (06/90) The concept of "worst month".

P.618-5 (05/97) Propagation data and prediction methods required for the design of Earth-space telecommunication systems.

P.619-1 (03/92) Propagation data required for the evaluation of interference between stations in space and those on the surface of the Earth.

P.620-3 (08/97) Propagation data required for the evaluation of coordination distances in the frequency range 0.85-60 GHz.

P.678-1 (03/92) Characterization of the natural variability of propagation phenomena.

P.832-1 (08/97) World atlas of Ground Conductivities

P.841 (03/92) Conversion of annual statistics to worst-months statistics.

P.1057 (11/93) Probability distributions relevant to radio-wave propagation modelling

P.1058-1 (05/97) Digital topographic databases for propagation studies

P.1144 (10/95) Guide to the application of the propagation methods of Radiocommunication Study Group 3

P.1145 (10/95) Propagation data for the terrestrial land mobile service in the VHF and UHF bands

P.1146 (10/95) The prediction of field strength for land mobile and terrestrial broadcasting services in the frequency range from 1 to 3 GHz.

Appendix to Section 2

A2.1. Data used in testing

(1). 11 GHz Slant-path data

The data here are the 11-GHz slant-path rain-attenuation data which were produced within the European co-operative project "Influence of the atmosphere on Earth-satellite radio propagation at frequencies above 10 GHz" (the COST205 project, ref. [COST205, 1985]) in the early 80's. The data came from experiments with the OTS and SIRIO satellites as well as radiometers. Although the number of sites involved is large (some 30 sites) the average measurement periods is about 2.5 years which falls short of the minimum requirement of 5 years established in Section 2.4.1. Therefore, in order to maximise the use the available data a pooling scheme must be applied.

In this scheme we have assumed that data obtained from different sites located 'not to far apart' and whose radio link parameter are 'similar' may be considered to be samples from the same random process and may therefore be pooled together. We also applied a small scaling of the data when required (frequency, polarisation and elevation scaling using methods given in Recommendation ITU-R P618. The maximum range of frequency scaling is from 11.4GHz (Vertical polarization) to 11.8GHz (Circular polarization), while the maximum range of elevation scaling is from 33 to 45 degrees. A small correction, of some 0.25 dB, is also applied when the satellite attenuation data does not include the background cloud and gaseous attenuation. The procedure of the scaling is to first calculate new values for (attenuation) threshold values using the aforementioned ITU models. Afterwards the values of the measured

cumulative distribution for the original (attenuation) threshold values are determined through interpolation.

This scheme results in 8 sets of pooled data as shown in Table 2.2 of Section 2. The size of the data sets varies from 5 years (Finland) to 12 years (Northern Italy). In total we have 69 site years of measurements. The thresholds considered in this thesis are: 1, 2, 3, 4, 6, 8, 10, 12 dB.

(2) Trans-horizon Data

The data here are those obtained during the European project "Influence of the atmosphere on interference between radio communication systems at frequencies above 1 GHz" (project COST210, ref. [COST210, 1990]) carried out in the period 1984 to 1990. Although the COST210 project has collected a very large amount of data only a small subset of this can be used for our investigation. This is due to two reason. The first is because many measurements do not reach our objective for a minimum observation period of 5 years. The second is because not all the data have been given in the form of individual monthly distributions. The COST210 data used in this paper are all obtained from long-distance measurements of a 1.3 GHz beacon transmitted from Martlesham Heath in the UK and from one link at 6.4 GHz (Irnsum – Leidschendam, 5 years of measurements). The 1.3GHz radio links involved all traverses for a large part the English Channel. The Irnsum – Leidschendam link traverses the Ijselmeer. For the Irnsum-Leidschendam data we have applied a replacement correction to correct for the fact that there has been one month of measurement failure. We took advantage of the fact that on exactly the same link measurements were at that time performed at 11 GHz. For the missing month we substituted the data from the 11 GHz measurements. According to a correlation analysis the statistical distributions at these two frequencies seems to be very similar.

Finally, the thresholds used are: -45, -35, -25, -15, -10, -5, 0, +5 dB with respect to the (theoretical) free space level. Table 2.3 in Section 2 summarises the data.

A2.2. Expressions for long-term average of X

As noted in Section 2.2 the long-term average, $\overline{X}$, of X can be calculated from the aggregate distribution P_X using Equation (2.17). By using this equation the expression for the long-term average for the four models considered in this thesis has been derived.

For the Conditional Exponential model (CEX):

$$\overline{X} = \begin{cases} C_0 C_1 & \text{(for } C_0 \leq 1) \\ C_1 + C_1 \ln(C_0) & \text{(for } C_o > 1) \end{cases}$$

(A2.1)

For the Conditional Lognormal model (CLN):

$$\overline{X} = C_0 C_1 \exp(C_2^2 / 2)$$

(A2.2)

For the Gamma type of model (GM and SGM):

$$\overline{X} = C_1 [C_2 \int_\kappa^\infty dt\, t^{C_2} \frac{e^{-t}}{\Gamma(1 + C_2)} - \kappa C_{0,eff}]$$

(A2.3)

where, by definition, κ is 0 for the GM and:

$$\kappa = C_2(1 - \sqrt{C_2}) H(C_2)$$

(A2. 4)

for the SGM model. H is the unit step function ($H(X) = 0$ for $X < 0$, and $H(X) = 1$ for $X \geq 0$). Further, $C_{0,eff} = 1$ in the case of the GM model, for the SGM model $C_{0,eff}$ is given by:

$$C_{0,eff} = \int_\kappa^\infty dt\, t^{C_2 - 1} \frac{e^{-t}}{\Gamma(C_2)}$$

(A2. 5)

A2.3. Bin construction for the χ^2 test

The χ^2 distribution is obtained as an asymptotic case ($N \to \infty$). According to *Cramer* [1945] the χ^2 square test is still accurate as long as we have at least 10 samples per bin, *Kreysig* [1970] takes 5 samples per bin as the required minimum. Further, since *Number of Degrees of Freedom (NDF)* must be ≥ 1 we have the following requirements on the number of bins to be used:

$$r \geq s + 2 \qquad (A2.6)$$

For the conditional lognormal model, for example, $s = 3$ so that the minimum number of bins to be used is $r = 5$. Notwithstanding these two restrictions there are usually still many ways to apply the test since many different bin constructions can be made. For this study we have chosen a bin construction which will ensure severe testing of the tail of the distribution: the highest bin (bin nr. r) shall be that which contains only the 5 largest samples, the second highest bin shall be that which contains only the 5 samples with ranks 6 to 10, etc.. The lowest bin (bin nr 1) shall contain the (discontinuity) point $X = 0$. If the lowest bin contains less than 5 points it shall be merged with the second lowest bin.

A2.4 Implementation of *Crane*'s method for determination of C_0

A method has devised by [*Crane*, 1991] to determine in a visual way the value of C_0 from a given sample distribution function, say $P_X(X_i)$ with X_1 the largest value etc. Crane's method consists of first plotting $1 - P_X(X_i)$ versus $\log(X_i)$. Obviously only points with $X_i > 0$ can be plotted this way. The method is based the observed fact that often the measured data, in such plots, would behave in a straight line across a large range and that for small values of X_i the data will strongly deviate from that straight line. See the example of Figure A2.1.
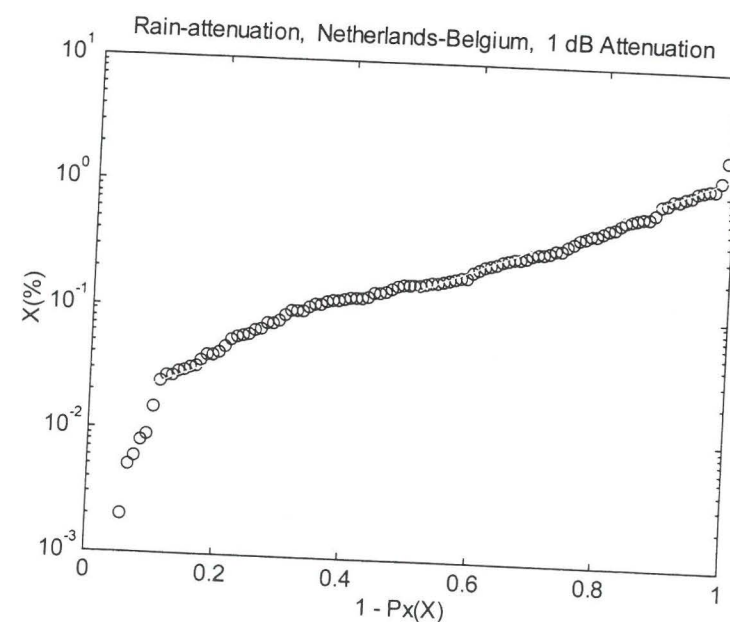

Figure A2.1 example of data plot for determination of C_0

According to Crane value of C_0 would correspond to the position of the "breakpoint". We have implemented this method in software. Basically we first determine a linear best-fit line. The "breakpoint" is then found by removing points below the line, starting with points with the smallest value, until the first point above the regression line is found.

Appendix to Section 3

A3.1 Ω_Y versus C_2 for the SGM model

For the SGM model the analytical relationship between Ω_Y and C_2 can easily be derived by using Equation (3.10) in Section 3, with $M = 12$, and calculating the 1st and 2nd moments of X etc.. The expression for the 1st moment of X has already been given in Section 2. The second moment of X is given by:

$$< X^2 > = C_1[(1+C_2)C_2 P_{GM}(\kappa, 2+C_2) - 2\kappa C_2 P_{GM}(\kappa, 1+C_2) + \kappa^2 P_{GM}(\kappa, C_2)] \tag{A3.1}$$

This equation, together with the equations for $<X>$ and Ω_Y, see Section 2 & 3, would result in a cumbersome equation for Ω_Y. A simple approximation has however been found, see Equation (3.16) in Section 3. Figure A3.1 shows the performance of this approximation.
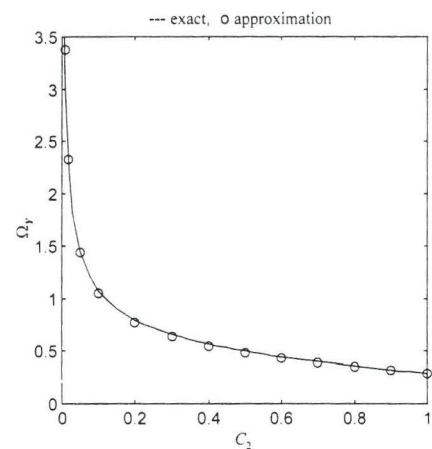

Figure A3.1 Ω_Y vs. C_2 for SGM model.

A3.2 The distribution P_Y for the SGM model

Here we determine the distribution of P_Y through the Monte-Carlo method.

The simulation has been done using MATLAB random generator for the Gamma variate g. For the SGM model X is related to a Gamma random variate g (with scale parameter C_1 and shape parameter C_2) through a shift and truncation process:

$$X = (g - \kappa C_1) H(g - \kappa C_1) \qquad\qquad (A3.2)$$

with H the Heaviside unit step function, and κ as given in Section 2.

We performed this simulation for various values of C_2 ranging from 0.01 to 1.0. For each chosen value of C_2 a run of 1000 years simulation was produced, from which the (empirical) distribution of P_Y were determined. Subsequent analysis of these distributions results in the approximation discussed in Section 3.3.4. The figures below give examples of comparisons of the approximation formula for P_Y and the distributions obtained through simulation.
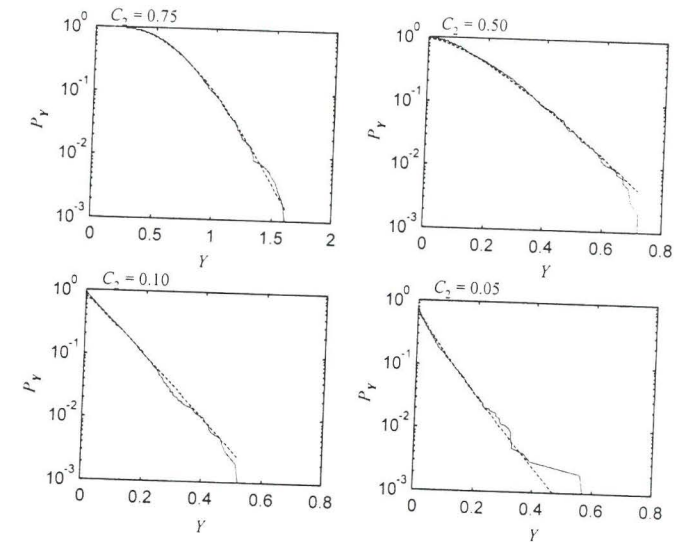
Figure A3.2. P_Y for SGM model. _____ from simulation, - - - approximation

Appendix to Section 4

A4.1 Equations for Q_1

Mathematically the ITU worst-month quotient Q_1 is defined as:

$$Q_1 = <\overline{W}> / <\overline{X}> \qquad (A4.1)$$

Here the average month value can be calculated using either the individual monthly probability density functions $f_{X|m}$ ($m = 1, 2, ..., 12$) or the average probability density function f_X:

$$<\overline{X}> = \frac{1}{12}\sum_{m=1}^{12}\int_0^\infty dX\, X\, f_{X|m}(X) = \int_0^\infty dX\, X\, f_X(X) \qquad (A4.2)$$

The average worst-month is calculated from

$$<W> = \int_0^\infty dW\, W\, f_W(W) \qquad (A4.3)$$

where

$$f_W(W) = \sum_{n=1}^{12} f_{X|n}(W) \prod_{\substack{m=1 \\ m \neq n}}^{12} (1 - P_{X|m}(W)) \qquad (A4.4)$$

which shows that for the exact determination of $<W>$ either the probability density function f_W must be known, or that all the 12 monthly probability density functions $f_{X|m}$ must be known. Assuming the simplified 2-population model, see Section 2.2 and 4.3, the expression for the probability density function of W simplifies to:

$$f_W(W) = 12 \, f_X(W) \, (1 - \tfrac{12}{M} P_X(W))^{M-1} \hspace{4cm} \text{(A4.5)}$$

so that only the aggregate statistics of X are needed to calculate Q_1. By using the models for P_X (resp. f_X) in these equations and performing some mathematics we can obtain the expressions for Q_1.

For the Conditional Exponential model we have:

$$Q_{1,CEX}(C_0, M) = 12 \sum_{k=0}^{M-1} \binom{M-1}{k} \left(-\frac{12}{M} C_0\right)^k \frac{1}{(1+k)^2} \hspace{2cm} \text{(A4.6)}$$

The mathematical expression for Q_1 was first derived by *Brussaard & Watson* [1979]. Equation (A4.6) is the slightly different formulation as given by *Mawira* [1980]. This formula is valid for $0 \le C_0 \le 1$. For the cases $C_0 \ge 1$ (the extended model introduced in Section 3) the expression for Q_1 can be easily derived by noting that in such cases X is the sum of a constant term $C_1 \ln(C_0)$ and an (unconditional) exponential random variate with scale parameter C_1. This leads to the following expression of Q_1 for $C_0 \ge 1$:

$$Q_{1,CEX} = (3.10 + \ln(C_0)) / (1 + \ln(C_0)) \hspace{3cm} \text{(A4.7)}$$

For the Conditional Lognormal model we have:

$$Q_{1,CLN}(C_0, C_2, M) = 12 \sum_{k=0}^{M-1} \binom{M-1}{k} \left(-\frac{12}{M} C_0\right)^k \int_{-\infty}^{\infty} dt \, \frac{e^{-t^2/2}}{\sqrt{2\pi}} \left(\int_{t+C_2}^{\infty} du \, \frac{e^{-u^2/2}}{\sqrt{2\pi}} \right)^k \hspace{1cm} \text{(A4.8)}$$

For the Shifted Gamma model we have:

$$Q_{1,SGM}(C_2, M) = \frac{12}{\eta} \sum_{k=0}^{M-1} \binom{M-1}{k} \left(-\frac{12}{M}\right)^k \int_{\kappa}^{\infty} dt \, (t-\kappa) \, t^{C_2-1} \frac{e^{-t}}{\Gamma(C_2)} \left(\int_{t}^{\infty} du \, u^{C_2-1} \frac{e^{-u}}{\Gamma(C_2)} \right)^k \hspace{0.5cm} \text{(A4.9)}$$

with

$$\eta = C_2 \left[\int_{\kappa}^{\infty} dt \, t^{C_2} \frac{e^{-t}}{\Gamma(1+C_2)} - \kappa \int_{\kappa}^{\infty} dt \, t^{C_2-1} \frac{e^{-t}}{\Gamma(C_2)} \right] \hspace{2cm} \text{(A4.10)}$$

κ is a function of C_2 as given by Equations (2.9) of Section 2. For the unconditional Gamma model the expression is the same but with $\kappa = 0$.

Note that Q_1 does not depend on the scale parameter C_1. Theoretically the conditional lognormal model has the largest possible ranges of values of Q_1. The upper bound is achieved when the shape parameter $C_2 = \infty$:

$$Q_{1,CLN}(C_0, \infty, M) = 12 \hspace{4cm} \text{(A4.11)}$$

The lower bound is achieved when the shape parameter $C_2 = 0$ and $M = 12$:

$$Q_{1,CLN}(C_0, \, 0, \, 12) = (1 - (1 - C_0)^{12}) / C_0 \hspace{3cm} \text{(A4.12)}$$

A4.2 Determination of the moments of q_N

In this section we shall derive the expressions for the moments of q_N, which is the random variate estimator of Q_1 obtained from N years of observation of the random process X monthly t.f.e.'s. A conditional exponential model (CEX) is assumed for the statistical distribution of X:

$$P_X(X) = C_0 \, e^{-X/C_1} \tag{A4.13}$$

with $0 < C_0 \leq 1$ and $C_1 > 0$. Without loss of generality we can take $C_1 = 1$ so that now we have for P_X:

$$P_X(X) = C_0 \, e^{-X} \tag{A4.14}$$

Furthermore we assume the simplified 2-population model discussed in Section 2.2 and 4.3. In this model the events $X > 0$ are only produced in a subset of M months of a year, the *Active* months, with probability distribution within each of this month given as:

$$P_X(X \mid X \in Active) = \text{probability}(X \geq X \mid X \in Active \text{ month}) = \frac{12}{M} C_0 \, e^{-X} \tag{A4.15}$$

for $X > 0$; for $X < 0$, $P_X = 1$.

A4.2.1 A generic case

Before solving the whole problem we shall first consider a similar but much more simplified situation. Here we consider an artificial world where the year is M months long ($= 1, 2, ...$) and where random process of the monthly t.f.e. X is statistically homogeneous, independent and of the exponential type:

$$P_X(X) = e^{-X} \tag{A4.16}$$

for $X > 0$; for $X <= 0$, $P_X(X) = 1$. The associated probability density function is:

$$f_X(X) = e^{-X} \, U(X) \tag{A4.17}$$

where U is the unit step function.

The worst month random process W here is defined analogous to that defined for real years (accounting for the deviant length of the year). We first consider the ratio of the worst-month to the year total, r_1:

$$r_1 = \frac{W}{\sum_{m=1}^{M} X_m} \tag{A4.18}$$

Our goal is to calculate the *(statistical) moments* of r_1, we need therefore first to obtain the joint probability of W and the other months. We notice that the worst month can occur, with equal probability, in any of the M months. The M possible cases (the worst-month is month 1, the worst-month is month 2, ...) are statistically equivalent, so that it is sufficient to analyse the structure of the joint-probabilities for the case that the worst-month occurs in the first month of the year, $W = X_1$. In this case, the probability density of observing the sequence of values $(W, X_2, X_3, ... , X_M)$:

$$f_X(W) \prod_{m=2}^{M} f_X(X_m) U(W - X_m) \tag{A4.19}$$

with U the unit step-function. For the case when the worst month is the month k, $W = X_k$, the expression for the joint-probability is similar as above, only some index is modified. To obtain a compact notation we should define the random process associated with the months "which are not the worst-month" V. We can obtain this by a process of renaming. The joint-probability density function, for the complete case (worst-month in 1st month, worst-month in 2nd month etc.) is now:

$$f_{W,V}(W, V_1, ..., V_{M-1}) = M f_X(W) \prod_{m=1}^{M-1} f_X(V_m) U(W - V_m) \tag{A4.20}$$

and for r_1 we now have:

$$r_1 = \frac{W}{W + \sum_{m=1}^{M-1} V_m} \qquad (A4.21)$$

for $M \geq 2$; for $M = 1$ we have the trivial case $r_1 = 1$.

The k-th *moment* of r_1 is defined as:

$$< r_1^{\;k} > \; = \; M \int_0^\infty dW \int_0^W dV_1 \; ... \int_0^W dV_{M-1} \; f_{W,\underline{V}}(W, V_1, ..., V_{M-1}) \left(\frac{W}{W + \sum_{m=1}^{M-1} V_m} \right)^k \qquad (A4.22)$$

Inserting Equation (A4.20) and Equation (A4.17) we obtain:

$$< r_1^{\;k} > \; = \; M \int_0^\infty dW \; e^{-W} \int_0^W dV_1 \; ... \int_0^W dV_{M-1} \; e^{-\sum_{i=1}^{M-1} V_i} \left(\frac{W}{W + \sum_{m=1}^{M-1} V_m} \right)^k \qquad (A4.23)$$

and by applying a variable transform $s_i = V_i/W$:

$$< r_1^{\;k} > \; = \; M \int_0^\infty dW \; W^{M-1} \; e^{-W} \int_0^1 ds_1 \; ... \int_0^1 ds_{M-1} \; e^{-W \sum_{i=1}^{M-1} s_i} \left(1 + \sum_{m=1}^{M-1} s_m \right)^{-k} \qquad (A4.24)$$

By carrying out the integration through W, and using well known properties of the Gamma function [Abramovitch, 1970], the equation is reduced to:

$$< r_1^{\;k} > \; = \; M! \int_0^1 ds_1 \; ... \int_0^1 ds_{M-1} \left(1 + \sum_{m=1}^{M-1} s_m \right)^{-M-k} \qquad (A4.25)$$

We solve this equation by using recurrence equations, to do this consider the following function:

$$I_M^{(k)}(\alpha) \; = \; \int_0^1 ds_1 \; ... \int_0^1 ds_{M-1} \left(\alpha + \sum_{m=1}^{M-1} s_m \right)^{-M-k} \qquad (A4.26)$$

with $\alpha > 0$; note that

$$< r_1^{\;k} > = I_M^{(k)}(1) \qquad (A4.27)$$

By performing the integration across s_{M-1} in Equation (A4.26) we obtain the following recurrence relationship:

$$I_M^{(k)}(\alpha) \; = \; \tfrac{1}{M-1+k}[I_{M-1}^{(k)}(\alpha) \; - \; I_{M-1}^{(k)}(1+\alpha)] \qquad (A4.28)$$

An initial value for solving the recurrence relationship can be found by solving the integral in Equation (A4.26) for $M = 2$, this leads to:

$$I_2^{(k)}(\alpha) \; = \; \tfrac{1}{1+k}[\tfrac{1}{\alpha^{1+k}} \; - \; \tfrac{1}{(1+\alpha)^{1+k}}] \qquad (A4.29)$$

By comparing this equation with Equation (A4.28) we can see that we can define in a consistent way:

$$I_1^{(k)}(\alpha) \; = \; \tfrac{1}{\alpha^{k+1}} \qquad (A4.30)$$

By applying the above recurrence successively relationship we obtain:

$$I_M^{(k)}(\alpha) \; = \; \tfrac{1}{(k+1)_{M-1}} \sum_{i=0}^{M-1} \binom{M-1}{i} (-1)^i \; \tfrac{1}{(\alpha+i)^{k+1}} \qquad (A4.31)$$

where $(u)_l$ is de Pochhammer's notation for $u(u+1)...(u+l-1)$, also $(u)_0 := 1$.

By setting $\alpha = 1$ we obtain:

$$< r_1^{\;k} > \; = \; \tfrac{M!}{(k+1)_{M-1}} \sum_{i=0}^{M-1} \binom{M-1}{i} (-1)^i \; \tfrac{1}{(1+i)^{k+1}} \qquad (A4.32)$$

for $M \geq 1$ and $k \geq 0$.

An interesting property which can be proven is that:

$$< r_1^{\,k} \; s_M^{\,n} > \;\; = \;\; < r_1^{\,k} > < s_M^{\,n} > \tag{A4.33}$$

for $n, k \geq 0$, and where

$$s_M \;\; = \;\; \sum_{m=1}^{M} X_m \tag{A4.34}$$

In other words r_1 and s_M are <u>statistically independent</u>; the formal proof of this statement is given in Section A4.3.

To prove Equation (A4.81) we first obtain the expression for $< r_1^{\,k} \; s_M^{\,n} >$ by performing the integrals using the probability function $f_{W,\underline{V}}$ in a similar as was done in the previous calculation of $< r_1^{\,k} >$. After the step where a variate transfom was performed, see Equation (A4.24), we arrive here at the following expression:

$$< r_1^{\,k} \; s_M^{\,n} > \;\; = \;\; M \int_0^{\infty} dw \, w^{M+n-1} e^{-w} \int_0^1 dt_1 \; \cdots \; \int_0^1 dt_{M-1} \; e^{-w \sum_{i=1}^{M-1} t_i} \left(1 + \sum_{m=1}^{M-1} t_m \right)^{-k+n} \tag{A4.35}$$

By carrying out the integration through w and using the known properties of Gamma function [*Abramovitch*, 1970] we obtain:

$$< r_1^{\,k} \, s_M^{\,n} > \;\; = \;\; M\Gamma(M+n) \int_0^1 dt_1 \; \cdots \; \int_0^1 dt_{M-1} \; \frac{1}{\left(1 + \sum\limits_{m=1}^{M-1} t_m \right)^{M+n}} \left(1 + \sum_{m=1}^{M-1} t_m \right)^{-k+n} \tag{A4.36}$$

which reduces further to:

$$< r_1^{\,k} \, s_M^{\,n} > \;\; = \;\; M\Gamma(M+n) \int_0^1 dt_1 \; \cdots \; \int_0^1 dt_{M-1} \left(1 + \sum_{m=1}^{M-1} t_m \right)^{-M-k} \tag{A4.37}$$

By using Equation (A4.26) and Equation (A4.27) we obtain:

$$< r_1^{\,k} s_M^{\,n} > \;\; = \;\; \frac{\Gamma(M+n)}{\Gamma(M)} \; < r_1^{\,k} > \tag{A4.38}$$

Finally we note that the n-th statistical moment of s_M is:

$$< s_M^{\,n} > \;\; = \;\; \frac{\Gamma(M+n)}{\Gamma(M)} \tag{A4.39}$$

since s_M is a Gamma variate with shape parameter M and scale parameter value 1 (this is a consequence of the fact that s_M is the sum of M independent and exponentially distributed stochastic variates with common shape parameter values 1). Combining the above two equations completes the proof of Equation (A4.33).

We note here that according to [*Shorack & Wellner*, 1986] r_1 is the maximum spacing sampled from M i.i.d. r.v.'s uniformly distributed between 0 and 1. Using this fact the moments of r_1 could be readily derived.

A4.2.2 $N = 1$ year, the case when $C_0 = M/12$

In this case it is assumed that the process X in the M active months of each year is purely exponential (e.i. unconditional), so that we can directly use the results of the previous section accounting for slight difference in definitions of q_1 and r_1:

$$< q_1^{\,k} > \;\; = \;\; 12^k \; < r_1^{\,k} > \tag{A4.40}$$

or, by using Equation (A4.32):

$$< q_1^{\,k} > \;\; = \;\; g_{k,M} \tag{A4.41}$$

where we have defined:

$$g_{k,M} = 12^k \frac{M!}{(k+1)_{M-1}} \sum_{i=0}^{M-1} \binom{M-1}{i} (-1)^i \frac{1}{(1+i)^{k+1}} \qquad (A4.42)$$

for $M \geq 1$ and $k \geq 0$. We further define $g_{k,0} = 12^k$.

For $k = 1$ we obtain:

$$<q_1> = 12 \sum_{i=0}^{M-1} \binom{M-1}{i} (-1)^i \frac{1}{(1+i)^2} \qquad (A4.43)$$

This expression is exactly the same as the expression for Q_1 (with $C_0 = M/12$), see Equation (A4.6) in section A4.1. This means that the random variate q_1 is an <u>unbiased</u> estimator for Q_1 for the cases where $C_0 = M/12$. Another interesting property is that in such cases q_1 and m_X are <u>statistically independent</u> (since r_1 and s_M are statistically independent, see Section A4.2.1).

A4.2.3 N = 1 year, the general case

In general C_0 does not have to be equal to $M/12$ so that the number of months per year with $X > 0$ may vary from year to year. The random process X, <u>within</u> the M active monthly population, can actually be considered to be the product between two random processes:

$$X = U \, X_e \qquad (A4.44)$$

where X_e represents an unconditional standard exponential process, while U is a binary random process (producing only values of 0 or 1) with probability:

$$p_U(1) = probability(U = 1) = \frac{12}{M} C_0 \qquad (A4.45)$$

and $p_u(0) = 1 - p_u(1)$.

From this we can see that the ensemble of yearly sequences of X may be composed of $M+1$ subsets of yearly sequences: those with M months having $X > 0$, those with $M-1$ having $X > 0$,

...., those with 0 months having $X > 0$. The probability of obtaining a yearly sequence with say m months having $X > 0$ is:

$$\text{probability}(m \text{ months per year with } X > 0) = B_M(m, p_U(1)) \qquad (A4.46)$$

where B_M is the well-known Binomial probability function [Abramovitch, 1970]:

$$B_M(m, p) = \binom{M}{m} p^m (1 - p)^{M-m} \qquad (A4.47)$$

Note: $B_M(M,1) = 1$.

Furthermore, for the subset of yearly sequences with m months having $X > 0$ the statistical moments can be determined in exactly the same way as in the previous section, we have here $<q_1^k> = g_{k,m}$. The statistical moments related to the whole population are obtained by weighted averaging:

$$< q_1^k > = \sum_{m=0}^{12} B_M(m, p_U(1)) \, g_{k,m} \qquad (A4.48)$$

A4.2.4 First & second moments of q_N for N > 1

When given N years of observation the N year estimation of Q_1 is given by:

$$q_N = \frac{\sum_{j=1}^{N} W(j)}{\sum_{j=1}^{N} m_X(j)} \qquad (A4.49)$$

where, for convenience sake, we have denoted the first year of observation with $j = 1$ etc.. In this section we shall calculate the mean and the second statistical moment of q_N.

First we rewrite the above expression by substituting $W(j) = q_1(j) m_X(j)$:

$$q_N = \sum_{j=1}^{N} q_1(j) \frac{m_X(j)}{\sum_{i=1}^{N} m_X(i)} \tag{A4.50}$$

The statistical averagings are performed in two steps. In the first step we restrict ourselves to the subset of the ensemble having m_1 months with $X > 0$ in the first year, m_2 months with $X > 0$ in the second year, ..., m_N months with $X > 0$ in the final year. Here we can make use of results obtained from the previous section: the statistical independence between the q_1's and the m_X's.(for the purely exponential cases)

In the second step we perform a weighted averaging using as weights the probability of occurrences on above mentioned sub-ensembles. This probability is a product of Binomial probabilities:

$$P_{m_1, m_2, ..., m_N} = B_M(m_1, \tfrac{12}{M} C_0) B_M(m_2, \tfrac{12}{M} C_0) .... B_M(m_N, \tfrac{12}{M} C_0) \tag{A4.51}$$

The mean of q_N for this subset is given by:

$$<q_N \mid m_1, m_2, ..., m_N> = \sum_{j=1}^{N} <q_1(j)|m_j> \; < \frac{m_X(j)}{\sum_{i=1}^{N} m_X(i)} \mid m_1, m_2, ..., m_N > \tag{A4.52}$$

or

$$<q_N \mid m_1, m_2, ..., m_N> = \sum_{j=1}^{N} g_{1,m_j} < \frac{m_X(j)}{\sum_{i=1}^{N} m_X(i)} \mid m_1, m_2, ..., m_N > \tag{A4.53}$$

where we have made use of the statistical independence of $q_1(j)$ and $m_X(j)$ proven in Section A4.2.2 and the expression for $<q_1|M>$ given by Equation (A4.41). Since g_{1,m_j} has already been determined in the previous section, see Equation (A4.42), we need here only to calculate the terms:

$$L_j = \; < \frac{m_X(j)}{\sum_{i=1}^{N} m_X(i)} \mid m_1, m_2, ..., m_N > \tag{A4.54}$$

To calculate this term we make use of the fact that m_X for the different years are independent. Furthermore they are Gamma distributed with shape parameter values equal to $m_1, m_2, ..., m_N$ for respectively the first year, the second year, ..., and the final year (the scale parameter value = 1 in all cases). The expression for L_j is now as follows:

$$L_j = \int_0^{\infty} ds \int_0^{\infty} dt \frac{s^{m_j-1} t^{n-1}}{\Gamma(m_j)\Gamma(n)} e^{-s-t} \frac{s}{s+t} \tag{A4.55}$$

with $n = m_1 + m_2 + ... + m_N - m_j$. Here we have made use of the fact that the sum of a number of independent Gamma variates with equal scale parameter values but with possibly differing shape parameter values is a Gamma variate with shape parameter value equal to the sum of the individual shape parameter values; the resulting shape parameter value is equal to the original scale parameter value.

To solve this integral we first perform a probability variate transform $s = u^2$ and $t = v^2$. Subsequently a polar transform is performed in the u, v space and the integral along the radius is performed, this results in:

$$L_j = 2 \frac{\Gamma(m_j + n)}{\Gamma(m_j)\Gamma(n)} \int_0^{\pi/2} d\phi \; \cos^{2m_j+1}(\phi) \; \sin^{2n-1}(\phi) \tag{A4.56}$$

The solution to the integral is readily found in [Abramovitch, 1975] in the section regarding Gamma and Beta functions:

$$L_j = \frac{\Gamma(m_j + n)}{\Gamma(m_j)\Gamma(n)} \frac{\Gamma(m_j + 1)\Gamma(n)}{\Gamma(m_j + n + 1)} = \frac{m_j}{m_j + n} = \frac{m_j}{\sum_{i=1}^{N} m_i} \tag{A4.57}$$

This results in the following expression for the conditional average:

$$<q_N | m_1, m_2, ..., m_N> = \sum_{j=1}^{N} g_{1,m_j} \left(\frac{m_j}{\sum_{i=1}^{N} m_i} \right) \quad \text{(A4.58)}$$

The unconditional mean is obtained by averaging of the above expression using the multiple Binomial expression given by Equation (A4.51):

$$<q_N> = \sum_{m_1=0}^{M} ... \sum_{m_N=0}^{M} P_{m_1, m_2, ..., m_N} <q_N | m_1, m_2, ..., m_N> \quad \text{(A4.59)}$$

For the special case when $C_o = M/12$ the expression can be simplified further. In this case $P_{m_1, m_2, ..., m_N} = 1$ when all the m_j's are equal to M and zero otherwise so that:

$$<q_N> = <q_N | M, M, ..., M> \quad \text{(A4.60)}$$

which as can be easily seen from Equation (A4.58) leads to:

$$<q_N> = g_{1,M} = Q_1 \quad \text{(A4.61)}$$

This shows that for these special cases q_1 is an unbiased estimator for the worst month quotient Q_1.

Now we shall determine the expression for the second statistical moment of q_N:

$$<q_N^2> = < \sum_{j=1}^{N} \sum_{k=1}^{N} q_1(j) q_1(k) \left(\frac{m_X(j) m_X(k)}{\sum_{i=1}^{N} m_X(i))^2} \right) > \quad \text{(A4.62)}$$

We follow here the same strategy as in the calculation of the mean of q_N. First we perform the statistical calculations for the sub-ensemble with m_1 month with $X > 0$ in the first year etc., and afterwards perform the overall averaging.

For this sub-ensemble we have:

$$<q_N^2 | m_1, m_2, ..., m_N> = \sum_{j=1}^{N} \sum_{k=1}^{N} <q_N(j) q_N(k) | m_j, m_k> L_{k,j} \quad \text{(A4.63)}$$

where

$$L_{j,k} = < \left(\frac{m_X(j) m_X(k)}{(\sum_{i=1}^{N} m_X(i))^2} \right) | m_1, m_2, ..., m_N >$$

where we have made use of the independence between the q_N's and m_X for the special situation of the sub-ensemble. The averaging of $q_N(j) q_N(k)$ is easily found using the results from Section A4.2.2:

$$<q_N(j) q_N(k) | m_j, m_k> = \delta_{j,k} g_{2,m_j} + (1 - \delta_{j,k}) g_{1,m_j} g_{1,m_k} \quad \text{(4.64)}$$

where $\delta_{j,k}$ is the Kronecker delta ($\delta_{j,k} = 1$ if $j = k$ otherwise $\delta_{j,k} = 0$).

By again exploiting the Gamma properties of the m_X's an integral expression for $L_{j,k}$ can be derived, which for the case j is not equal to k is given by:

$$L_{j,k} = \int_{0}^{\infty} \int_{0}^{\infty} \int_{0}^{\infty} dt ds du \frac{t^{m_j-1} s^{m_k-1} u^{n-1}}{\Gamma(m_j) \Gamma(m_k) \Gamma(n)} e^{-s-t-u} \frac{s t}{(s+t+u)^2} \quad \text{(4.65)}$$

where $n = m_1 + m_2 + ... + m_N - m_j - m_k$. We note here that the above equation is valid only when both m_j and m_k are > 0, in other cases $L_{j,k} = 0$.

To solve the above integral we first perform the variate transform $s = x^2$, $t = y^2$, $u = z^2$. After that we perform a polar transform and carry out the integration along the radials. This results in:

$$L_{j,k} = \frac{4\Gamma(m_j + m_k + n)}{\Gamma(m_j)\Gamma(m_k)\Gamma(n)} \int_0^{\pi/2} d\theta\, S_\theta^{2(m_j+m_k-n)-1} C_\theta^{2n-1} \int_0^{\pi/2} d\phi\, C_\phi^{2(m_j+1)-1} S_\phi^{2(m_k+1)-1} \tag{4.66}$$

with $S_q = \sin(q)$ etc.. The solution of the above integrals can be found in [*Abramovitch*, 1970] in the section on Gamma and Beta functions. Using these results we obtain:

$$L_{j,k} = \frac{\Gamma(m_j + m_k + n)}{\Gamma(m_j)\Gamma(m_k)\Gamma(n)} \frac{\Gamma(m_j + m_k + 2)\Gamma(n)}{\Gamma(m_j + m_k + n + 2)} \frac{\Gamma(m_j + 1)\Gamma(m_k + 1)}{\Gamma(m_j + m_k + 2)} \tag{4.67}$$

which can be reduced to:

$$L_{j,k} = \frac{m_j m_k}{(m_j + m_k + n)(m_j + m_k + n + 1)} = \frac{m_j m_k}{(\sum_{i=1}^N m_i)(\sum_{i=1}^N m_i + 1)} \tag{4.68}$$

valid for j not equal to k. Somewhat surprisingly the above equation produces the right answer for $m_j = 0$ and/or $m_k = 0$.

For j equal to k the integral expression is as follows:

$$L_{j,j} = \int_0^\infty \int_0^\infty ds\,dt \frac{s^{m_j-1} t^{n-1}}{\Gamma(m_j)\Gamma(n)} e^{-s-t} \frac{s^2}{(s+t)^2} \tag{4.69}$$

where $n = m_1 + m_2 + \ldots + m_N - m_j$. In a similar way as above the integral can be solved to produce:

$$L_{j,j} = \frac{m_j(m_j + 1)}{(\sum_{i=1}^N m_i)(\sum_{i=1}^N m_i + 1)} \tag{4.70}$$

We can combine Equation (4.68) and Equation (4.70) in one equation:

$$L_{j,k} = \frac{m_j(m_k + \delta_{j,k})}{(\sum_{i=1}^N m_i)(\sum_{i=1}^N m_i + 1)} \tag{4.71}$$

with $\delta_{j,k}$ the Kronecker delta. Using these expressions we obtain:

$$<q_N^2 \mid m_1, m_2, \ldots, m_N> = \sum_{j=1}^N \sum_{k=1}^N [\delta_{j,k}\, g_{2,m_j} + (1 - \delta_{j,k})\, g_{1,m_j} g_{1,m_k}] \frac{m_j\,(m_k + \delta_{j,k})}{\sum_{i=1}^N m_i(\sum_{i=1}^N m_i + 1)} \tag{4.72}$$

The unconditional values are obtained by averaging of the expression using the multiple Binomial expression given by Equation (A4.51):

$$<q_N^2> = \sum_{m_1=1}^N \cdots \sum_{m_N=1}^N P_{m_1,m_2,\ldots,m_N} <q_N^2 \mid m_1, m_2, \ldots, m_N> \tag{4.73}$$

For the special cases when $C_0 = M/12$ this equation can be simplified further in the same way as was done previously for $<q_N>$, leading to:

$$<q_N^2> = \frac{M+1}{MN+1} g_{2,M} + \frac{M(N-1)}{MN+1} g_{1,M}^2 \tag{4.74}$$

Using this equation and Equation (A4.61) we can find a rather simple expression, for these special cases, of the standard deviation for q_N. Inserting these equations in the definition for the variance

$$\sigma_{q_N}^2 = <q_N^2> - <q_N>^2 \tag{4.75}$$

we obtain:

$$\sigma_{q_N}^2 = \tfrac{M+1}{MN+1} g_{2,M} + \tfrac{M(N-1)}{MN+1} g_{1,M}^2 - g_{1,M}^2 \qquad (4.76)$$

or

$$\sigma_{q_N}^2 = \tfrac{M+1}{MN+1}(g_{2,M} - g_{1,M}^2) = \tfrac{M+1}{MN+1}\sigma_{q_1}^2 \qquad (4.77)$$

A4.3 Formal proof of independence

In this section we give the proof of the statement that if for two random variates, say X and Y, their joint moments are separable:

$$< X^m Y^n > \;=\; < X^m >< Y^n > \qquad (A4.78)$$
$$(\text{for } n, m = 0, 1, 2, \ldots\ldots)$$

the X and Y are statistically independent:

$$f_{X,Y}(X,Y) \;=\; f_X(X) f_Y(Y) \qquad (A4.79)$$

where $f_{X,Y}$ is join-probability density function, while f_X and f_Y the respective marginal probability density functions.

To do this we use the characteristic functions Φ_X, Φ_Y, $\Phi_{X,Y}$ defined as [Papoulis, 1984]:

$$\Phi_X(\omega_X) = \int_{-\infty}^{\infty} dX \, f_X(X) \, e^{j\omega_X X} \qquad (A4.80)$$

$$\Phi_Y(\omega_Y) = \int_{-\infty}^{\infty} dY \, f_Y(Y) \, e^{j\omega_Y Y} \qquad (A4.81)$$

$$\Phi_{X,Y}(\omega_X,\omega_Y) = \int_{-\infty}^{\infty} dX \int_{-\infty}^{\infty} dY \, f_{X,Y}(X,Y) \, e^{j(\omega_X X + \omega_Y Y)} \qquad (A4.82)$$

According to [Papoulis, 1984] if the joint characteristic function $\Phi_{X,Y}$ equal the product of the marginal characteristic functions Φ_X and Φ_Y:

$$\Phi_{X,Y}(\omega_X,\omega_Y) \quad = \quad \Phi_X(\omega_X)\Phi_Y(\omega_Y) \tag{A4.83}$$

then X and Y are statistically independent.

To prove this we first expand the above expressions as power series:

$$\Phi_{X,Y}(\omega_X,\omega_Y) = \sum_{n=0}^{\infty}\frac{1}{n!}\sum_{k=0}^{n}\binom{n}{k}<X^kY^{n-k}>(j\omega_X)^k(j\omega_Y)^{n-k} \tag{A4.84}$$

$$\Phi_X(\omega_X) = \sum_{k=0}^{\infty}\frac{1}{k!}<X^k>(j\omega_X)^k \tag{A4.85}$$

$$\Phi_Y(\omega_Y) = \sum_{n=0}^{\infty}\frac{1}{l!}<Y^l>(j\omega_Y)^l \tag{A4.86}$$

Now by using Equation (A4.78) in the above expression for $\Phi_{X,Y}$ we obtain:

$$\Phi_{X,Y}(\omega_X,\omega_Y) = \sum_{n=0}^{\infty}\frac{1}{n!}\sum_{k=0}^{n}\binom{n}{k}<X^k><Y^{n-k}>(j\omega_X)^k(j\omega_Y)^{n-k} \tag{A4.87}$$

Now we can rearrange the order of summation:

$$\Phi_{X,Y}(\omega_X,\omega_Y) = \sum_{k=0}^{\infty}\sum_{n=k}^{\infty}\frac{1}{n!}\binom{n}{k}<X^k><Y^{n-k}>(j\omega_X)^k(j\omega_Y)^{n-k} \tag{A4.88}$$

or if we apply the explicit expression for the binomial factor (.):

$$\Phi_{X,Y}(\omega_X,\omega_Y) = \sum_{k=0}^{\infty}\sum_{n=k}^{\infty}\frac{1}{k!(n-k)!}<X^k><Y^{n-k}>(j\omega_X)^k(j\omega_Y)^{n-k} \tag{A4.89}$$

The next step is to replace n with a new summing index $l = n-k$:

$$\Phi_{X,Y}(\omega_X,\omega_Y) = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\frac{1}{k!l!}<X^k><Y^l>(j\omega_X)^k(j\omega_Y)^l \tag{A4.90}$$

By comparing this equation with Equations (A4.85) and (A4.86) we can easily see that:

$$\Phi_{X,Y}(\omega_X,\omega_Y) \quad = \quad \Phi_X(\omega_X)\Phi_Y(\omega_Y) \tag{A4.91}$$

which completes the formal proof of statistical independence. Note that this proof requires that the series must be convergent. E.g. for heavy tailed distributions this proof is not valid because the higher moments becomes infinite for finite (moment) order.

Samenvatting

Een belangrijke eis ten aanzien van troposferische propagatie modellen is hun vermogen om de lange duur statistieken van diverse propagatie grootheden te kunnen voorspellen. Voorts moet de voorspelling over een zeer groot kansbereik werken. Net zo belangrijk, zeker voor systeemontwerpers, is het vermogen van een dergelijk model om de variaties aan te kunnen geven van de statistische verdeling bij kortere duur observaties. Het modelleren van dergelijke variaties is het onderwerp van dit proefschrift. Hier wordt aandacht besteed aan een bijzondere klasse van propagatie grootheden, namelijk de zogenaamde maandelijkse overschrijdingsduur; in het Engels de "monthly time fraction of excess" (TFE). Dergelijke TFE speelt een belangrijke rol bij de zogenaamde "Grade of Services" specificaties van de ITU (International Telecommunication Union). Het onderzoek wordt beperkt tot de TFE geassocieerd met ten eerste regen-demping op satellietpaden en ten tweede ducting op over-de-horizon radioverbindingen.

Ter inleiding wordt een studie gemaakt naar de algemene statistische aard van de maandelijkse TFE. Er wordt een theoretisch raamwerk opgesteld waarbij de TFE als statistisch proces wordt gemodelleerd. Hierbij fungeert als "ensemble" van het proces de hypothetische verzameling van alle verbindingen in een, klimatologisch gezien, homogeen gebied. Alle verbindingen hebben uiteraard een zelfde parametrische instelling (frequentie, padlengte etc.). Analyses, die gepleegd zijn met behulp van de meetgegevens, laten zien dat als werkhypothese jaarlijkse cyclostationariteit van het proces mag worden verondersteld. Verder rechtvaardigen de resultaten van correlatie analyses de aanname van onafhankelijkheid van het proces. De seizoensafhankelijkheid van dergelijke propagatie processen, die meestal complex is, wordt opgevangen door invoering van het zogenaamde duale-populatie model.

In het tweede deel van het onderzoek wordt aandacht besteed aan het verkrijgen van een goed model voor de statistische verdeling van de maandelijkse TFE. Vier kandidaatmodellen worden onderzocht. Het betreft het (geconditioneerde) exponentiële model (CEX), het lognormale model, het gamma model (GM), en het verschoven gamma model (SGM). Diverse klassieke toetsingmethoden worden toegepast zonder dat het duidelijk uitsluitsel oplevert. Een nieuwe toetsingmethode is derhalve ontworpen; deze methode is gebaseerd op invoering van een zodanige normalisatie dat alle data als één geheel beschouwd kunnen worden. De resultaten van de test met behulp van deze nieuwe methode laten overduidelijk de superioriteit van het SGM model zien; het CEX model is een goede tweede.

Vervolgens wordt aangetoond dat het gevonden statistische model voor de maandelijkse TFE toegepast kan worden om de gedragingen van zowel de jaarlijkse TFE als de slechtste maand TFE te kunnen bepalen. Interessant om te noemen is dat theoretische analyses laten zien dat de vereenvoudiging tot het duale-populatie model niet tot grote fouten zal leiden. Verder, werd de grootte van de te verwachten variabiliteit van beide grootheden bepaald en gerelateerd aan de omkeerperiode (return period). Tot slot is een classificatie van de variabiliteit eigenschappen in Europa opgesteld op basis van de regen-demping gegevens. Er worden vijf klimaatzones onderscheiden, namelijk Scandinavië, Noordwest Europa, Centraal Europa, de Alpen en tot slot het Middellandse Zee gebied.

In het laatste deel van het proefschrift wordt ingegaan op mogelijke innovaties ten aanzien van manieren om kwaliteit en prestatie van radioverbindingen te kwantificeren. Besproken wordt de mogelijke merites van methodieken en concepten uit de "reliability engineering" en meteorologie. Tot slot wordt een voorbeeld uitgewerkt van een toepassing van het gevonden statistisch model voor de maandelijkse TFE op een verzekeringsgeval voor een telecommunicatiesysteem.

Curriculum Vitae

Agus Mawira was born in Jakarta, Indonesia in 1947. He graduated with a degree in Electrical Engineering at the Eindhoven University of Technology in 1974. From 1974 to 1975 he was with the Radio Communication Department of the Eindhoven University of Technology where he worked on the theory of the polarisation of thermal emission of rain at microwave frequencies. He joined KPN Research, then called the Dr. Neher Laboraries, in 1975 where he started with work on the experimental and theoretical characterisation of rainfall attenuation at microwave links. He has since worked for KPN Research on various propagation research ranging from satellite and slant-path propagation, rain-scatter, mobile propagation at VHF and UHF. He has been active in various European co-operative projects: the COST25/5 project for microwave terrestrial propagation, the COST205 project for slant-path propagation, the COST207 project for mobile communications, the COST210 project on interference between fixed radio systems, the RACE project 2029 on UMTS and also in projects headed by the European Space Agency such as OPEX. Within the project COST205 he was chairman on the working group dealing with precipitation aspects. Within OPEX he was chairman of the working group dealing with data acquisition and pre-processing. From 1990 to 1994 he was National co-ordinator and representative for the Netherlands for ITU CCIR Study Groups 5 & 6. During the last few years his research activities has been focused on the development of propagation models for the planning of cellular systems, at first for the GSM systems, now more for the future UMTS systems.

Acknowledgements

The work presented in this thesis is related to activities that has spanned more than a decade and has involved many people and institutes. Continuous during this period has been the support of various divisions of KPN for these activities. These divisions are the operating Telecom divisions of KPN and also KPN Research. I would like to thank all the decision-makers who had the foresight and patience in supporting our research activities.

Many of the research have been carried out within the framework of co-operations. Such co-operation range from national (Netherlands) through the European (e.g. the COST projects) and global (through the ITU – R Studygroups). I would like to thank all the people and institutes involved. In particular I would like to thank all the people who worked for the COST205 and COST210 projects, without their effort the valuable and extensive propagation database would not be available. I would also like to mention my appreciation for the special long term co-operation between KPN Research and Delft University of Technology, through prof. Leo Ligthart, and the Eindhoven University of Technology, through prof. Gert Brussaard.

I am also indebted to many people within KPN and KPN Research with whom I have had a fruitful co-operation over the many years. On a personal note I would like to thank Jan Neessen who has been a mentor to me during my starting year at KPN Research. He also was the one who got my interest for the subject of "worst-month". Special thanks also to Ad Labrujere who helped me get started to work on my thesis and to Cees van Bochove who as my boss has throughout given me support for my work.

I would also like to thank every person on the PhD commission for their efforts. I would especially like to thank those members, who caused me a lot of additional work, they are Gert

Brussaard, John Einhmal, Matti Herben and Leo Ligthart. I would like to thank Gert Brussaard, not only for accepting me as his PhD student, but also for the very many, sometimes heated but mostly fruitful, discussions we have had on various propagation subjects in the last decennia.

With regard to the manuscript I would like to thank Jan Trompert of the infrastructure department for design of the cover and for complying with my unreasonably tight schedule.

Finally, I would like to thank my family and friends who in the last year patiently put up with my excuses: "so sorry, I am too busy with my thesis".

Stellingen behorende bij het proefschrift

Variability of monthly
time fraction of excess of
atmospheric propagation parameters

door

A. Mawira

Eindhoven, 13 September 1999

1. Van propagatie achter de horizon, zoals die door ducting wordt veroorzaakt, is de invloed op cellulaire radiosystemen nog onvoldoende onderzocht.

2. De benaderingsformule gegeven door [*Abramowitz*, 1978] om de inverse van de Onvolledige Beta-functie, $I_X(a, b)$, te berekenen, levert verkeerde waarden op voor de gevallen $a = 1$ en/of $b = 1$. In deze gevallen kunnen de verkregen waarden van X met een factor 0,4 tot 1,4 afwijken van de werkelijke waarden.

3. De uitvinder van het wiel met as is homo ludens.

4. Als geld de wortel van alle kwaad is dan moeten wij de overgang naar een geldloze maatschappij verwelkomen.

5. Een alternatieve aanpak voor de methode aangegeven in deze thesis is om uit te gaan van de variabiliteit van propagatiegroootheden in plaats van de fractionele overschrijdingsduur. Deze methode leidt tot complicaties en een zeer beperkte model, maar kan wel nuttige diensten bewijzen bij de presentatie van resultaten voor een specifieke toepassing.

6. Het Teledesic-project, waarmee breedband Internet zal worden verzorgd met behulp van 288 LEO satellieten, is een zakelijk risico van 9 miljard dollar.

7. Het statistisch proces van de maandelijkse TFE van atmosferische propagatiegroootheden vertoont cyclostationariteit en onafhankelijkheid. [*Dit proefschrift, Hoofdstuk* 2].

8. De afhankelijkheden van de genormeerde jaarlijkse en maandelijkse TFE van de omkeertijden en de locatie parameter C_0 zijn nagenoeg gelijk. [*Dit proefschrift, Hoofdstuk* 5].

9. Er is geen indicatie van enig fundamenteel verschil tussen de maandelijkse TFE behorende bij regendemping en die behorende bij ducting. Dit is een verrassende uitkomst, aangezien de onderliggende fysische mechanismen van de betrokken propagatiegroootheden zeer verschillend zijn. [*Dit proefschrift, Hoofdstuk* 6].

10. In een vrije-markt maatschappij moet het onderwijs zodanig zijn dat iedere burger de financiële pagina van de krant kan lezen en daarnaast zakelijke vaardigheden bezit.

11. Volgens [*Hagel* III en *Singer*, 1999] zal in de toekomst elke onderneming zich concentreren op een van de drie basisprocessen (Klantenmanagement, Innovatie, Productie/infrastructuur). Dit om een maximale efficiency te bereiken en om conflictueuze eisen te vermijden. Een probleem met deze theorie is de impliciete aanname dat Innovatie (en R&D) processen met gemak kunnen functioneren als een onafhankelijke zakelijke entiteit (b.v. het kunnen verkrijgen van investeringsgelden van de financiële markt). Innovatie behoeft vaak langdurige financiële investeringen, zonder uitzicht op korte-termijn winsten. Dit is tegengesteld aan de behoefte van de financiële markt op winst op korte termijn.
   [*Hagel III, J., Singer, M., 1999, Unbundling the corporation, Harvard Business Review, March-April, pp 133 – 141.*]