

DISSERTATION
for the degree of
DOCTOR OF PHILOSOPHY
in
PHYSICS

**Analysis of the European
Wind Power Climatology and the
Possible Cosmic Radiation Forcing on
Global Lightning Activity**

PÉTER KISS

Supervisor: Prof. Imre M. Jánosi, D.Sc.

Eötvös Loránd University, Faculty of Science
Graduate School in Physics

Head: Prof. Zalán Horváth, MHAS

Statistical Physics, Biological Physics and
Physics of Quantum Systems Program

Head: Prof. Jenő Kürti, D.Sc.



Department of Physics of Complex Systems
Eötvös Loránd University
Budapest, 2009

Foreword

The understanding of our climate system is one of the most important scientific issues today. Besides the complicated computer models, which at some point are supposed to give a reliable description of the whole system, there is a great need for accurate measurements and studies, which can identify and quantitatively assess the various physical processes. Furthermore the results of the climate related studies pose some practical questions as well.

An important issue, which is tightly linked to climate speculations is our energy usage and its obvious unsustainability. Regardless of the possible effects on our climate, it seems inevitable to change fossil fuels for different energy sources in the future because of the finite reserves. Out of the few already existing alternatives, wind energy – and also solar thermal and photovoltaic energy – seem to be the most promising technologies. There is already a significant amount of wind energy integrated into the power grids all over the world, but the real breakthrough is still far away. However the fossil fuel burning power plants cannot be simply replaced by wind farms because of the wildly fluctuating wind fields. In this work we try to quantitatively assess the fluctuations of various hypothetical all-European wind power networks using different models and data-mining techniques. Our aim is to demonstrate, what the limits of the European wind resource are. We also give a detailed statistical analysis of Europe's wind climatology.

After the first part of the thesis dealing with the European wind energy possibilities, which is essentially applied research, we turn our attention to a more fundamental physical question in connection with lightning activity.

There is a long standing debate on how and to what extent cosmic radiation affects our climate. In the second part of the thesis we try to identify the connection between cosmic rays and lightning activity using similar statistical and data-mining tools as we use in the case of European wind energy calculations. Lightning activity is obviously only a tiny sub-process in the climate system, however the cosmic rays – lightning activity connection seems to be relatively well supported theoretically and its clarification could possibly lead to a better understanding of more general processes.

Contents

1 Wind energy	7
1.1 Introduction	7
1.2 The data used	10
1.2.1 Reanalysis data	10
1.2.2 Tower measurements and wind farm data	13
1.3 Models for wind speed histograms	15
1.3.1 Rayleigh distribution	15
1.3.2 Weibull distribution	18
1.3.3 Generalized gamma distribution	20
1.4 Temporal behavior of wind speed records	26
1.5 Wind profiles - height dependence of wind speed	31
1.5.1 Wind data at different levels - a consistency check	31
1.5.2 Direct comparison of 10 m - 1000 hPa data	32
1.5.3 Power law wind speed profile	37
1.5.4 Empirical wind profile	40
1.5.5 Comparison of tower measurements and reanalysis wind speeds – a Hungarian case study	43
1.6 Wind power estimations	50
1.6.1 Transformation of surface wind speed to wind power	50
1.6.2 Comparison of empirical wind power histograms and transformed generalized gamma wind speed distributions	51
1.6.3 Comparison of reanalysis-based wind power data with tower measurements	53
1.7 Wind power networks	59
1.7.1 Wind energy production in a fully connected static network	59
1.7.2 Energy production in limited area static networks	62
1.7.3 Output Variability in Growing Static Networks	65
1.7.4 Static network with different capacity limits	66
1.7.5 Static network based on real wind farm data	67

1.7.6	Network “disintegration” test for wind field persistence	71
1.7.7	Energy production with full dynamic control	73
1.7.8	Matching demand and wind power supply – a case study	75
1.8	Conclusions	84
2	Lightning - Cosmic Rays	87
2.1	Introduction	87
2.2	Overview of terrestrial lightning activity	88
2.2.1	General features of lightning activity	88
2.2.2	The global electrical circuit	89
2.2.3	The breakdown processes during lightning flashes . .	93
2.2.4	Description of one distinct lightning flash	94
2.3	Cosmic radiation – the missing link in climate?	96
2.3.1	General properties of cosmic radiation	96
2.3.2	Modulations of cosmic radiation flux	98
2.3.3	Climatic connections - Clouds	100
2.3.4	Connection with lightning activity	102
2.4	Datasets used	103
2.4.1	Lightning data	103
2.4.2	Cosmic radiation data	109
2.4.3	Cloud data	109
2.5	Construction of time series	110
2.5.1	Lightning statistics of Central Africa – an example .	110
2.5.2	Global time series of lightning activity	113
2.5.3	Cosmic radiation time series	116
2.6	Lightning activity during Forbush decreases	119
2.7	Lightning activity in a Solar Proton Event	120
2.8	Further attempts	123
2.8.1	Lightning activity in hurricanes	123
2.8.2	Analysis of reflectivity data	123
2.9	Summary	123
A	Maximum likelihood estimation	127
B	Acronyms	129

Chapter 1

Estimating the limitations of the European wind energy resource

1.1 Introduction

The growth of worldwide wind power capacity is almost exponential with 121.2 GW total installed nameplate capacity at the end of 2008 [1]. Europe has a pioneering role in this development with more than half of this capacity (64.9 GW at the end of 2008) [2]. European leaders signed up in March 2007 to a binding EU-wide objective to supply 20 % of their energy needs from renewable sources such as biomass, hydro, wind and solar power by 2020.

The twitching of oil prices throughout the year 2008 or the latest Eastern European gas crisis in January 2009 has boosted discussions again on the future role of renewable energy sources. Over the past decade, the governments of leading industrial countries have appreciably supported the development of both solar and wind energy.

In spite of these facts, renewable energy represents only 5-6 % of the total energy consumption among OECD (Organization for Economic Cooperation and Development) countries [3]. Besides economic and technical questions, there is a primary interest in estimating the physical potential and limitations of various resources.

Recent comprehensive studies on global wind energy potential have shown that 1–5 times the global energy usage (7–40 times the global electricity consumption)¹ could be supplied from wind energy on a yearly basis [4, 5],

¹The world total primary energy consumption was ~ 15.5 TW in 2005 (1.9 TW were electricity out of this), according to the U.S. Energy Information Administration.

provided that wind turbines were installed at every possible location (urban and forested areas were excluded).

These estimations are based on either direct meteorological observations (surface or sounding) or reanalysis datasets. An optimal global evaluation would require a dense measuring tower network and a waiting time of 5–10 years, but of course, this is not possible in practice. Landberg et al. listed eight alternative methods for wind resource estimation, starting from “folklore” to combined meso/micro scale modeling [6]. Global databases of winds (method 4 in Ref.[6]) have become available within the past decade as a result of huge reanalysis efforts by various institutions such as NCEP/NCAR (National Centers for Environmental Prediction/National Center for Atmospheric Research) [7, 8] or ECMWF (European Centre for Medium-Range Weather Forecasts) [9]. Measured surface wind data (at standard 10 m) over the ocean by buoys and satellites are assimilated in the reanalyzes [9], however validation projects have revealed some statistical deficiencies in the extrapolated fields over the Arctic Ocean [11], equatorial Indian Ocean [12], or over a complex terrain [13].

Wind power scepticism is mostly based on the fact that energy supply is often not available from any single cluster of turbines. Contrarily, proponents argue that “wind always blows somewhere”, thus energy generation integrated over a large area is less variable than generation from a single region. This statement is supported by several studies [14, 15, 16, 17], nevertheless it is also a fact that the medium range (hours to days) variability cannot be decreased by distributed generation as effectively as the short range (seconds to minutes) fluctuations [15]. The integration of electricity production over increasing areas has natural barriers. At present, wind turbine installation in EU countries is confined to the continent and an off-shore belt, lowered by urban areas, various land-use functions (e.g., bioreserves) and mountain regions of high elevation. Assessment studies over Europe [14, 17] analyzed the main aspects of large scale distributed wind power production. Giebel [17] modeled the whole electrical generation system of Europe with varying penetrations of wind power. The main conclusion was that wind energy can contribute more than 20 % of the European demand without significant changes in the system and can replace conventional sources worth about 10 % of the installed wind power capacity.

In order to achieve a substantial penetration of distributed energy resources, it is necessary to address the key issues related to their integration into existing and future energy systems. One of the most important challenges seems to be the management of fluctuations in the electricity produc-

tion from renewable energy sources [18]. In the power network supply has to match consumption exactly in order to prevent the destruction of grid components and major blackouts [19]. The incorporation of higher levels of stochastic generation requires the development of “horizontally operated” power systems [20], where dispatchable generation units adapt to the variations of the non-dispatchable ones. Traditional power systems based on thermal plants have limited wind power acceptance capacity constrained by an adequate level of system stability [21]. However, the development toward the integration of distributed generation is continuous, including a number of new technologies such as intelligent switches, adaptive protection, or web based information and control systems [22, 23].

In this work we intend to contribute to a better characterization of the constraints arising from the statistical properties of wind field over Europe. After a thorough examination of wind speed statistics we extrapolate surface wind fields to turbine level and use the latter to approximate wind power output. We construct model time series for wind power based on ERA-40 reanalysis [9] wind data covering 44 whole years and the geographic area of the continent and the surrounding maritime regions. We also use high frequency turbine data to link the reanalysis based estimates to real measurements, and also to construct a local model in which we try to supply a real consumer with wind farm generated electricity.

In all cases we consider “perfect” grid architectures, where all the generated electricity is integrated without any loss or technical restrictions. Our approach is certainly far from any technical actuality, however it helps to demonstrate the unavoidable limitations coming from the intermittent properties and spatio-temporal correlations of the wind field. We show that the large wind power fluctuations cannot be eliminated even by integrating over the whole continent.

In spite of the difficulties we strongly believe, that wind power will be a key energy resource in the future, as it is already very important now, however the large scale implementation is by far not as simple as that of coal or gas or hydro plants. The optimal exploitation of wind power cannot be realized without taking into account the physical constraints coming from the intermittency of wind fields.

1.2 The data used

1.2.1 Reanalysis data

We evaluated European Centre for Medium-Range Weather Forecasts' (ECMWF) **ERA-40** reanalysis data [9] consisting of the u (eastward) and v (northward) orthogonal components of the horizontal wind field at 10 m above ground level. The data base covers a time period of 44 whole years between 1 September 1958 and 31 August 2002. Four instantaneous values are recorded each day for the main synoptic hours of 00, 06, 12 and 18 UT at each location. The ERA-40 reanalysis is a dynamical model using the three dimensional variational technique. The model is fed with a great number of archive observational data. The spatially spectral model was gridded to yield a $1^\circ \times 1^\circ$ latitude by longitude resolution. A given value for an atmospheric variable is considered to be representative for the whole cell. Our analysis is restricted to a geographical area covering Europe: 2501 grid-points between 35° N and 75° N latitude and 20° W and 40° E longitude. The ERA-40 data are freely available <http://data.ecmwf.int/data/> at a coarser resolution of $2.5^\circ \times 2.5^\circ$ (latitude by longitude). ²

The average value \bar{s} and standard deviation σ_s of ERA-40 surface wind speeds $s = \sqrt{u^2 + v^2}$ computed over the whole period of 44 years (see Figure 1.1) illustrate the gross features of wind climatology over Europe. The strong coupling between the values of average speed and standard deviation is apparent, the coefficient of variation is around $\sigma_s/\bar{s} \approx 0.5$, except for a few isolated regions (for example around Corsica).

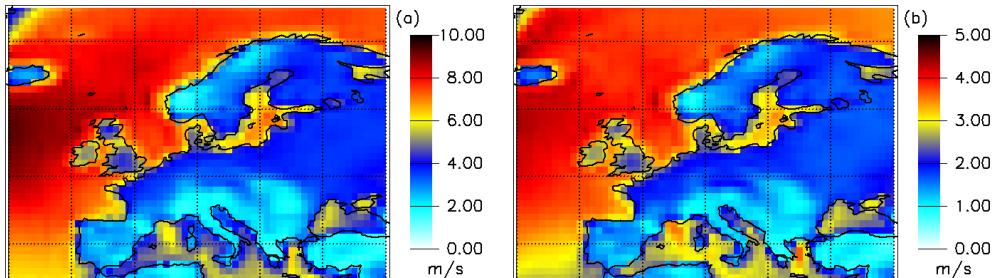


Figure 1.1: (a) Average value, and (b) standard deviation of ERA-40 surface wind speeds in the period 1958-2002, in units of m/s. Note that the color scales are different by a factor of 2.

In order to be able to extract wind profiles and estimate wind speeds

²Here I would like to thank the Hungarian Meteorological Office (OMSZ) to provide us the full resolution ERA-40 data.

higher above the ground, we used wind field data at another level as well. Namely we used the u (eastward) and v (northward) orthogonal components of the ERA-40 horizontal wind field at the 1000 hPa pressure level and the geopotential Φ of the 1000 hPa pressure level. The spatial and temporal coverage of these data is identical to the 10 m wind speed data. In the following calculations let z denote the elevation above mean sea level and h denote elevation above the ground. The geopotential Φ at a given latitude φ and elevation z is given by the following equation:

$$\Phi = \int_0^z g(\varphi, z') dz' , \quad (1.1)$$

where g is the local gravitational acceleration which depends on latitude φ and the vertical coordinate - elevation z' measured from mean sea level. The geopotential Φ is a measure of gravitational potential energy per unit mass at a given elevation measured from mean sea level.

It is common to define a “gravity adjusted” height coordinate, the so-called geopotential height z_g (above mean sea level) by the following expression:

$$z_g = \frac{\Phi}{g_0} , \quad (1.2)$$

where $g_0 = 9.80665$ is the standard gravitational acceleration at mean sea level. We used the geopotential height of the 1000 hPa pressure level which is on the average about $z_g \approx 100$ m. Gravitational acceleration changes very little on the scale of a few hundred meters around sea level, therefore the integral in Equation (1.1) simplifies to a multiplication by z . Furthermore if we neglect the latitude dependence of g , and approximate it with $g(\varphi) \approx g_0$ then the errors we get are less than 0.2 % (g_0 is measured at $\varphi = 45^\circ$ N). Thus we get that in the case of the 1000 hPa level and above Europe the geopotential height is a good approximation of geometrical height (elevation):

$$z_g \approx z . \quad (1.3)$$

The average and the standard deviation of the geopotential height z_g of the 1000 hPa pressure level can be seen in Figure 1.2.

The reanalysis provides continuous meteorological fields by extrapolations, thus the 1000 hPa geopotential height often exhibits negative (“underground”) values (see Figure 1.2). Clearly, the wind field at the 1000 hPa pressure level provides limited information, especially over high geographic elevations, where the pressure height $h = z_g - z_o$ (where z_o is the elevation of the ground, the orography) is almost always negative. Despite this fact the

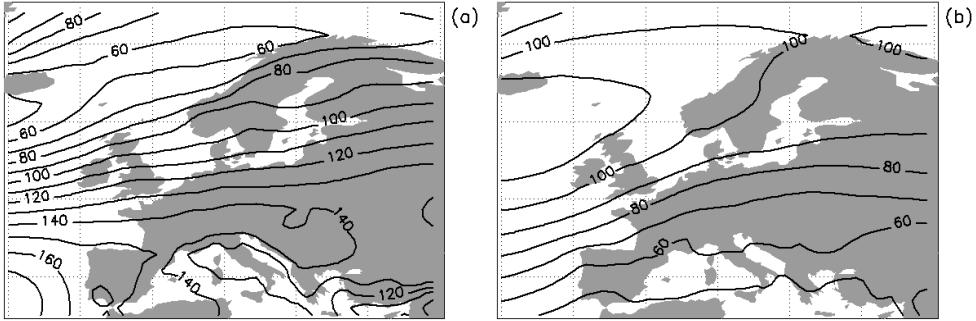


Figure 1.2: (a) Average value, and (b) standard deviation for the 1000 hPa geopotential height for the 44 years covered by the ERA-40 data bank. Contour lines are drawn with a spacing of 10 m, see the labels.

geopotential data are still very useful after a filtering procedure to calculate wind profiles.

Data of the European orography z_o were obtained from <http://earthsystematlas.sr.unh.edu>. Data with $0.5^\circ \times 0.5^\circ$ (latitude by longitude) resolution were averaged to give a $1^\circ \times 1^\circ$ (latitude by longitude) grid corresponding to the ERA-40 data. The topographical map constructed from the above mentioned elevation data can be seen in Figure 1.3.

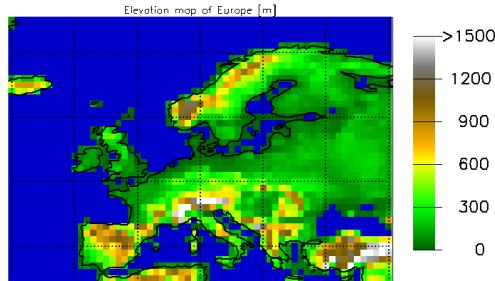


Figure 1.3: Topographical map of Europe based on the data obtained from <http://earthsystematlas.sr.unh.edu>. Elevation (z_o) is given in meters.

We used the 10 m wind velocity fields above Europe from the **ERA-Interim** reanalysis as well. There are some differences in data assimilation and use of observations between ERA-40 and ERA-Interim [10], furthermore the spatial resolution decreased to $1.5^\circ \times 1.5^\circ$ (latitude by longitude), however the 6 hour sampling remained the same as in ERA-40. We used the freely downloadable data between 01/01/2000 and 12/31/2005 retrieved from <http://data.ecmwf.int/data>.

1.2.2 Tower measurements and wind farm data

High frequency wind speed (nacelle anemometer readings) and output power data of two wind turbines at the geographic location 47.816° N, 17.174° E (near Mosonszolnok, Hungary) were available to us (see Figure 1.4). ³ 10 minute average values were recorded for two neighboring Enercon E-40 wind turbines in the period from 01/01/2004 to 12/31/2006. The proximity of the locations (the distance is 370 m) yielded very similar time series, significant differences originated from measurement errors, or hardware breakdowns indicated most often by zero power reading during high wind periods, or finite electric output at zero recorded wind speeds. These turbines will hereafter be referred to as $M1$ and $M2$.

Similar wind speed and power readings of another Enercon E-40 turbine at the location 47.057° N, 18.914° E, in Kulcs, Hungary were available to us. ⁴ The period of available measurements spans 01/01/2005 to 06/30/2008. The data set of the Kulcs turbine will be referred to as K .

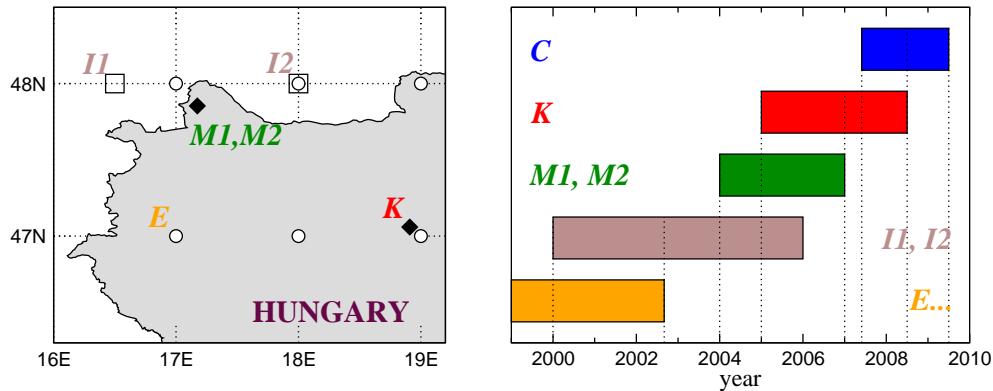


Figure 1.4: (a) The geographic map of North-Western Hungary with the positions of the turbines $M1$, $M2$ (Mosonszolnok) and K (Kulcs) indicated. The two ERA-Interim grid-points $I1$ and $I2$ are also indicated with open squares along with the ERA-40 grid-points (open circles). (b) The time period of the turbine measurements $M1$, $M2$ and K is shown along with the consumption data C , the ERA-40 data E and the ERA-Interim data $I1$ and $I2$.

We used the electricity consumption (referred to as C) data of a large consumer, a factory in Hungary that kindly provided us their high quality data. The total record length covers more than 2 years (see Figure 1.4) with a temporal resolution of 15 minutes without a sole missing point. The power

³Here I would like to thank László Varga for kindly providing us the wind turbine data.

⁴Here I would like to thank Balázs Stelczer for kindly providing us the wind turbine data.

consumption record for the calendar year 2008 is shown in Figure 1.5, the grey intervals are magnified in the middle and bottom panels. The pattern is composed of highly regular daily and weekly cycles interrupted by short national holidays and two longer periods of decreased activity in the Summer and around Christmas.

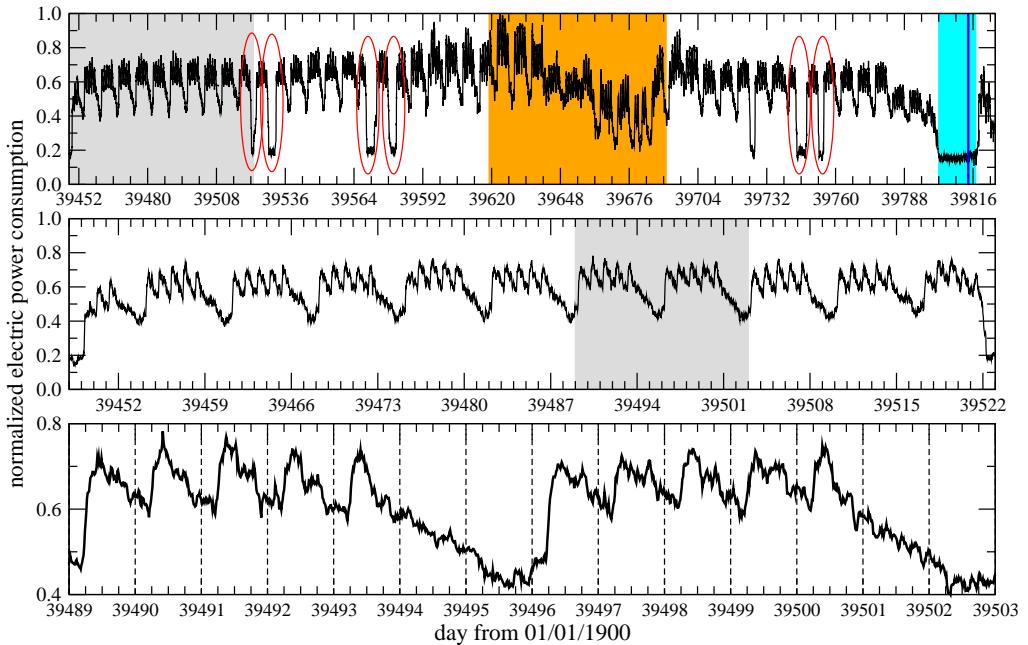


Figure 1.5: Time series C: relative electric power consumption (instantaneous value normalized by the measured maximum) as a function of time in the year of 2008 (top). Grey shading indicates the period magnified in the middle panel. The orange interval is the main holidays season (from mid-June to end August), light-blue section shows the Christmas-New Year break. The ellipses denote major national holidays (15th of March, Easter, 1st of May, Pentecost, 23th of October, and Hallowmas). The grey shading in the middle panel denotes the two weeks magnified at the bottom.

We used data on the distribution of European wind farms corresponding to the situation of October 2007. These data give cumulative information on the total operating nominal capacity of wind farms in each $1^\circ \times 1^\circ$ (latitude by longitude) ERA-40 grid cell.⁵

⁵Here I would like to thank Flavio Bono and Eugenio Gutierrez at the Joint Research Centre, Ispra, Italy for kindly providing us the data on the distribution of European wind turbine nameplate capacities.

1.3 Models for wind speed histograms

The atmospheric flows are strongly volatile therefore the average wind speed at a given location is a very poor predictor of the energy output of a wind turbine. The basic requirement for wind power estimates is an adequate characterization of the empirical probability distribution of wind speeds since wind direction is less important because of the well developed methods of yaw control for modern turbines [24]. The statistical description is highly simplified when a measured histogram can be accurately fitted by an analytical probability density function (PDF) with a few parameters.

The traditional approach of modeling wind speed PDF is based on the Rayleigh and the more flexible Weibull distributions [24, 25, 26, 27, 28, 29, 30]. However, several authors noted that Weibull fits of empirical data have low quality at several locations, mostly over land. Various analytical forms of skewed distributions were proposed as possible alternatives, such as the log-normal [31, 32], square-root-normal [33, 34], chi [31], inverse Gaussian [35], generalized gamma [31], generalized extreme value [36] or extended exponential functions [37, 38].

In this section we report on a detailed analysis of surface wind speed distribution over Europe [39]. The main goal was to find an effective and optimal description of the PDF both for onshore and offshore locations. Our tests unambiguously demonstrate that the generalized gamma (GG) distribution provides an improved fit for an overall statistical characterization of surface wind speed.

1.3.1 Rayleigh distribution

The most transparent model for scalar wind speed distribution is based on the assumptions that the orthogonal u and v components are independent and identically distributed (iid) Gaussian random variables with zero means and equal standard deviations of $s_0/\sqrt{2}$ (we adopt this notation to get simpler mathematical formulas below). Of course, all the higher moments (skewness, kurtosis, etc.) are identically zero. In this case $s = \sqrt{u^2 + v^2}$ obeys Rayleigh probability density distribution [40] of the functional form

$$P_R(s; s_0) = \frac{2}{s_0} \left(\frac{s}{s_0} \right) \exp \left[-\left(\frac{s}{s_0} \right)^2 \right], \quad (1.4)$$

where the only free parameter is s_0 (the so called scale parameter).

A trivial consequence of the basic assumptions behind a Rayleigh distribution is that the mean vector wind should be zero as well. However, it is

well known that the long range vectorial averages are significantly different from zero, in particular over the oceans (see e.g. [41, 42]). Actually, these nonzero values define the prevailing wind systems (e.g. trade winds). This is the first reason why the Rayleigh distribution has a limited applicability, especially for sea winds [30, 43].

The mean vector wind, however, is often close to zero over land [41, 42]. Therefore, a next plausible test on the validity of the basic assumptions is to determine the normalized third and fourth central moments, the skewness (Sk) and kurtosis (K) for the individual wind components u and v :

$$Sk(x) \approx \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n\sigma_x^3}, \quad K(x) \approx \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{n\sigma_x^4} - 3, \quad (1.5)$$

where x_i is either u_i or v_i , and n is the number of observations. Equations (1.5) are not exact equalities because they are biased estimators of skewness and kurtosis but in this case the sample size is very large ($n = 64240$) and Equations (1.5) can be considered to be very good approximations. The results are shown in Figure 1.6. The maps clearly illustrate that very few geographical locations exhibit pure Gaussian probability distribution ($Sk = 0, K = 0$) for the individual wind vector components.

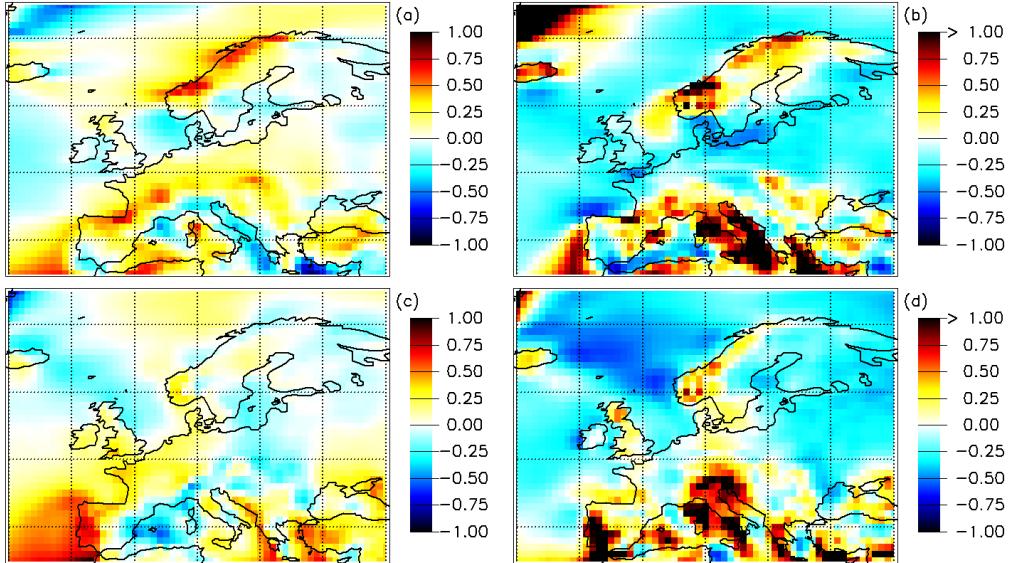


Figure 1.6: Geographical distribution of empirical skewness Sk and kurtosis K (see Eq. (1.5)) for the wind components u and v . (a) $Sk(u)$, (b) $K(u)$, (c) $Sk(v)$, and (d) $K(v)$. White color emphasizes $Sk = 0$ and $K = 0$ levels.

The standard method to test interdependence of the components u and

v is based on computing the correlation coefficient r_{uv} defined as

$$r_{uv} = \frac{\sum_{i=1}^n (u_i - \bar{u})(v_i - \bar{v})}{(n-1)\sigma_u\sigma_v}, \quad (1.6)$$

where over-line indicates average value, and σ_u and σ_v are corresponding standard deviations, as before. Figure 1.7 shows that the assumption of independence fails in general, strong correlations of magnitude 0.6 – 0.8 are present at several geographical locations. (Note that $r_{uv} \approx 0$ does not necessarily mean statistical independence.)

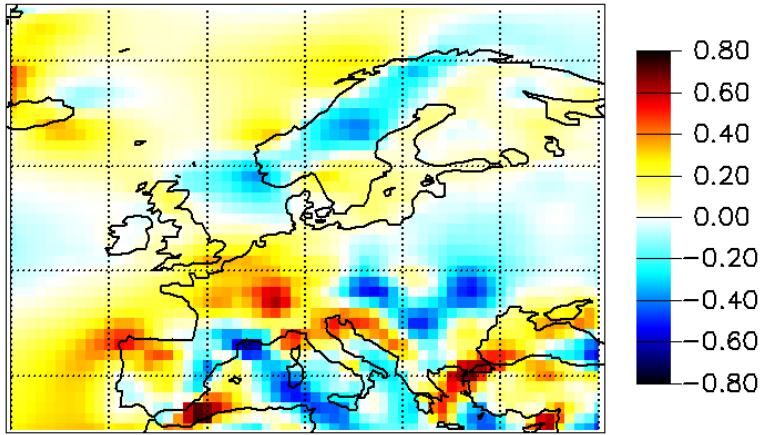


Figure 1.7: Correlation coefficient r_{uv} (see Eq. 1.6) for the u and v wind velocity components. White color emphasizes $r_{uv} = 0$ level.

Nonzero correlations are usually taken into account by considering joint probability distributions. For example, when u and v are assumed to be Gaussian random variables with mean values \bar{u} and \bar{v} , standard deviations σ_u and σ_v , and correlation coefficient r_{uv} , then the joint PDF can be written as

$$P(U, V) = \frac{1}{2\pi\sigma_u\sigma_v\sqrt{1-r_{uv}^2}} \exp\left(-\frac{U^2 - 2r_{uv}UV + V^2}{2(1-r_{uv}^2)}\right), \quad (1.7)$$

where $U = (u - \bar{u})/\sigma_u$ and $V = (v - \bar{v})/\sigma_v$ denote standardized variables.

There are two plausible methods to obtain standardized velocity components U and V . First, the mean values \bar{u} and \bar{v} can be computed over the whole length of the time series assuming a well-defined prevailing wind. The second way is to consider velocity fluctuations around the annual periodic background signal, which is determined by the average value for a given synoptic hour and given calendar day. (For further discussion see Section 1.4). We tested both cases.

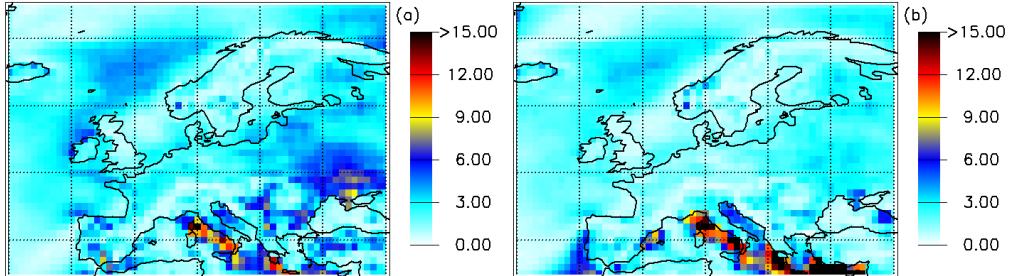


Figure 1.8: Unexplained percentage variance $100(1 - R^2)$ of joint Gaussian PDF fits by Equation (1.7). (a) Full time average values removed. (b) Daily and annual periodic cycle removed.

Figure 1.8 illustrates the quality of bivariate Gaussian model fits for the two methods of standardization. The relative low values of the unexplained percentage variance $100(1 - R^2)$ calculated between the empirical and the model PDF, indicate where the model performs well. We note that most of the deviations between the model and data are due to a higher empirical probability of large wind speeds with respect to a Gaussian distribution. At several locations exponential tails $P(s) \propto e^{-s}$ for large values of s provide a better fit than the Gaussian decay $P(s) \propto e^{-s^2}$. Higher values of kurtosis in Figure 1.6 refer to this feature as well.

In view of the above results, we should not expect that the Rayleigh or Gaussian models provide an adequate universal description for the empirical wind speed distributions.

1.3.2 Weibull distribution

The most widely accepted model for wind speed probabilities is the two-parameter Weibull distribution [44]:

$$P_W(s; s_0, k) = \frac{k}{s_0} \left(\frac{s}{s_0} \right)^{k-1} \exp \left[-\left(\frac{s}{s_0} \right)^k \right], \quad (1.8)$$

where s_0 and k denote the scale and the shape parameters, respectively. It is easy to see that the Rayleigh distribution (Equation (1.4)) is a special case of Weibull with $k \equiv 2$, while $k \equiv 1$ gives the simple exponential distribution. The Weibull PDF (Equation (1.8)) can be derived theoretically as a form of extreme value distributions [45], and it is the most popular model for failure rate distributions [40, 44]. In the context of wind speed histograms, we can rather consider the functional form Eq. (1.8) as a generalization of the Rayleigh distribution, which provides an increased flexibility to fit empirical

data. The properties of the Weibull distribution have been thoroughly studied and numerous studies show that it works well for modeling wind speeds at several locations [24, 25, 26, 27, 28, 29, 30, 33, 43, 46, 47, 48, 49, 50, 51, 52].

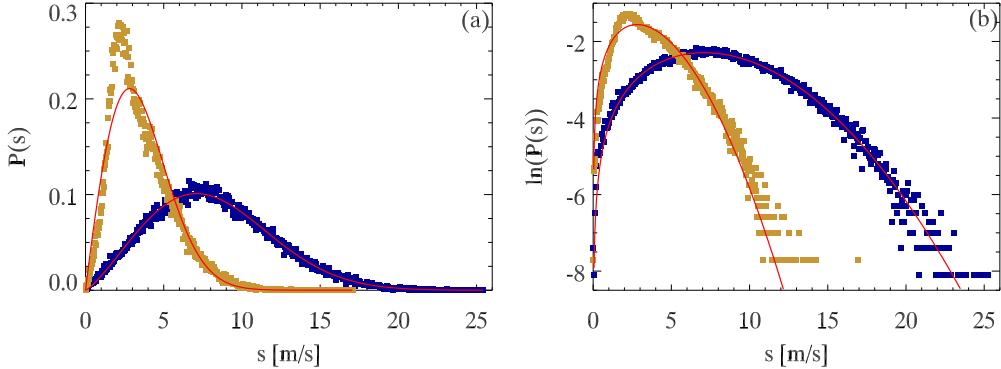


Figure 1.9: Normalized histogram of wind speed s at two distinct locations: Atlantic Ocean ($66^\circ N$ $4^\circ W$, dark blue), and northern Germany ($52^\circ N$ $11^\circ E$, light brown). Maximum likelihood Weibull fits are shown with straight lines. (a) Linear scale, and (b) semi-log scale.

Examples of Weibull fits are shown in Fig. 1.9, where the geographical locations were chosen to illustrate particularly good and poor fits. In general, the Weibull PDF is a reasonably good model over the ocean and seas, however histograms for large areas over land can be fitted with rather large errors, as it has been noted by other authors as well [53]. Fig. 1.9 illustrates that neither the central part nor the large speed tail of the empirical histograms are close enough to a Weibull distribution. The latter is more problematic in practice, because wind energy is only produced above the cut-in speed of wind turbines (typical values are 3-5 m/s), therefore an adequate fit of the large speed tail is crucial for wind energy estimates.

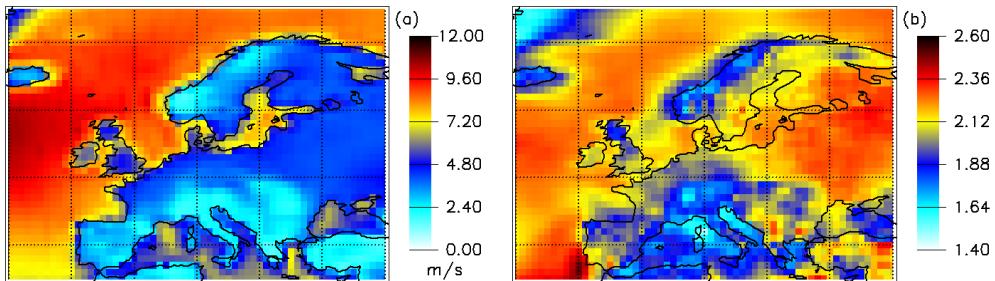


Figure 1.10: Geographical distributions of the (a) scale s_0 and (b) shape k parameters of maximum likelihood Weibull fits (see Eq. (1.8)).

We estimated Weibull parameters using the maximum likelihood method [54], which lead to a transcendent equation [55], that we solved iteratively (see Appendix A). The geographical distribution of fitted Weibull parameters is shown in Fig. 1.10. The scale parameter s_0 shows almost the same pattern as the mean wind speed \bar{s} in Fig. 1.1a. This is plausible, because the mean value of a Weibull distribution is given by $s_0\Gamma(1+1/k)$, where the correction factor of the Gamma function changes in the interval $\Gamma \in [0.888, 0.911]$ for the fitted shape parameter regime $k \in [1.4, 2.6]$. The shape parameter k exhibits a more interesting spatial pattern in Fig. 1.10b. Lower numerical values indicate a slower decay of the large speed tail, and such shape is characteristic for areas of very low average speed \bar{s} or scale parameter s_0 , however the relationship is not entirely strict.

A straightforward generalization of the two-parameter Weibull PDF is the three-parameter Weibull distribution

$$P_{W3}(s; \mu, s_0, k) = \frac{k}{s_0} \left(\frac{s - \mu}{s_0} \right)^{k-1} \exp \left[- \left(\frac{s - \mu}{s_0} \right)^k \right] , \quad (1.9)$$

where the third (location) parameter μ shifts the Weibull peak horizontally. This helps to model wind speed histograms at calm local climate, where $s = 0$ has a significant positive probability. However, such places are exceptional, therefore the general improvement of fits by Eq. (1.9) is marginal.

Since the Weibull distribution does not provide a universally good description of wind speeds everywhere, we tested other empirical distributions as well. For example, the log-normal distribution ($\ln(s)$ exhibits Gaussian PDF) was extensively used to fit wind speed histograms over land [32, 50]. Our results show that the log-normal distribution is an inappropriate model over most of the areas examined. In general, a log-normal distribution decays much slower towards high wind speeds than the empirical data. Nevertheless, at a few grid-points (e.g., northern Germany) we found that the log-normal model fits better than the Weibull, see Fig. 1.11. A typical histogram at these grid-points is characterized by a sharp peak and an exponential-like right tail, which is closer to the log-normal behavior (Fig. 1.11).

1.3.3 Generalized gamma distribution

In order to overcome the constraint that both tails of a Weibull peak are described by a single shape parameter k , we can further generalize Eq. (1.8) to get the generalized gamma (GG) distribution also suggested by Auwera

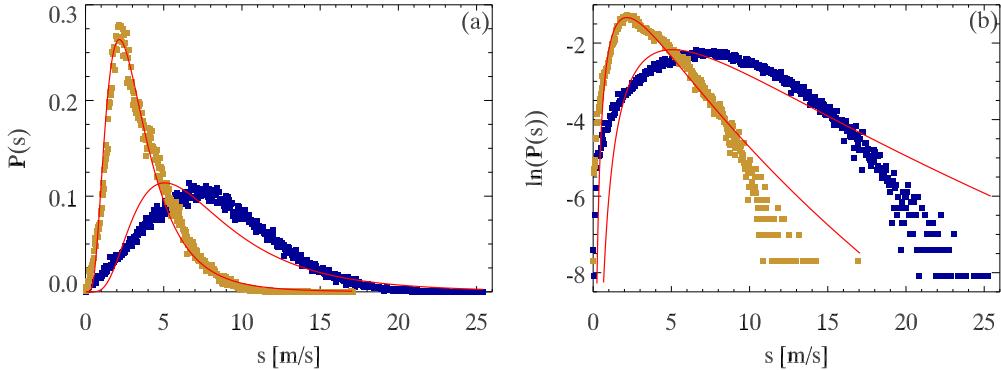


Figure 1.11: Log-normal fits for the same normalized histograms as shown in Fig. 1.9. (a) Linear scale, and (b) semi-log scale.

et. al. [31]:

$$P_{GG}(s; s_0, k, \epsilon) = \frac{k}{s_0 \Gamma(\epsilon)} \left(\frac{s}{s_0} \right)^{\epsilon k - 1} \exp \left[-\left(\frac{s}{s_0} \right)^k \right] , \quad (1.10)$$

where the Gamma correction $\Gamma(\epsilon)$ is required for normalization, and the new parameter ϵ improves the shape flexibility. Special cases are $\epsilon \equiv 1$ (the original Weibull distribution), and $k \equiv 1$ which gives the gamma distribution. The log-normal distribution can also be obtained as a limiting distribution when $\epsilon \rightarrow \infty$ [56]. The fixed parameter value $k \equiv 2$ defines a sub-family of GG which is known as the generalized normal (GN) distribution. The GN is itself a flexible family and includes the Half-normal ($\epsilon = 1/2$), Rayleigh ($\epsilon = 1$), Maxwell-Boltzmann ($\epsilon = 3/2$), and chi ($\epsilon = n/2$; $n = 1, 2, 3, \dots$) distributions.

The function P_{GG} (Eq. (1.10)) has a single maximum (mode m) at

$$m = s_0 \left(\epsilon - \frac{1}{k} \right)^{\frac{1}{k}} . \quad (1.11)$$

The effects of the two shape parameters k and ϵ cannot be fully separated, nevertheless the asymptotic behavior on the left side is a power law with exponent $q = \epsilon k - 1$, while it is a stretched exponential on the right:

$$\begin{aligned} P_{GG}(s) &\propto s^q & 0 \leq s \ll m , \\ P_{GG}(s) &\propto e^{-(s/s_0)^k} & s \gg m . \end{aligned} \quad (1.12)$$

The α -th non-central moment of the GG distribution can be obtained easily:

$$\langle s^\alpha \rangle = s_0^\alpha \frac{\Gamma(\epsilon + \alpha/k)}{\Gamma(\epsilon)} . \quad (1.13)$$

An important property of GG distributions is that the family is closed under power transformation. That is, if $s > 0$ obeys $P_{GG}(s; s_0, k, \epsilon)$, then the corresponding PDF for $z = s^p$ has the form $P_{GG}(z; s_0^p, k/p, \epsilon)$ [56]. This property can be immediately exploited for estimating potential wind power, since it is proportional to the cube of the wind speed s^3 [24, 26].

GG parameters were obtained for the empirical data by the maximum likelihood method. We derived the corresponding set of nonlinear equations, which were solved by the Newton-Raphson algorithm [54], initial guess for the parameters was provided by Weibull fits. For more details the reader is referred to Appendix A.

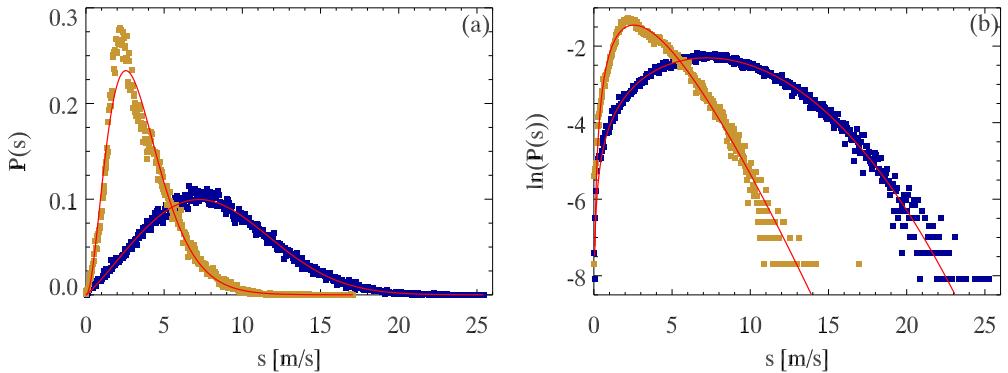


Figure 1.12: Maximum likelihood GG fits (see Eq. (1.10)) for the same normalized histograms as shown in Figs. 1.9 and 1.11. (a) Linear scale, and (b) semi-log scale.

Our results show that the GG PDF (Eq. (1.10)) provides a significantly improved fit compared to the Weibull model (Eq. (1.8)). The improvement is especially spectacular at the right tail (high wind speeds), which is correctly fitted in each case (Fig. 1.12). Nevertheless, even the generalized gamma distribution cannot capture all the features of the measured histograms over a few geographical areas. Such a problematic region is the northern coast of the Black sea, especially around the Crimean peninsula. Further details and a possible climatological explanation are given in Section 1.4.

The values of the fitted parameters at different locations can be seen in Fig. 1.13. The peak maximum m given by Eq. (1.11) exhibits practically the same pattern as the average wind speed \bar{s} (Fig. 1.1a) or the Weibull scale parameter s_0 (Fig. 1.10a), therefore we do not show the map again. The left tail shape parameter $q = \epsilon k - 1$ (Fig. 1.13a) is approximately 1 over the seas which implies a close to linear increase of probabilities for low wind speeds. The corresponding characteristic values are definitely larger over land, typically around $q \approx 2$. The geographical pattern for the right

tail shape parameter k (Fig. 1.13b) has the opposite tendency: typical values around 2 are characteristic over most of the offshore areas, whereas smaller values closer to 1 over land indicate a slower decay of large wind speed probabilities.

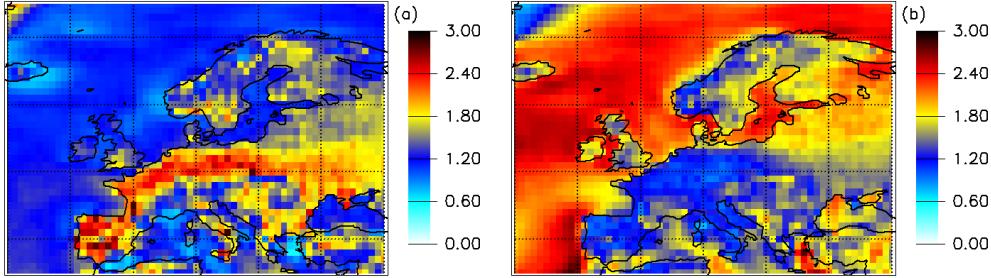


Figure 1.13: Geographical distribution of fitted GG parameters (see Eq. (1.10)). (a) Left tail parameter $q = \epsilon k - 1$, and (b) right tail shape parameter k .

A possible way of confronting the goodness of fit for Weibull and generalized gamma distributions is to compare the spatial patterns of an error parameter, such as the residual (or unexplained) percentage variance $100(1 - R^2)$ between the empirical and model PDFs. The results shown in Fig. 1.14 convincingly demonstrate that the generalized gamma distribution provides a better tool to fit wind speed histograms over various surface conditions. Note that the range of color scale in Fig. 1.14 is identical to the scale in Fig. 1.8, which illustrates that both the Weibull and GG models are superior compared to any Gaussian description that assumes linear correlation between u and v wind velocity components.

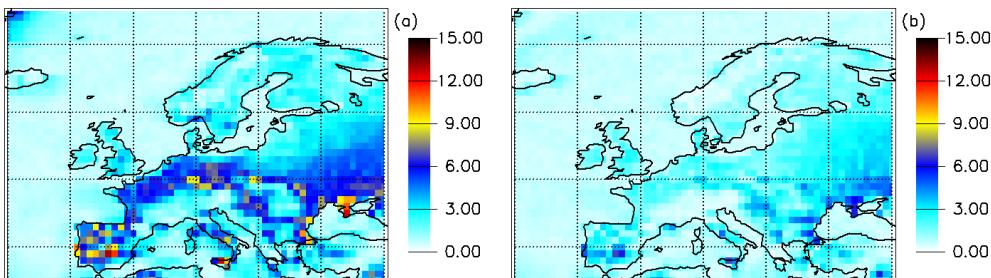


Figure 1.14: Unexplained percentage variance $100(1 - R^2)$ for (a) Weibull (Eq. (1.8)), and (b) GG (Eq. (1.10)) fits.

The maps of GG parameter distribution in Fig. 1.13 showing large coherent geographical areas suggest that the different shapes for wind speed histogram have a physical origin. Further insight can be gained by inspecting

the correlation plots of the fitted GG parameters $q = \epsilon k - 1$, k and m , the results are shown in Fig. 1.15.

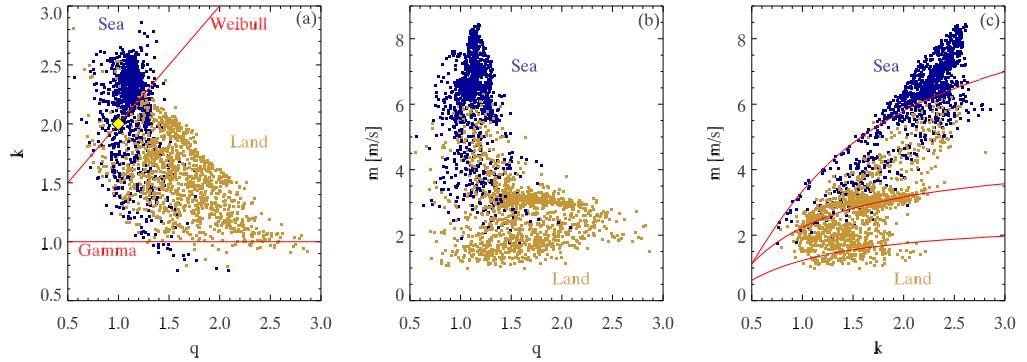


Figure 1.15: Correlation plots for the fitted parameters of GG distribution (see Eq. (1.10)): (a) $q = \epsilon k - 1$ vs. k , straight lines denote the special cases of gamma and Weibull distributions, and a diamond symbol denotes the Rayleigh distribution; b) q vs. m ; and (c) k vs. m . For an explanation of the continuous lines in (c) see the text. Land and sea surface areas are distinguished by different colors.

The first inference to be drawn in Fig. 1.15a is that only a minority of histograms coincide with the constrained distributions of Weibull ($k = q+1$), gamma ($k = 1$) or Rayleigh ($q = 1$, $k = 2$) forms, the rest has significantly different shape parameters for the left and right tails. Furthermore, the separation of clusters in parameter space corresponding to histograms over sea surface and over land is remarkable. The overlap between the shape parameters is characteristic for relatively small islands (e.g. Ireland) and restricted maritime regions (e.g. Adriatic sea). Fig. 1.15b illustrates the interdependence between the left tail shape parameter q and mode m , the separation for sites over sea and land is clear again. From the point of view of wind power generation, the correlation plot in Fig. 1.15c is the most interesting. This illustrates the relationship between the right tail shape parameter k and the mode m (essentially the mean wind speed). The optimal parameter regime is located somewhere in the upper right quarter of the diagram Fig. 1.15c, where the mean wind speed is well over the cut-in value of 5 m/s, and the decay of probabilities of very large speeds is fast. The latter property is beneficial because of the smaller frequency of high loads on the wind turbine's tower.

For a further analysis of Fig. 1.15c, let us suppose that the probability of wind speeds obeys GG distribution (Eq. (1.10)) with a constant second (non-central) moment $\langle s^2 \rangle = C$ characterizing large coherent geographical regions. An invariant second moment means that the average momentum flow rate is

conserved. (The instantaneous mass flow rate is given by $\dot{m} = \varrho A s_i$, where ϱ is the density, A is the cross-sectional area, and s_i is the instantaneous flow speed. The momentum flow rate is simply $\dot{m}s_i$, its average value over a given time interval is clearly related to $\langle s^2 \rangle$.) In order to check the behavior of the second moment, we determined its geographical distribution (Fig. 1.16a) and the corresponding histograms separately over land and seas (Fig. 1.16b). The histograms in Fig. 1.16b reveal a characteristic second moment $\langle s^2 \rangle \approx 67 \text{ m}^2/\text{s}^2$ for maritime locations (with much lower values around the coastlines), and a bimodal distribution for land with two local maxima at $\sim 5 \text{ m}^2/\text{s}^2$ for high altitudes and $\sim 15 \text{ m}^2/\text{s}^2$ for the rest of the continent.

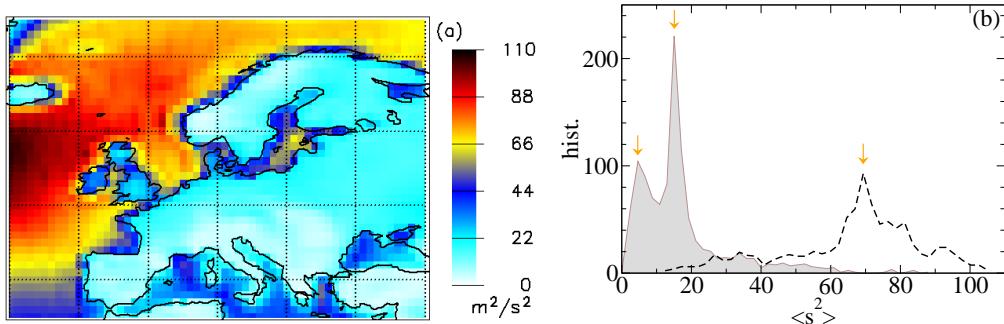


Figure 1.16: The second non-central moment $\langle s^2 \rangle$ calculated from Eq. (1.13) using GG fitted parameters. (a) Geographical distribution. (b) Histograms (not normalized) for maritime (dashed line) and terrestrial (gray shading) locations. Characteristic values $\langle s^2 \rangle$ of the peaks are indicated with arrows (see text).

Next we return to Fig. 1.15b. As it is already noticed, the left tail shape parameter $q = \epsilon k - 1$ is close to 1 for most maritime locations, irrespective of the other parameter values. Much larger variations are observed over land, nevertheless we can fix a typical value somewhere around $q = 2$. This permits a consistency check of the results: when the second moment $\langle s^2 \rangle$ (see Fig. 1.16b) and shape parameter q are fixed, Eqs. (1.11) and (1.13) provide a relationship to express the mode m as a function of k . The thin curves in Fig. 1.15c illustrate the expected behavior with the fixed parameter values for sea and land, the agreement is quite satisfactory. Note that this procedure does not favor the second moment of $\alpha = 2$ in Eq. (1.13), practically any of them could theoretically be used. However, we found that the histogram of the second moments (Fig. 1.16b) exhibits a good separation of peaks, furthermore $\langle s^2 \rangle$ can be associated with the momentum flow rate, therefore we chose the second moment.

1.4 Temporal behavior of wind speed records

The probability density distribution of wind speed represents an aggregated statistical information, however it contains no information about possible temporal patterns which are equally important for practical purposes. The standard method to reveal such patterns is the Fourier spectral analysis. Since the ERA-40 records contain no missing intervals, we could exploit the speed of the FFT procedure [54] in our analysis.

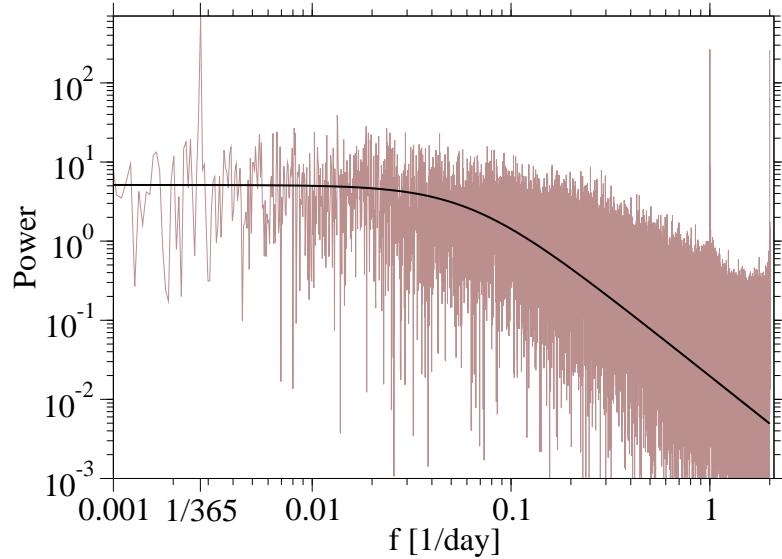


Figure 1.17: Normalized power density spectrum of the wind speed record over northern Germany ($52^\circ N, 11^\circ E$). The solid line shows a Lorentz fit [Eq. (1.14)].

The power spectra of wind records exhibit a fairly simple structure, an example is shown in Fig. 1.17. Only two characteristic cycles could be extracted from all of the data: annual and daily periods. The wide continuum background can be fitted by a Lorentzian spectrum (solid line in Fig. 1.17)

$$S(f) = \frac{2\tau}{1 + (2\pi f)^2\tau^2} , \quad (1.14)$$

where τ is a characteristic time related to an exponential decay of the auto-correlation function as

$$A(\Delta t) = \langle s(t)s(t + \Delta t) \rangle \propto \exp\left(-\frac{\Delta t}{\tau}\right) . \quad (1.15)$$

Typical values for the exponential decay time τ are between 1.5 and 4.0 days.

As for the identified periodicities, the daily cycle is missing from most of the records, even the annual course cannot be resolved at several places. In order to visualize the strength of the periodicities, we determined the partial power of the periodic components by integrating the area under the peaks shown in Fig. 1.17 for normalized power spectra. The geographical distributions are illustrated in Fig. 1.18. Numerical values close to zero mean that the given peak fades into the background. However, the maps in Fig. 1.18a should be considered with care, because the low sampling frequency and the low spatial resolution can obviously hinder to resolve the daily cycle at some places.

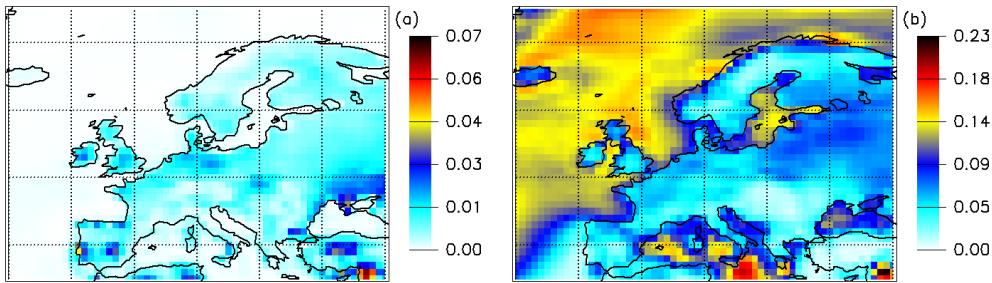


Figure 1.18: Partial power for the (a) daily and (b) annual periodicities extracted from normalized FFT power spectra of wind speed records. Note the different color scales. (By definition, the total integrated power has unit value.)

Figure 1.19 illustrates typical annual cycles computed over the 44 years. Remarkable features are that the yearly cycle is weaker over the onshore site (as it is seen in Figure 1.18 as well), and that it is loaded with considerable day-to-day fluctuations at both sites despite the fact that 44-year averages are shown. It is also remarkable that winters are much windier than summers.

In view of the above results, we can consider the time series of wind speeds as a superposition of three signals related to different physical mechanisms (the sub-hour timescale turbulence is not resolved by the data). The slowest component is an annual periodic background determined by the changing global insolation. Far the strongest component is determined by synoptic scale meteorological features (extra-tropical cyclones and anticyclones) and loaded with short range turbulent fluctuations excited by surface roughness and thermal convection. The third component representing the daily periodicities is characteristic to coastal locations, where the well known land-sea temperature contrast often induces daily periodic winds.

This picture helps to explain the apparent breaks in many histograms, clearly visible also in the near-coast record in Figs. 1.9, 1.11 and 1.12 at $s = 4.5 \text{ m/s}$. Such histograms can be perfectly decomposed into a superposi-

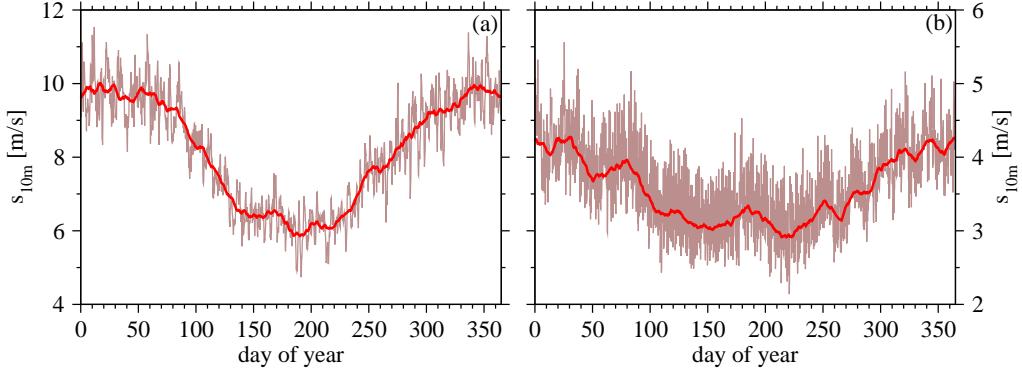


Figure 1.19: Annual wind speed cycles (a) over the Atlantic Ocean (66°N , 4°W) and (b) over northern Germany (52°N , 11°E). Brown lines represent the 44 year average for a given synoptic hour and calendar day and a thick red line shows the 15-day moving average.

tion of a narrow Gaussian (daily coastal wind), and a wide GG distribution representing the large scale flow. Since this decomposition includes 6 free parameters to be fitted, its usefulness is limited in the practice.

Nevertheless it is worth to check the effects of removing the background periodicities, thus considering wind speed fluctuations s' around various trends. The first procedure we implemented removes constant average values \bar{u} and \bar{v} separately from the velocity components, and produces the fluctuation series

$$s'_c(t) = \sqrt{[u(t) - \bar{u}]^2 + [v(t) - \bar{v}]^2} . \quad (1.16)$$

Note that the results shown in Fig. 1.8a were obtained by the same method (Eq. (1.16)). Next, the daily cycle can be removed by computing the average values for each synoptic hour separately as $\bar{u}(h) = 44^{-1} \cdot 365^{-1} \sum_{y,d} u(y, d, h)$ and $\bar{v}(h)$ similarly, and obtaining

$$s'_d(y, d, h) = \sqrt{[u(y, d, h) - \bar{u}(h)]^2 + [v(y, d, h) - \bar{v}(h)]^2} , \quad (1.17)$$

where the time variable t is replaced by the full calendar indices $y = 1 \dots 44$ (year), $d = 1 \dots 365$ (day), and $h = 1 \dots 4$ (hour). Finally, the daily and annual periodicity can be removed by obtaining average values for the velocity components as $\bar{u}(d, h) = 44^{-1} \sum_y u(y, d, h)$ and $\bar{v}(d, h)$ similarly for a given day and hour in a year. The fluctuations around the daily and annual cycles are given as

$$s'_a(y, d, h) = \sqrt{[u(y, d, h) - \bar{u}(d, h)]^2 + [v(y, d, h) - \bar{v}(d, h)]^2} . \quad (1.18)$$

Note that the results shown in Fig. 1.8b were obtained by this method (Eq. (1.18)).

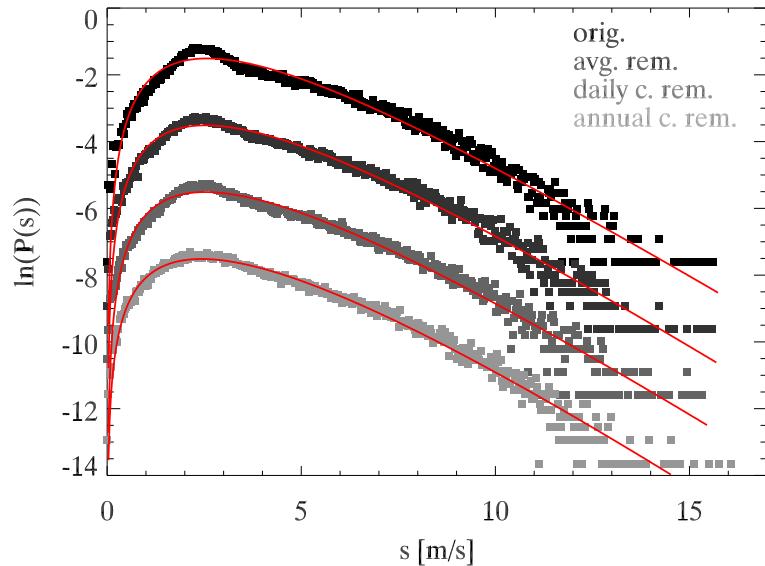


Figure 1.20: Semi-log scale wind speed histograms at the Crimean peninsula ($45^{\circ}N$, $34^{\circ}E$). From top to bottom: unfiltered data, constant removed (Eq. (1.16)), daily cycle removed (Eq. (1.17)), and daily and annual cycle removed (Eq. (1.18)). Continuous lines represent generalized gamma fits. The curves are vertically shifted by a value of 2 with respect to each other. The correct scale corresponds to the uppermost curve.

The effects of gradually removing a constant, daily, and daily and annual cycles from the original records is demonstrated in Fig. 1.20. It is obvious that wind speed fluctuations are much closer to a single generalized gamma distribution than the original wind speed s . This indicates that the main component of wind determined by the synoptic scale atmospheric patterns obeys a GG probability density distribution, indeed. The improvement works almost globally, as illustrated in Fig. 1.21.

The spatial distribution of fitted parameters and the correlation plots between them are very similar to the results obtained without the removal of averages (see Figs. 1.13 and 1.15), therefore the corresponding figures for wind speed fluctuations are not shown.

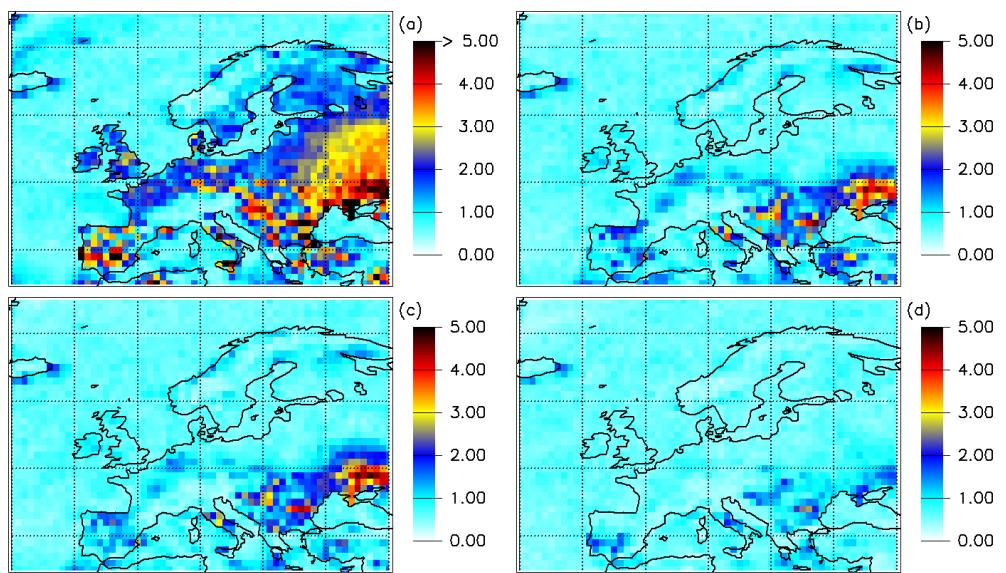


Figure 1.21: Residual percentage variance $100(1 - R^2)$ of generalized gamma fits. (a) Unfiltered wind speed s , (b) constant removed s'_c (Eq. (1.16)), (c) daily cycle removed s'_d (Eq. (1.17)), and (d) daily and annual cycle removed s'_a (Eq. (1.18)). Note that the color scale is much narrower than in Fig. 1.14.

1.5 Wind profiles - height dependence of wind speed

Wind speed (and direction as well) change with elevation above the ground. In the absence of turbine-level (60–120 m) observations or model data, it is common practice to extrapolate surface wind data upward [4, 57, 58, 59, 60] or free-atmosphere fields downward [6, 61, 62].

1.5.1 Wind data at different levels - a consistency check

Since wind components u and v at 1000 hPa pressure level and the geopotential Φ of the 1000 hPa pressure level were available to us from the ERA-40 data base, it was possible to perform a consistency check between the 10 m and the 1000 hPa wind speed data. The elevation of the 1000 hPa pressure level changes with the meteorological conditions but it can be approximated using the geopotential data set by calculating the geopotential height z_g using Equation (1.3) and Equation (1.19). In order to be able to compare 1000 hPa data with 10 m data we need to calculate the elevation of the 1000 hPa level above the ground which is given by the following equation.

$$h = z - z_o \approx z_g - z_o = \frac{\Phi}{g_0} - z_o , \quad (1.19)$$

where z is the elevation of the 1000 hPa level above mean sea level, z_g is the geopotential height of the same level, g_0 is the standard gravitational acceleration and z_o is the orography, the elevation of the ground. For the 10 m data evidently $h = 10$ m.

In the consistency check we used those data which fulfilled the condition $8 \text{ m} < h < 12 \text{ m}$ at the 1000 hPa level. Obviously there are some geographical locations (at higher elevations) where the ground level pressure is always smaller than 1000 hPa so this type of comparison is impossible. At each possible geographical location we examined the statistics of $s_{1000\text{hPa}} - s_{10m}$. The results are displayed in Figure 1.22. It can be seen that the two wind speed data do not match exactly. 1000 hPa wind speeds are about 0.6 m/s higher above the seas and about 0.75 m/s lower above terrestrial locations than 10 m wind speeds (see Figure 1.22a). These differences are not constant but have a standard deviation of about 0.4 m/s above the seas and about 0.8 m/s above land (see Figure 1.22b). The average of the absolute differences between 1000 hPa and 10 m wind speeds ($\langle |s_{1000\text{hPa}} - s_{10m}| \rangle$) can be seen in Figure 1.22c. In agreement with the calculations above, the mean absolute differences are about 0.7 m/s. Though at many locations the mean absolute

differences are about 1.0 m/s and at exceptional sites above 2.0 m/s. The number of data used in the previous calculations can be seen in Figure 1.22d.

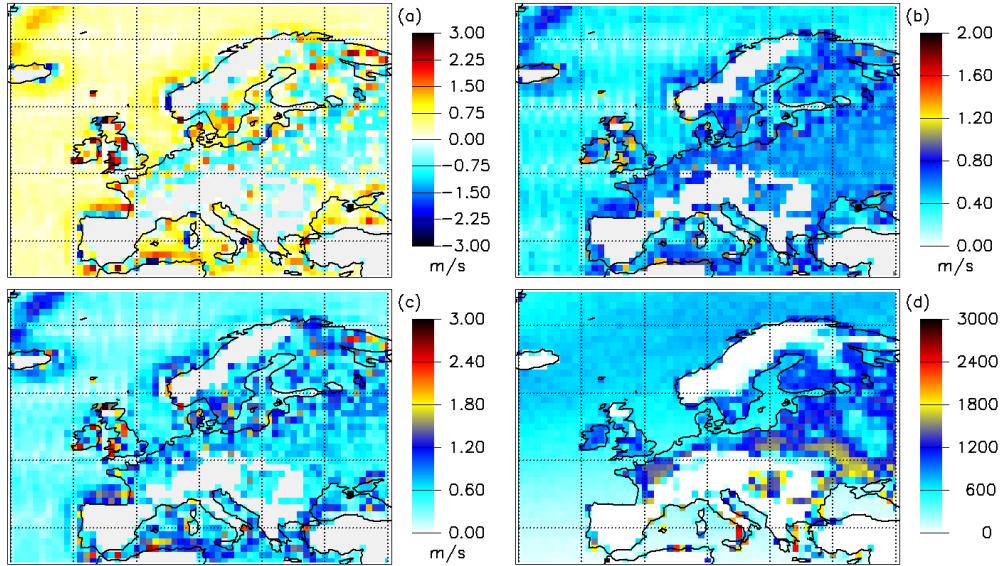


Figure 1.22: Consistency check between 1000 hPa and 10m wind speed data. Only those data were used which met the condition $8 \text{ m} < h < 12 \text{ m}$ at the 1000 hPa level. (a) Average of $(s_{1000\text{hPa}} - s_{10\text{m}})$, (b) standard deviation of $(s_{1000\text{hPa}} - s_{10\text{m}})$, (c) average of $|s_{1000\text{hPa}} - s_{10\text{m}}|$, and (d) the number of data used.

In general we can say that the 10 m and the 1000 hPa data are not inconsistent with each other but they do not match perfectly. This can introduce considerable errors in calculating the height dependency of wind speed or, as a next step, calculating wind speed at different elevations.

1.5.2 Direct comparison of 10 m - 1000 hPa data

Though the height difference between the two levels of available wind speed data changes as the 1000 hPa pressure level varies in elevation, a direct comparison might still be reasonable. Scatter plots between 10 m and 1000 hPa wind components and wind speed can be seen in Figures 1.23 and 1.24. Sites with good correlation are plotted on the former figure and sites showing poor correlation can be seen on the latter one. If the height separation were constant between the two levels, a perfectly linear relationship could be expected between wind speeds assuming a permanent height dependence (under constant stability conditions of the boundary layer). The fact that the height level separation depends on the meteorological conditions and we can still observe good linear correlations (the correlation coefficient $r \approx 0.98$ in

1.5. WIND PROFILES - HEIGHT DEPENDENCE OF WIND SPEED 33

some cases) between wind speeds at different levels might be surprising. Such strong correlations can be explained by an extremely weak height dependence of wind speeds.

Sites with strong correlation between 10 m and 1000 hPa wind speeds (weak height dependence) are mainly offshore locations, whereas sites with poor correlation are mainly located along coastlines or in regions of higher mountains indicating the important role of surface variability (see Figure 1.25). Besides orographic effects, weak correlations can be caused by a stronger height dependence of wind speed and thus higher sensibility to changes in elevation of the 1000 hPa level.

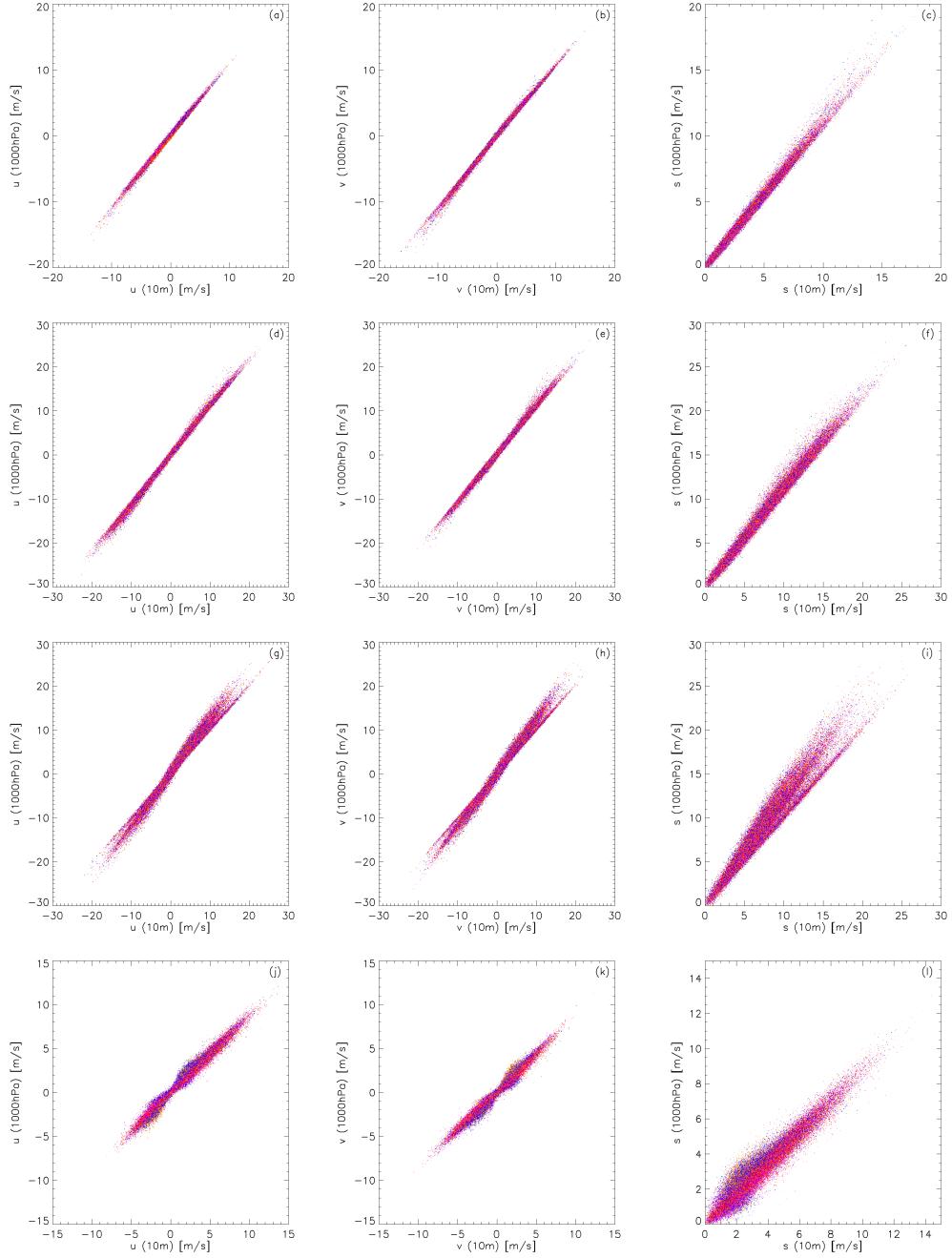


Figure 1.23: Scatter plots for 1000 hPa and 10 m wind component u (a,d,g,j) wind component v (b,e,h,k) and wind speed s data (c,f,i,l). Figures in different rows correspond to different geographical locations: figures (a,b,c) to [43N, 15E], figures (d,e,f) to [63N, 19W], figures (g,h,i) to [57N, 2E], figures (j,k,l) to [49N, 16E]. Different colors correspond to different hours of the day: blue 0 UT, purple 6 UT, red 12 UT, orange 18 UT.

1.5. WIND PROFILES - HEIGHT DEPENDENCE OF WIND SPEED 35

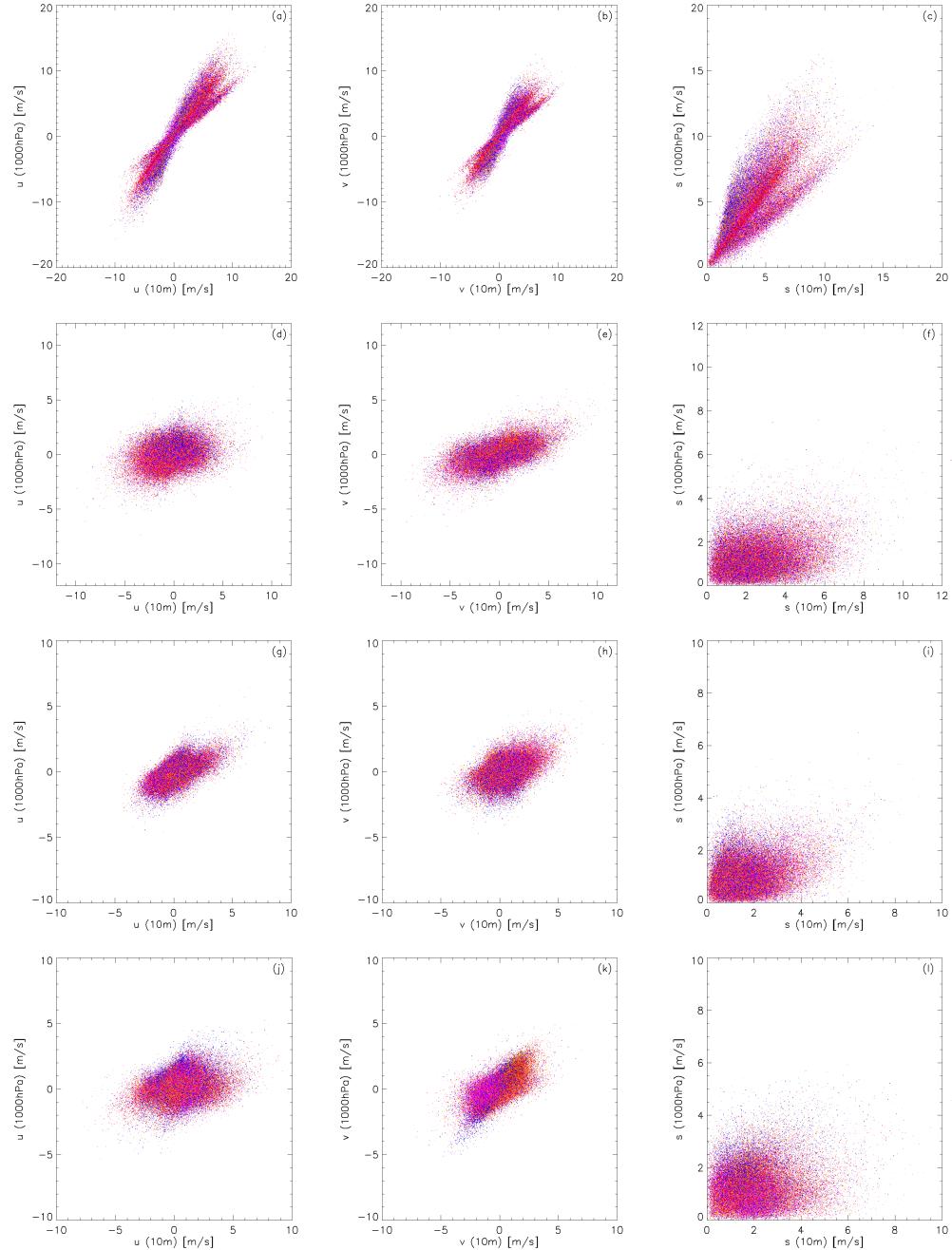


Figure 1.24: The same as figure 1.23 but for the locations: [53N,10E] for figures (a,b,c), [65N,17W] for figures (d,e,f), [62N,7E] for figures (g,h,i), [38N,2W] for figures (j,k,l).

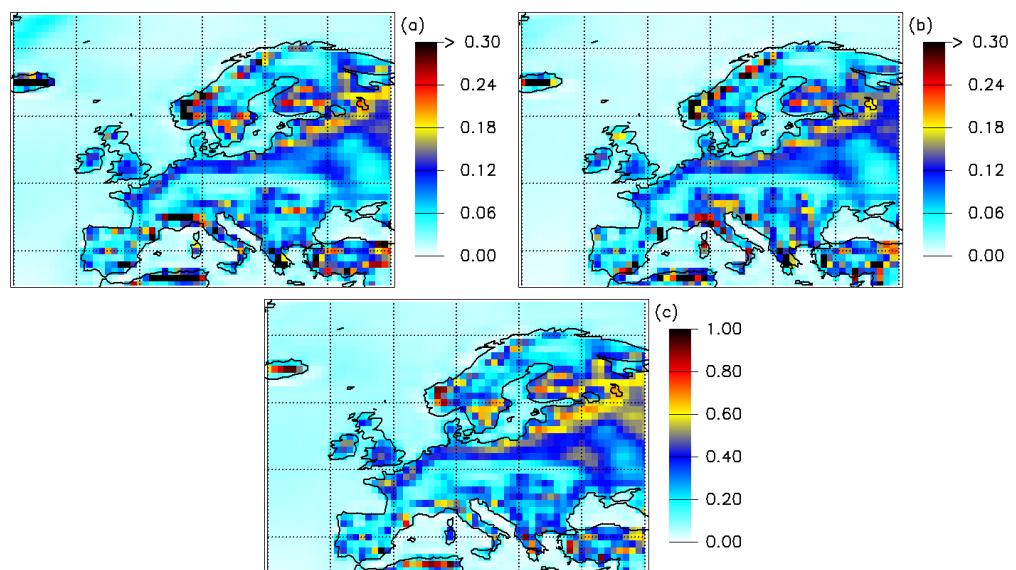


Figure 1.25: $1 - r^2$ calculated from the coefficient of correlation r between the (a) u component, (b) v component and (c) speed s of 10 m and 1000 hPa data.

1.5.3 Power law wind speed profile

In the engineering practice, it is common to use a simple power law approximation [4, 24, 26, 63, 64, 65, 66] for the wind profile:

$$\frac{s_2}{s_1} = \left(\frac{h_2}{h_1} \right)^\alpha, \quad (1.20)$$

where the indices denote two different levels with elevations h_1 and h_2 above the ground. The formula gives a relationship between wind speeds s_1 and s_2 at different levels using only one parameter α , the Hellmann exponent.

A more realistic expression for the mean wind speed at height z is provided by the logarithmic wind profile with atmospheric stability correction [4, 60, 66, 67]. Since stability considerations are beyond the scope of this work, we refer to the simplified logarithmic profile

$$\frac{s(z_2)}{s(z_1)} = \frac{\log(z_2/z_0)}{\log(z_1/z_0)}, \quad (1.21)$$

where z_0 denotes a surface roughness length. The global average [4] is around $\bar{z}_0 \approx 0.7$ m with a maximum of 3.5 m over complex terrains and with a minimum of a fraction of 1 mm over smooth orographies, like sea surface.

It is well-known that the power law relationship is only an approximation of real wind profiles, e.g. Archer and Jacobson checked 6 different formulations for each location to obtain an acceptable fit [4]. Nevertheless several measurements illustrated that a power law itself is a good approximation [63] but the Hellmann exponent α varies with stratification and atmospheric stability [68], therefore with time (season, part of day) and even with wind speed. Theoretically, a separation of different meteorological circumstances could be performed by using other ERA-40 variables (pressure, temperature, etc.) in a comprehensive analysis. However, even such a procedure cannot eliminate the mentioned main deficiency, namely that the interpolated 1000 hPa pressure level surface is often located below the physical ground level (that $h < 0$), see Figure 1.22d.

When fitting the α parameter we used a least square linear fit of the logarithm of Equation (1.20):

$$\ln \left(\frac{s_2}{s_1} \right) = \alpha \ln \left(\frac{h_2}{h_1} \right), \quad (1.22)$$

where $h_1 = 10$ m, $h_2 = h$ was calculated using Equation (1.19) for the 1000 hPa level, $s_1 = s_{10m}$ represents 10 m and $s_2 = s_{1000hPa}$ 1000 hPa wind speed. Because of the possible wind speed dependence of α we fitted the model in

various bins (e.g. data at separate locations which obeyed $0 \text{ m/s} < s_1 < 1 \text{ m/s}$; $1 \text{ m/s} < s_1 < 2 \text{ m/s}$; etc.). We found a weak wind speed dependence, that is α slightly decreases over land and slightly increases over the seas as wind speed increases. The model fits relatively poorly in most wind speed regimes, even $r \approx 0$ coefficients of regression are obtained at very low wind speeds. The model fits best for s_1 slightly larger than the mode ($r > 0.8$). Hellmann exponents for two selected wind speed regimes can be seen in Figures 1.26 (4 m/s $< s_{10m} <$ 6 m/s) and 1.27 (8 m/s $< s_{10m} <$ 10 m/s).

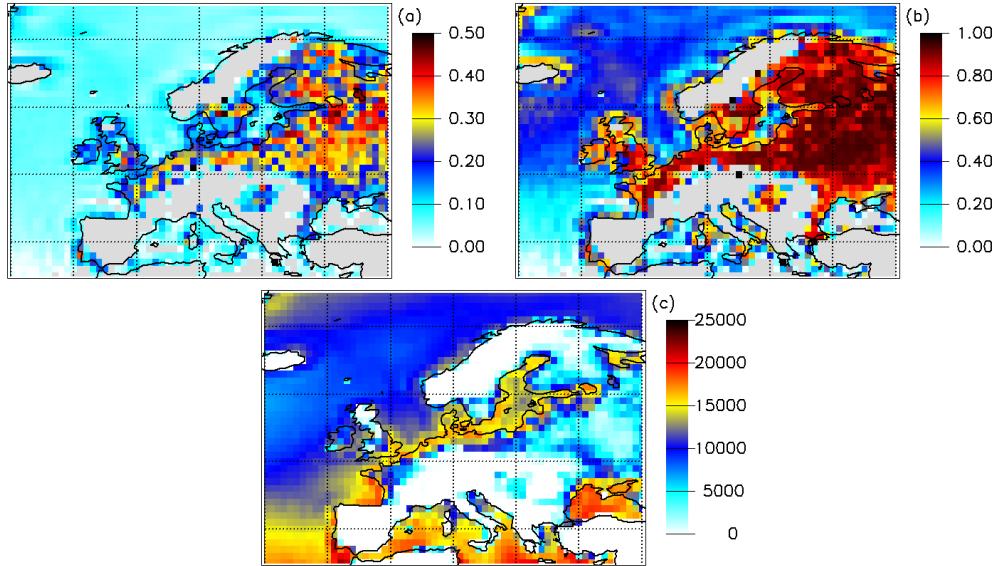


Figure 1.26: (a) Hellmann exponent α , (b) the coefficient of correlation r and (c) the number of data used in the fit. Data were restricted to the $4 \text{ m/s} < s_{10m} < 6 \text{ m/s}$ wind speed regime. Grey color indicates areas where the fit was not possible.

Because of the observed weak wind speed dependence it is plausible to assume that α is independent of wind speed. To fit this model we performed the procedure as described before, then averaged the fitted α exponents but only those bins were used in averaging where there were enough data points (at least 10) and the fit was sufficiently good ($r > 0.8$). These values are somewhat arbitrary, but they are based on values obtained above. Wind speed independent Hellmann exponents fitted by this way can be seen in Figure 1.28

A plausible way to get an impression about the quality of the above fits is to compare original data at the varying 1000 hPa level and those obtained from 10 m data using the power law relationship. In case of a perfect agreement, we would be able to say that the height dependence is adequately captured and the effects of varying meteorological conditions in Figure 1.23c,

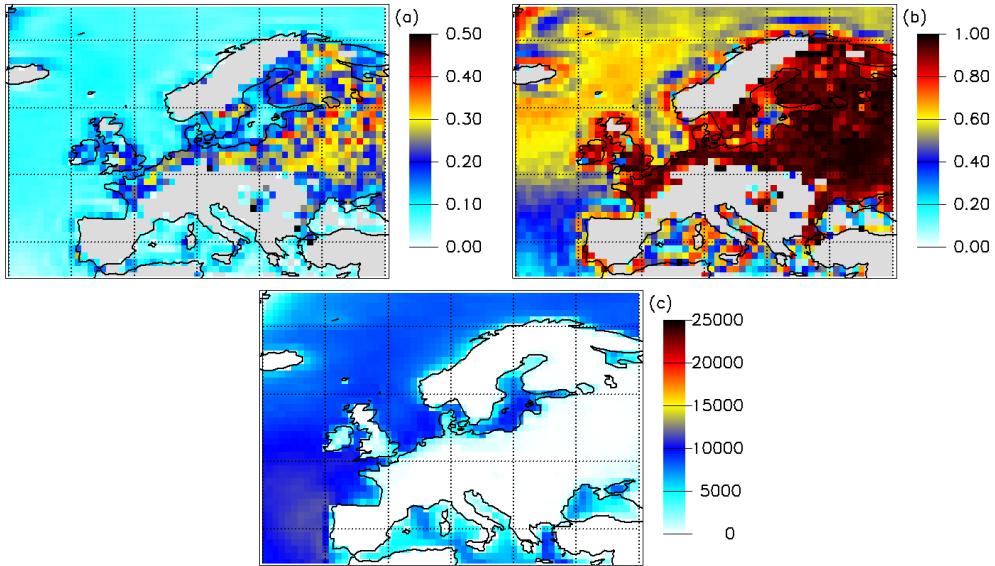


Figure 1.27: (a) Hellmann exponent α , (b) the coefficient of correlation r and (c) the number of data used in the fit. Data were restricted to the $8 \text{ m/s} < s_{10m} < 10 \text{ m/s}$ wind speed regime. Grey color indicates areas where the fit was not possible.

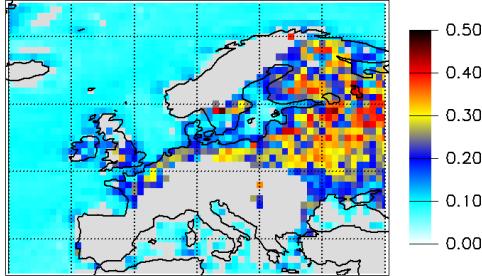


Figure 1.28: Hellmann exponent α supposing it is independent of wind speed. Grey color indicates areas where the fit was not possible.

f, i, l and Figure 1.24c, f, i, l are completely eliminated. As it can be seen in Figure 1.29a, the agreement is not perfect at a selected location in northern Germany [53° N , 10° E] in the case of constant Hellmann exponent (constant both in speed and time). But the major effects of meteorology visible in Figure 1.24c have been transformed out. Height dependence is described only with considerable errors. In the case of a speed dependent Hellmann exponent, the basically linear relationship of Figure 1.29a is replaced by a nonlinear function (not shown here), but we do not get a better data collapse. On the other hand, by fitting power law relationships separately to different parts of the day, thus considering the time dependence of α , we get slightly

better results (see Figure 1.29b). The large error of the fit is unfortunately still not reduced significantly. α exponents fitted to different hours of the day can be seen in Figure 1.30.

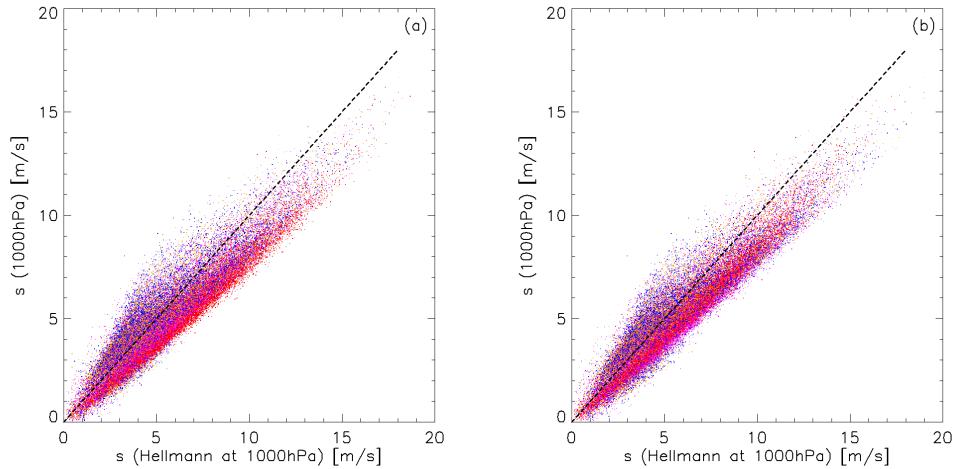


Figure 1.29: Scatter plots between 1000 hPa wind speed and calculated wind speed for the 1000 hPa level based on 10 m data using (a) constant (in speed and time) Hellmann exponent and (b) time-dependent (but constant in speed) Hellmann exponent. Plots correspond to the location [53N,10E]. Compare figures with figure 1.24c.

What we can say is that the power law relationship can describe the height dependence of the two levels of ERA-40 wind speeds, but with considerable errors. The wind speed dependence of the only parameter, the Hellmann exponent α proved to be negligible. The time dependence of α is somewhat stronger, but since we do not get significantly better fits when taking it into account, it can be neglected as well. Hellmann exponents can be extracted at most locations except for sites with higher elevation (greater than a few hundred meters). In general, weaker height dependence of wind speed s is found above the seas ($\alpha \approx 0.1$) and stronger above land ($\alpha \approx 0.3$).

1.5.4 Empirical wind profile

The power law wind profile fits relatively poorly at some locations. Furthermore, wind profile data cannot be extracted at several locations based on measured data at two levels only. Therefore we adopted another method and calculated the average empirical wind profile by aggregating all available data above terrestrial sites. Later we use this profile to estimate potential wind power production.

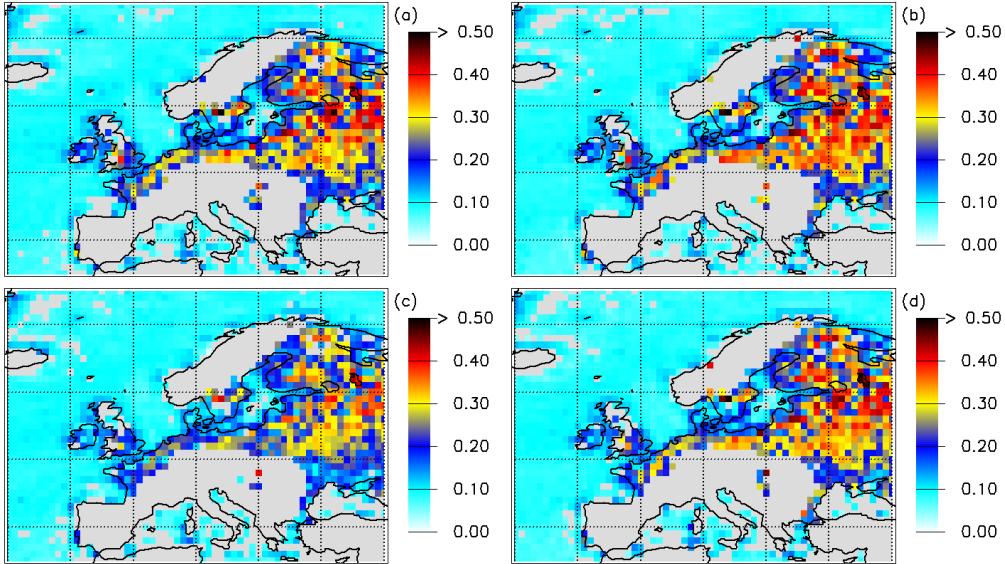


Figure 1.30: Wind speed independent Hellmann exponent α fitted separately for different parts of the day: (a) 0 UT, (b) 6 UT, (c) 12 UT, (d) 18 UT. Grey color indicates areas where the fit was not possible.

In order to improve statistics, we performed spatial averaging over the continent, the results are shown in Figure 1.31. At a given bin of 1000 hPa pressure height $h \pm 2.5$ m, the speed ratio histograms exhibit well defined mode (most probable) values (Figure 1.31a). The mode wind speed ratio profile (Figure 1.31b) cannot be perfectly fitted neither by the logarithmic [69, 70] nor the power-law [63, 64, 65] assumptions used widely in the literature.

Consequently, we introduced the simplest possible estimate of wind speed at the elevation of h above ground level based on the available data through multiplying 10 m instantaneous wind speeds by the corresponding value of the empirical wind speed ratio profile $\eta(h)$:

$$s_h \approx \eta(h)s_{10} . \quad (1.23)$$

Setting $h = 100$ m results in $\eta = 1.28$ and Equation (1.23) gives a crude measure of wind speed for 100 m over land (see Figure 1.31a) and 100-130 m over sea surface. As it can be seen in Figure 1.31b, the empirical wind profile is loaded with considerable errors. Different multipliers η obviously change the numerical values for wind power estimates, but do not change the main conclusions of the following sections.

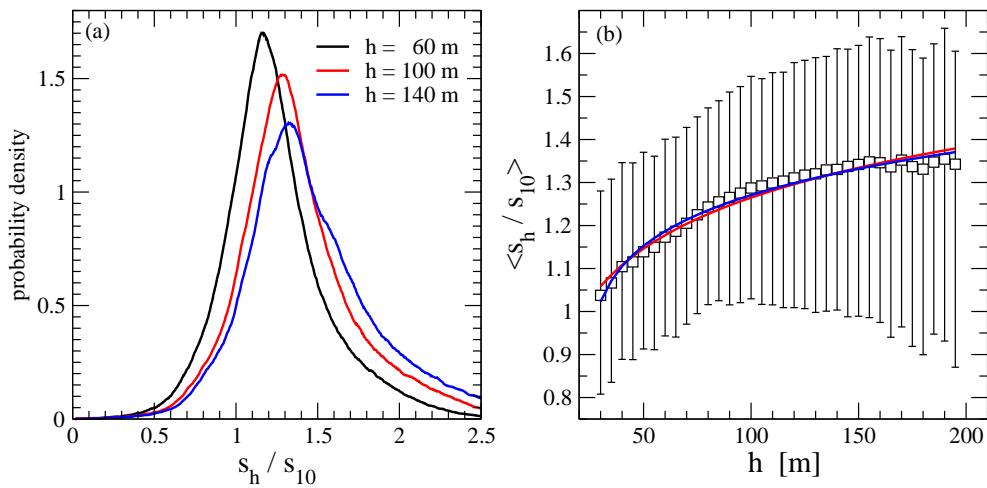


Figure 1.31: (a) Normalized histogram of wind speed ratio s_h / s_{10} for three different bins $h \pm 2.5$ m (see labels), where h is the average 1000 hPa pressure height. The histograms contain aggregated data over the continental geographic area. (b) The mode of wind speed ratio profile $\eta(h)$ as a function of 1000 hPa pressure height (white symbols). The error bars indicate the half width of histograms shown in (a). Red line is a logarithmic fit, blue line illustrates a power-law with an exponent ~ 0.1 .

1.5.5 Comparison of tower measurements and reanalysis wind speeds – a Hungarian case study

In this section our aim is to compare ERA-40 wind speed data and turbine measurements at Mosonszolnok, Hungary, and thus characterize the appropriateness of reanalysis wind speed data for local power estimation (see also [71]).

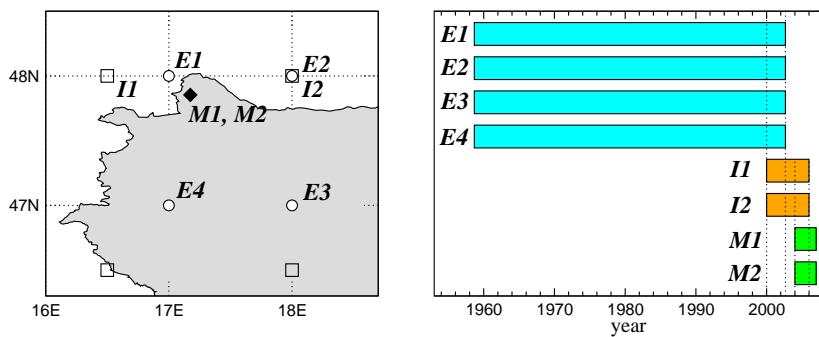


Figure 1.32: Sketch of the geographic setting of the study area and time-line of the records. Heavy diamond indicates the location of the two Enercon E-40 wind turbines M1 and M2 (47.816° N, 17.174° E), gray shading signs Hungary, E1–E4 label the nearby ERA-40 grid points (empty circles), I1 and I2 (empty squares) locate the grid points from the ERA-Interim data base used for comparison. The time-line illustrates the overlapping periods: 01/01/2000 - 08/31/2002 for E and I records, and 01/01/2004 - 12/31/2005 for I and M time series.

Since there is no temporal overlap between the ERA-40 and turbine time series (see Fig. 1.32 and Fig. 1.4 for a more complete picture on dataset time-lines), a direct comparison is not possible. Fortunately, the 3rd generation reanalysis, ECMWF ERA-Interim [10] has become available till 2009. Two grid points are close enough to the location of turbines (see Fig. 1.4), thus time series in the period 01/01/2000 - 12/31/2005 are used to bridge the temporal gap between ERA-40 and the turbine measurements (Fig. 1.32).

Figure 1.33 shows wind speed time series for the initial 100 days of the overlapping period at the sites I1, I2, and E1...E4 (note that I2 and E2 refer to the Interim and ERA-40 records for the same geographic grid point, see Fig. 1.32). It is easy to see that the wind speed changes quite homogeneously over the given area, the gross dynamical features are almost identical. The apparent synchrony can arise, on one hand, from the smooth orography: the area belongs to the Little Hungarian Plain, a low lying tectonic basin of approx. 8000 km^2 in northwestern Hungary, southwestern Slovakia, and eastern Austria. On the other hand, the 6 h time resolution of the records

can easily hide shorter temporal shifts between the wind speed signals.

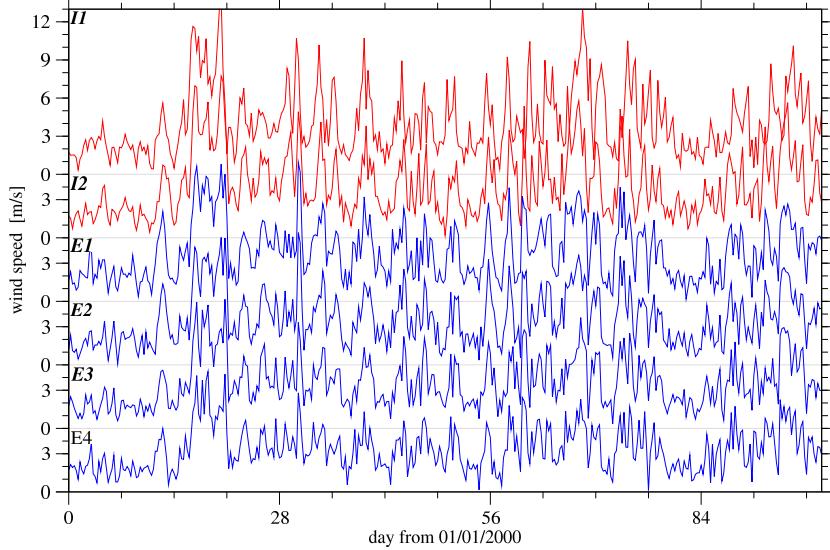


Figure 1.33: Wind speed records for the I_1 , I_2 , and $E_1 - E_4$ grid points (see Fig. 1.32) in the first 100 days of the overlapping period.

The same is true for the wind speed records at sites I_1 , I_2 , and M_1 , M_2 in the overlapping period of two years, see Fig. 1.34 (since M_1 and M_2 speeds are practically identical in absence of technical failures, therefore only the former is shown). Note that 6-hour average values are determined from the turbine data series, where the time stamps are centered to the UT sampling time of ERA-40 and Interim records. This is because the reanalysis wind fields are quite smooth (sub-grid scale turbulence is not resolved), wind gusts and lulls are not represented. In Fig. 1.34, the agreement between the turbine measurements ($h = 65$ m) and surface wind speeds ($h = 10$ m) is not as strong as in Fig. 1.33, nevertheless the temporal evolution of the turbine record is properly reflected by the reanalysis.

In order to quantitatively characterize the strength of synchrony, we computed the usual two-point correlation matrix for the overlapping periods by

$$\text{Corr}(s_i, s_j) = \frac{\langle (s_i(t) - \bar{s}_i)(s_j(t) - \bar{s}_j) \rangle_t}{\sigma_i \sigma_j}, \quad (1.24)$$

where $i, j \in \{E_1, E_2, E_3, E_4, I_1, I_2, M_1, M_2\}$, $s(t)$ denotes the wind speed of average value \bar{s} and standard deviation σ , and $\langle \cdot \rangle_t$ indicates temporal averaging. The results are shown in Table 1.1.

The lower diagonal represents the geographic distances between the sites in units of kilometers. It is worth to mention that the corresponding cross

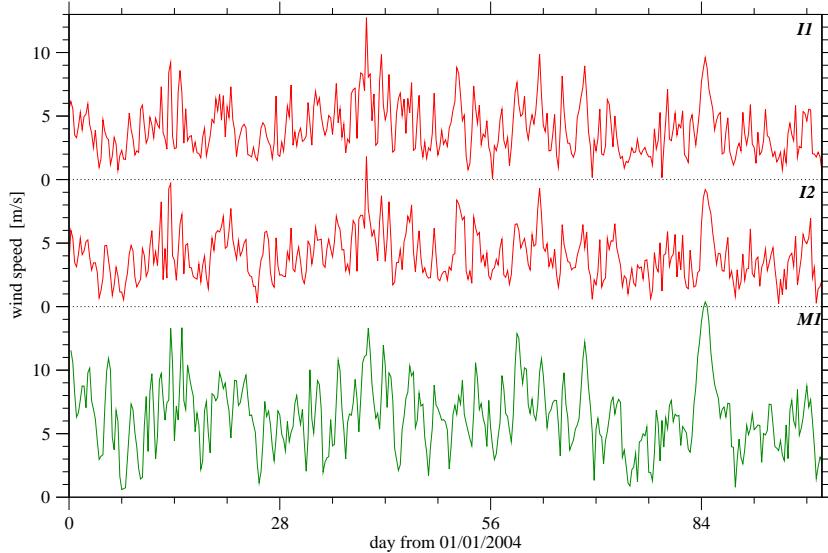


Figure 1.34: The same as Fig. 1.33 for the I1, I2 (red) and M1 (green) records.

	<i>E1</i>	<i>E2</i>	<i>E3</i>	<i>E4</i>	<i>I1</i>	<i>I2</i>	<i>M1</i>	<i>M2</i>
<i>E1</i>	1	0.929	0.824	0.873	0.797	0.772	—	—
<i>E2</i>	(74.6)	1	0.902	0.841	0.765	0.847	—	—
<i>E3</i>	(134.3)	(111.2)	1	0.913	0.716	0.820	—	—
<i>E4</i>	(111.2)	(134.3)	(76.1)	1	0.742	0.757	—	—
<i>I1</i>	(37.3)	(111.9)	(158.5)	(117.4)	1	0.880	0.768	0.743
<i>I2</i>	(74.6)	(0)	(111.2)	(134.3)	(111.9)	1	0.753	0.734
<i>M1</i>	—	—	—	—	(53.7)	(63.3)	1	0.971
<i>M2</i>	—	—	—	—	(53.7)	(63.3)	(0.4)	1

Table 1.1: Equal time two-point correlation (see Eq. (1.24)) matrix for the time series in the overlapping periods (upper diagonal), and geographic distance in units of km (lower diagonal, in parentheses).

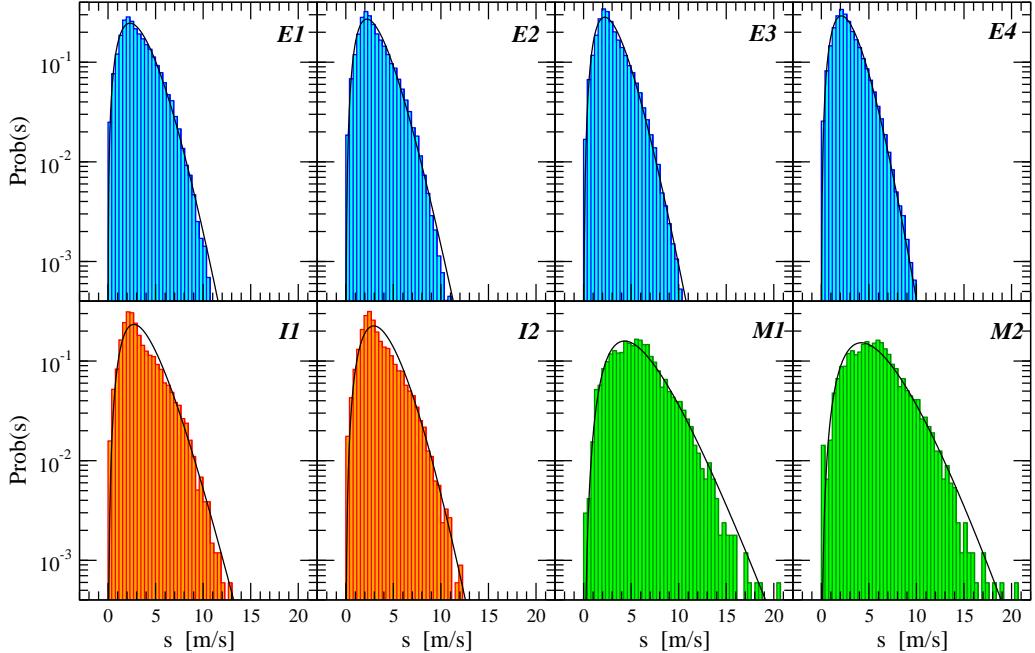


Figure 1.35: Normalized histogram of wind speed s for the time series $E1, E2 \dots M2$, and fitted generalized gamma (see Eq. (1.10)) probability density functions (thin black lines). The empirical parameters are summarized in Table 1.2.

correlation functions (not shown here) exhibit a quick drop for nonzero time lags both in the negative and positive directions, which means that the wind field is essentially coherent in time windows of 6 hours over the given area. The correlation matrix (Table 1.1) seems to be consistent in the sense that larger spatial separation usually entails a lower correlation coefficient. The differences between ERA-40 and Interim records are characterized by the matrix element $\text{Corr}(s_{E2}, s_{I2}) = 0.847$ for the same grid point.

We tested the effect of spatial interpolation with the records $I1$ and $I2$ by means of inverse squared distance weights [72] for the turbine location. Since the improvement was negligible, we used the original time series for further analysis.

Statistical differences are also present when the histograms of wind speeds are compared. Figure 1.35 illustrates the normalized empirical probability densities of wind speeds for each record with the corresponding generalized gamma fits (see Equation (1.10)). The values of the generalized gamma parameters are listed in Table 1.2 for each record. The fitted curves in Fig. 1.35 have very similar shapes, nevertheless the parameters can be quite different. This is because s_0 , $q = \epsilon k - 1$, and k are very sensitive to small,

	\bar{s} [m/s]	R_{M1}	σ [m/s]	s_0 [m/s]	q	k	m [m/s]
E1	3.19	1.79	1.75	2.315	1.404	1.411	2.306
E2	3.06	1.87	1.64	1.409	1.990	1.172	2.216
E3	2.99	1.91	1.56	1.345	2.145	1.184	2.219
E4	2.82	2.02	1.48	1.753	1.734	1.352	2.106
I1	3.43	1.66	1.96	1.291	2.423	1.084	2.712
I2	3.51	1.63	1.90	2.310	1.952	1.391	2.947
M1	5.71	1	2.67	1.273	3.195	0.975	4.302
M2	5.70	1.00	2.72	2.448	2.241	1.182	4.207

Table 1.2: Fitted parameters of the normalized wind speed probability distributions for the time series E1, E2 ... M2, see Fig. 1.35. The mean value \bar{s} and standard deviation σ are taken over the entire available record lengths (see Fig. 1.32), the parameters of the generalized gamma distribution (see Eq. (1.32)) are denoted by s_0 , $q = \epsilon k - 1$ and k , and the mode (most probable value) is $m = (q/k\beta)^{1/k}$. The second column ($R_{M1} = \overline{s_{M1}}/\bar{s}$) is the ratio of average wind speed $\overline{s_{M1}}$ at the turbine M1 and at the given site \bar{s} .

sometimes hardly visible changes at the tails of the histograms.

The main difference between reanalysis and turbine histograms is that the later exhibits significantly higher average value and width, due to the increased elevation. The overlapping time interval between the ERA-Interim reanalysis and turbine records allows to test the assumption of a constant profile – a linear relationship between wind speed values at the surface (10 m) and at the hub height of 65 m:

$$s(h = 65 \text{ m}) = \eta^*(h = 65 \text{ m}) s_{10} . \quad (1.25)$$

The equal time scatter plot for I1 and M1 time series is shown in Fig. 1.36a. A linear relationship (Eq. (1.25)) seems to be plausible as a first approximation with $\eta^*(h = 65 \text{ m}) = 1.529$, however the errors involved are far from being negligible. The effect of the simple linear transformation $s'_{I1} \rightarrow 1.529 \times s_{I1}$ is illustrated in Figs. 1.36b and 1.36c. While the fitted generalized gamma distributions collapse almost perfectly (Fig. 1.36c), the empirical histograms (Fig. 1.36b) exhibit characteristic differences already apparent in Fig. 1.35. The most probable empirical value (the peak) is shifted to larger speeds for turbine data with respect to an “ideal” generalized gamma curve, contrarily, the peak is shifted toward lower values for the reanalysis histogram. Actually, all of the empirical histograms in Fig. 1.35 seem to be a mixture of at least two unimodal distributions, when a fit of really good quality is attempted. Such attribute can be a consequence of the strong seasonality common in mid

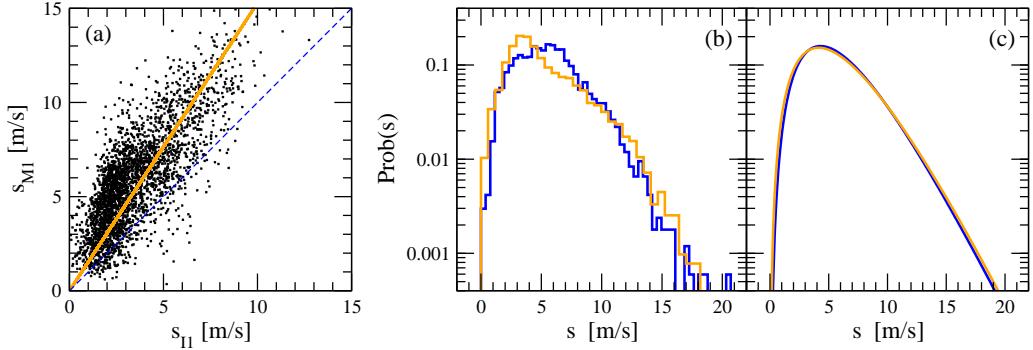


Figure 1.36: (a) Scatter plot for the equal time wind speeds measured at site I1 and M1. The regression line has the slope of 1.529. (b) Normalized empirical wind speed histograms for the M1 turbine measurements (dark, blue) and the transformed I1 record (light, orange): $s'_{I1} \rightarrow 1.529 \times s_{I1}$. (c) The same as (b) for the fitted probability distributions.

latitudes, where the different seasons are connected with different prevailing winds [41].

The conversion factor $\eta^*(h = 65 \text{ m}) = 1.529$ between surface and hub height wind speeds obtained for Mosonszolnok, Hungary is considerably higher, than the average conversion factor $\eta = 1.28$ derived from the ERA-40 reanalysis. Furthermore the latter average η is thought to correspond to hub heights of about 100 m.

The big difference can be due to (1) the effect of averaging used to calculate η , because the coastal areas have a great weight in the average, and in these areas the wind profile is less steep, than at inland sites. Due to this effect the extrapolated inland hub height wind speeds can be underestimations of the real situation (and offshore wind speeds are probably overestimations). (2) Another phenomenon which can create a considerable difference between η and η^* can be the presence of sub-grid effects not captured by the reanalysis. Wind turbines are - for economical reasons - installed at the windiest possible spots, where the local wind speeds may be significantly higher due to orographic enhancements.

In an ideal situation one should be able to measure η^* for every wind farm - a dense network across Europe - and thus “calibrate” the reanalysis wind fields. However because of the lack of turbine data we cannot assess this issue for more than a few sites in Hungary.

In the following analyzes we decided to use the average $\eta = 1.28$ conversion to get hub height wind speeds from surface reanalysis fields, however in some cases we tested the effects of higher, but still spatially constant conversion factors, $\eta_1 = 1.50$ and $\eta_2 = 1.79$. In one case we examined the effects

1.5. WIND PROFILES - HEIGHT DEPENDENCE OF WIND SPEED 49

of using different conversion factors at onshore and offshore areas. For these tests see Section 1.7.5.

1.6 Wind power estimations

1.6.1 Transformation of surface wind speed to wind power

As a first step, we transformed surface (10 m) wind speed s_{10} values to hub height (typically 100 m at on-shore sites and 100-130 m at offshore sites) wind speed. This was done by using the empirical wind speed profile (Section 1.5.4) at $h = 100$ m. Thus we obtained that the wind speed at hub height equals to the instantaneous surface (10 m) wind speed multiplied by $\eta = 1.28 : s_{100} \approx 1.28s_{10}$. As mentioned earlier (Sections 1.5.4 and 1.5.5) in this process considerable errors are involved, but we believe that these errors do not affect the main conclusions of Section 1.7.

The second step is to calculate wind power estimates from wind speed data. The standard method for a given turbine is based on the power curve [24] provided by the producers (see the power curves, e.g. at <http://www.enercon.de>, <http://www.nordex-online.com>, <http://www.vestas.com>). The rated power can be very different for smaller and larger generators, however they share several other aspects. The cut-in wind speed is usually 3-5 m/s, then the power curve exhibits a power-law range with an exponent value close to 3. There is a crossover at around 11-15 m/s to a plateau (representing the rated power regime of active blade pitch control), and energy production falls at the cut-off wind speed at about 20-25 m/s for most commercial constructions (refer to the power curves mentioned above). An idealized power curve is depicted in Figure 1.37, where the following simple form is used:

$$p(s) = \begin{cases} a_0(s - s_{ci})^\beta & \text{if } s_{ci} < s \leq s_x \\ \frac{a_1}{1 + \exp[-(s - b)^c]} & \text{if } s_x < s \leq s_{co} \\ 0 & \text{if } s \leq s_{ci}, s > s_x \end{cases} \quad (1.26)$$

Here $s_{ci} = 1.2$ m/s is the cut-in, $s_x = 10$ m/s is the crossover, $s_{co} = 25$ m/s is the cut-off speed, $b = 2.8$ is the exponent, $a_0 = 0.166$, $a_1 = 103.3$, $b = 9.16$, and $c = 1.05$ are empirical parameters obtained by fitting high resolution ($\Delta t = 10$ minute) measured data of two Enercon E-40 turbines (rated power 600 kW) installed near Mosonszolnok, Hungary (see Figure 1.37).

Here we emphasize again that fine details can depend on the precise mathematical form of the power curve used, however they cannot significantly affect the main conclusions.

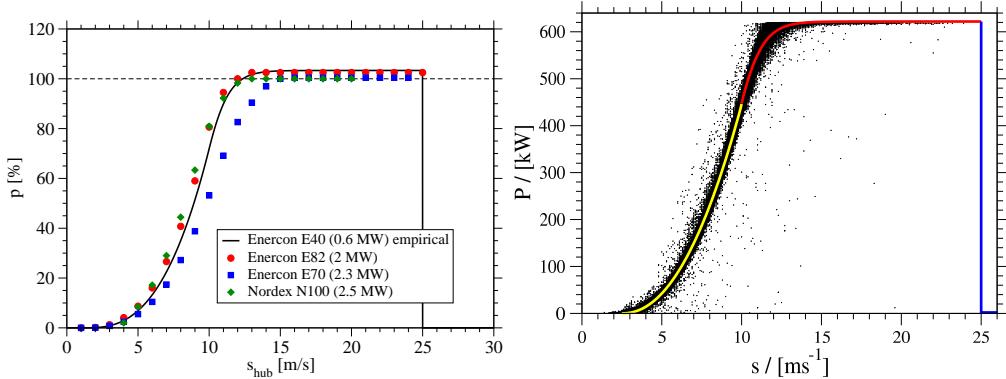


Figure 1.37: Left panel: the power curve used for wind power estimates (see Equation (1.26)) expressed in capacity factor. The brochure data for different turbines given by the manufacturers are shown as well. Right panel: the same curve, as shown in the left panel, with measurements for the M1 turbine, Mosonszolnok, Hungary. The different segments of the curve (Equation (1.26)) are shown with different colors.

1.6.2 Comparison of empirical wind power histograms and transformed generalized gamma wind speed distributions

As concluded earlier (Section 1.3), the generalized gamma model gives the best representation of wind speed histograms. Since wind power estimates are analytically calculated from wind speed data, it is possible to transform the wind speed distributions analytically to obtain the corresponding wind power distribution. The question arises whether the transformed generalized gamma distribution gives an appropriate representation of wind power histograms. Two examples of empirical wind power probability density functions can be seen in Figure 1.38. As it is apparent, the transformed generalized gamma distribution gives a perfect representation of the power histograms. The very good agreement between transformed generalized gamma and empirical wind power distributions illustrated in Figure 1.38 is valid at other geographical locations as well. The good fit of the curves in Figure 1.38 confirms however only the goodness of the generalized gamma fits to the reanalysis wind speeds. Note that both the curves and histograms in Figure 1.38 are based on the reanalysis data. The comparison with real measurements is discussed in the next section (Section 1.6.3).

The transformed probability density function of wind power p ($0 \leq p \leq a_1$) takes the following form:

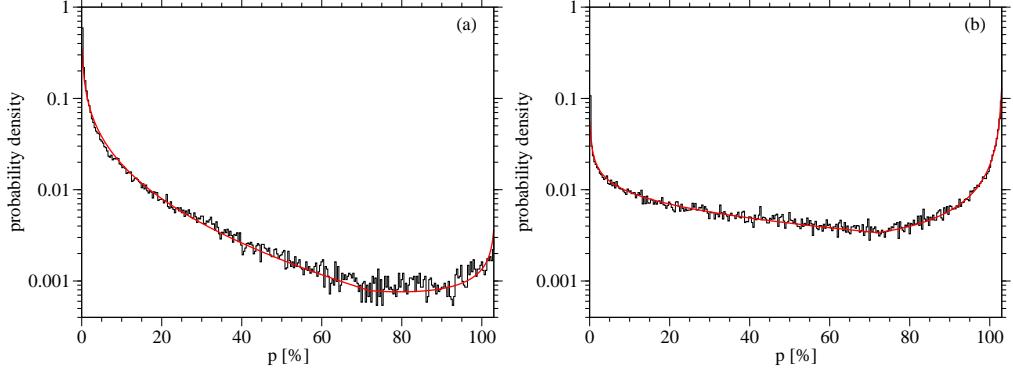


Figure 1.38: Empirical wind power histograms based on calculations described in Section 1.6.1 for (a) a terrestrial location in Northern Germany (52° N, 11° E) and (b) a marine location (66° N, 4° W). Red line indicates the transformed generalized gamma distribution (fitted to corresponding surface wind speed data).

$$P(p) = \begin{cases} \left(\int_0^{\eta s_{ci}} P_{GG}(s; s_0, k, \epsilon) ds + \int_{\eta s_{co}}^{\infty} P_{GG}(s; s_0, k, \epsilon) ds \right) \delta(p) & , \text{if } p = 0 \\ P_{GG} \left(\left[\left(\frac{p}{a_0} \right)^{1/\beta} + s_{ci} \right] / \eta; s_0, k, \epsilon \right) / \left[\eta \beta a_0 \left(\frac{p}{a_0} \right)^{\frac{\beta-1}{\beta}} \right] & , \text{if } 0 < p \leq p_x \\ P_{GG} \left(\left\{ \left[-\ln \left(\frac{a_1}{p} - 1 \right) \right]^{1/\epsilon} + b \right\} / \eta; s_0, k, \epsilon \right) / \\ / \left\{ \eta c \frac{p^2}{a_1} \left(\frac{a_1}{p} - 1 \right) \left[-\ln \left(\frac{a_1}{p} - 1 \right) \right]^{\frac{c-1}{c}} \right\} & , \text{if } p_x < p \leq a_1 , \end{cases} \quad (1.27)$$

in which we used the generalized gamma distribution $P_{GG}(s; s_0, k, \epsilon)$ (see Equation (1.10)) of $s = s_{10}$ wind speed with parameters s_0 , k and ϵ . Parameters s_{ci} , s_x , s_{co} , a_0 , b , a_1 , b and c are used in the transformation of wind speed at hub height to wind power (see Section 1.6.1). Parameter η is the ratio of hub height wind speed (100 m) and surface wind speed (10 m) (see Section 5.1). $\delta(\dots)$ represents the Dirac- δ function. The integrals in Equation 1.27 (in the case of $p = 0$) can be evaluated numerically using incomplete gamma functions.

1.6.3 Comparison of reanalysis-based wind power data with tower measurements

As an addition to wind speed measurements at Mosonszolnok, wind power output measurements were also available to us. This permits a comparison of actual wind power and reanalysis-based estimates at one site.

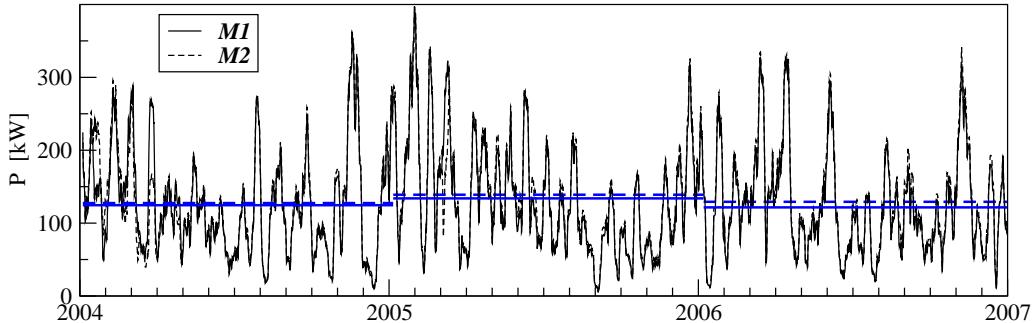


Figure 1.39: Electric power output measured at the turbines M1 (thin black line) and M2 (dashed black line) in three consecutive years, high frequency (10 min) data are smoothed by 1008 point (1 week) running average. Horizontal (blue) lines indicate the annual averages. (Percentage values with respect to the rated power are for M1: $20.76 \pm 11.55\%$, $22.33 \pm 12.58\%$, $20.27 \pm 11.24\%$, and for M2: $21.26 \pm 11.66\%$, $23.16 \pm 12.24\%$, $21.55 \pm 11.58\%$, respectively.)

Figure 1.39 illustrates the smoothed output power measured directly at the turbines M_1 and M_2 in three working years. The two curves are very similar, significant differences are due to technical problems. It is remarkable that the apparent seasonality is very weak, and the fluctuations are huge in spite of the smoothing by a wide window of one week. The average load factor or capacity factor (the ratio of mean and rated output power) around 21-22 % is considerably lower than at optimal off-shore sites [73, 74, 75], in contempt of the geographic location of turbines belonging to the windiest region of Hungary [76, 77, 78, 79, 80].

We used the standard method described in Section 1.6.1 for estimating wind power from (reanalysis) wind speed data. The only difference was that we used $\eta^* = 1.529$ and other conversion factors instead of $\eta = 1.28$, as used in Section 1.7.

Based on the fitted curve Eq. (1.26), an estimate for output power can be given at the sites where wind speed records are available. A comparison between turbine measurements and modeled time series at seemingly distant geographic locations seems to be reasonable because of the strong spatial correlations revealed in Section 1.5.5.

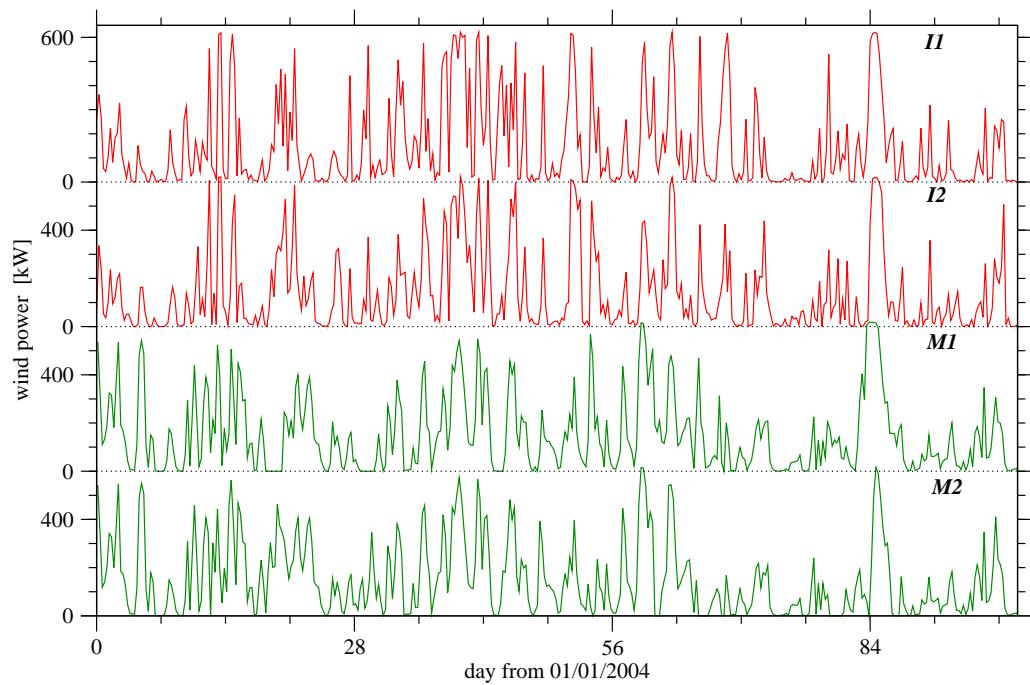


Figure 1.40: Comparison of estimated electric output based on $\eta^ = 1.529$ (see Fig. 1.36) and the power curve Eq. (1.26) from 6 hourly wind speed data for sites I1 and I2 (red) with direct measurements at the turbines M1 and M2 (green). Note that the later two curves are 6 h averages of 10 min records.*

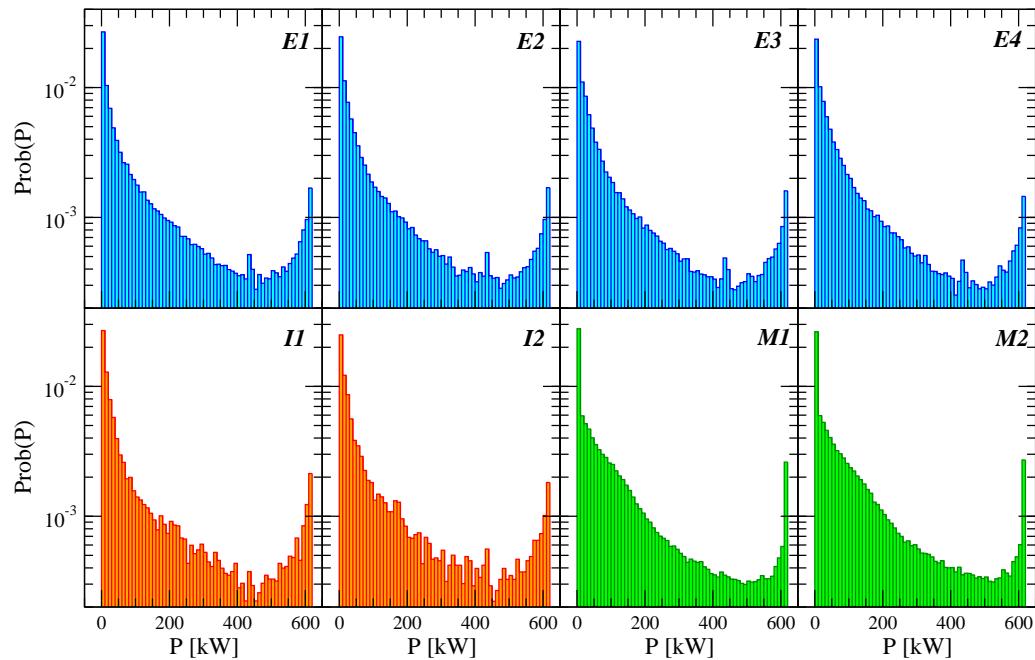
A point-by-point comparison of reanalysis-based power estimates with turbine data is possible only for sites $I1$ and $I2$ in the overlapping period of two years (see Fig. 1.32). Time series for the first 100 days are shown in Fig. 1.40. The general behavior is similar to the wind speed curves in Fig. 1.34: the gross features are reproduced with apparent deviations in fine details.

Collating of statistical properties has shown that model time series produce systematically lower mean (annual) capacity factors than the measured values in Fig. 1.39. The reason is the marked difference in the histograms of the turbine and scaled reanalysis data shown in Fig. 1.36b: the probability of wind speed in the range of 3–10 m/s is skewed to larger values for the turbines, while an opposite skew is apparent in the reanalysis data. Note that the very wind speed interval is critical, because small differences in numerical values are strongly exaggerated by the cubic behavior of the power curve (Fig. 1.37).

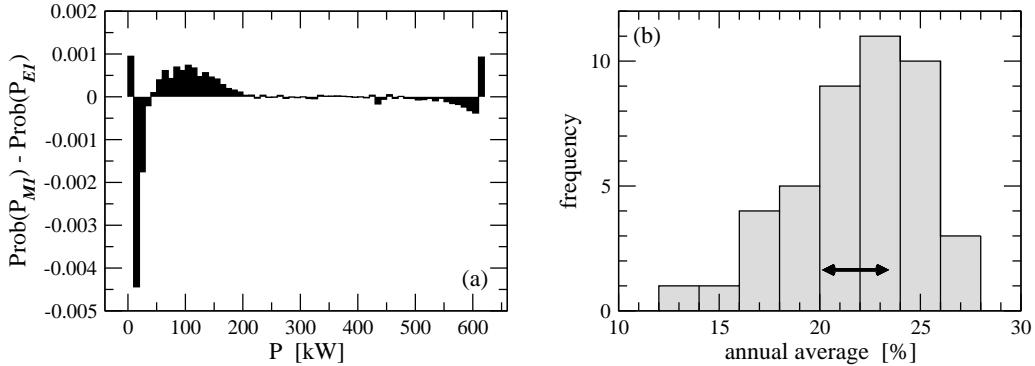
As we mentioned in Section 1.5.5, the shape of wind speed histograms in Figs. 1.35 and 1.36b cannot be perfectly fitted by simple functional forms, because they seem to be mixed (see also Section 1.4). For the same reason, an uncomplicated transformation resulting in an adequate collapse of the histograms in Fig. 1.36b is not possible. Nevertheless Fig. 1.36a suggests a relatively simple relationship between the heights of 10 and 65 m, therefore a systematic search of a proper multiplier is performed to obtain the best agreement between model and empirical capacity factors.

We have tested the the mean, mode and median of the wind speed histograms as possible scaling factors. Simple statistical comparison of measured and modeled power outputs revealed that the best result is provided by matching the long term average wind speed values \bar{s} listed in Table 1.2, first column, numerical values of the appropriate multiplier $R_{M1} = \bar{s}_{M1}/\bar{s}$ are given in the second column. Note that these factors impair the agreement for the wind speed histograms by giving too large frequencies to large values, however this is fully indifferent from the point of view of power estimation because of the plateau regime (see Fig. 1.37).

Model and empirical output power histograms are shown in Fig. 1.41. The shape of the histograms is highly nontrivial, it is produced by the nonlinear transform of Eq. (1.26) from the empirical wind speed distributions in Fig. 1.35. It is no wonder that the best matching of turbine and reanalysis data is found for the closest site $E1$. Figure 1.42a illustrates the difference between the histograms of $E1$ and $M1$, the agreement is fairly good. The largest deviation is around very small but nonzero output power, where the frequency of appropriate wind speeds is overrepresented in both ERA-40 and ERA-Interim data. The length of ERA-40 records permits to compute es-



*Figure 1.41: Normalized histograms of electric power P for the time series $E1, E2 \dots M2$. The last two (green) are based on direct measurements, the others are estimated by the power curve Eq. (1.26) from the rescaled surface wind speeds $s' = \eta^{**} s$, where the numerical value of $\eta^{**} = R_{M1}$ is given in the second column of Table 1.2. (The vertical scale is logarithmic.)*



*Figure 1.42: (a) The difference between the histograms for sites M1 and E1 shown in Fig. 1.41. (b) Histogram of annual average output power estimated from the E1 record by $\eta^{**} = R_{M1} = 1.79$ and the power curve Eq. (1.26). The black arrow indicates the range of measured averages at M1 and M2.*

timated annual capacity factors for 43 whole years, the related histogram is shown in Fig. 1.42b. The width of this histogram is surprisingly large, however a comparison with reality would need much longer turbine records.

We can say, that the comparisons of reanalysis wind fields and direct measurements at two turbines revealed that the dynamics is appropriately reproduced, apart from magnitudes. Note that this can be a consequence of the essentially flat surface configuration around the test site, much larger deviations between reanalysis and measured data are obtained over a complex terrain [13].

A key point of wind power estimation from surface speed values is the approximation of speeds at hub heights. The scatter plot Fig. 1.36a cannot disqualify the assumption of a linear relationship between 10 and 65 m data. The slope of the regression line provides the simplest tool of speed estimates for the increased altitude. However, scaling with this factor results in an underestimated capacity profile, because shape anomalies of the wind speed histograms are amplified by the cubic section of the power curve Eq. (1.26). The best agreement for modeled and empirical histograms of output power is achieved when the long time average values of wind speed are matched, the corresponding multipliers are listed in Table 1.2, second column. The apparent reason of this success is that such rescaling transforms the reanalysis histograms in a way that the best agreement is obtained for the modeled and measured wind speeds over the sub-plateau range of the power curve (0–13 m/s interval).

The matching factors η^{**} used in the turbine - reanalysis comparison (Table 1.2, second column) are consistent with Hellmann exponent values

$\alpha \approx 0.23 - 0.37$ or with surface roughness length values in the range of $z_0 \approx 0.3 - 1.6$ m. These values are also consistent with expectations.

An important consequence of the considerations above is that a uniform scaling factor cannot provide a satisfactory approximation of high altitude wind speeds for each location, irrespective of the geographic conditions. It is very probable that the constant factor of $\eta = 1.28$ used in Section 1.7 seriously underestimates speed values over land, while it is too high for offshore locations. The main problem is that changing wind speed probabilities in the range of most frequent moderate values (2–12 m/s) affect drastically an estimated capacity factor by cubic amplification, see Eq. (1.26) and Fig. 1.37. On the other hand, the present analysis confirms that the key parameter of wind power modeling is the long time average wind speed at potential hub heights, in agreement with Archer and Jacobson [4].

In spite of the difficulties described above, we think that reanalysis wind data can provide reliable wind power estimates for extended geographic regions after proper parameter matching. And even when such accurate matching is impossible, a great part of the wind power statistics can be extracted from the reanalysis wind fields.

1.7 Wind power networks

1.7.1 Wind energy production in a fully connected static network

In order to produce reference data for further analyzes, we generated hypothetical power time series by the two steps described in Section 1.6.1 (see also [81, 82]). Figure 1.43 illustrates the long time average value and the coefficient of variation for the locations where wind turbines might be practically installed. Possible off-shore sites are simulated by an envelope of 1° around the coastlines (Iceland and the Faroe Islands are not considered). The total number of geographic cells is 1325 over this restricted area. Figure 1.43 indicates the well known fact that the best locations for wind energy extraction are off-shore sites near the Atlantic ocean [83, 84]. Solid surfaces do not only moderate wind speeds but increase the variability of energy production over land as well (Figure 1.43b).

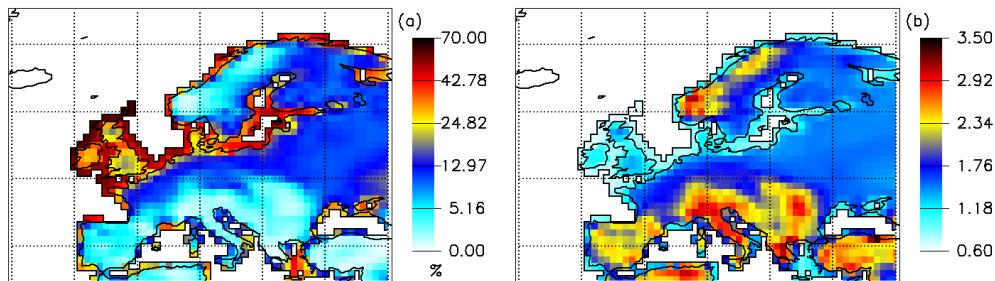


Figure 1.43: (a) Average wind power in units of percentage capacity factor (logarithmic scale), and (b) coefficient of variation (standard deviation over average value) for each individual geographic cell and for the whole temporal period of 44 years (linear scale).

Next we can imagine that the whole area depicted in Figure 1.43 is connected by a very efficient electric network, where fluctuations do not limit the integration of wind energy. This assumption is far from reality, however our aim is to evaluate a reference situation of no losses. Figure 1.44a shows the time series of fully aggregated energy production over the 1325 sites in units of total capacity factor (this is a convenient measure allowing any values for rated power at different locations). The fluctuations are highly non-Gaussian (see Figure 1.44b), however we could not find a simple analytical form for an adequate fit. The long time average value for the aggregated output is 14.4 % (with a standard deviation of 6.8 %), which is considerably lower than the sustained average capacity factor at optimal off-shore sites [73, 74].

The time series in Figure 1.44a exhibits a marked annual periodicity and a much weaker daily cycle revealed by the power spectrum in Figure 1.44c. The total aggregated output never drops to absolute zero, nevertheless it is less than 10 % of the total rated power for 104-105 days in a year. It is also remarkable that the absolute minimum is as low as 1.15 %, and the absolute maximum does not exceed 53.05 % during the period of 44 years. These facts clearly indicate that a fully connected static network over the whole continental Europe is highly suboptimal with huge idle capacity and strong seasonality.

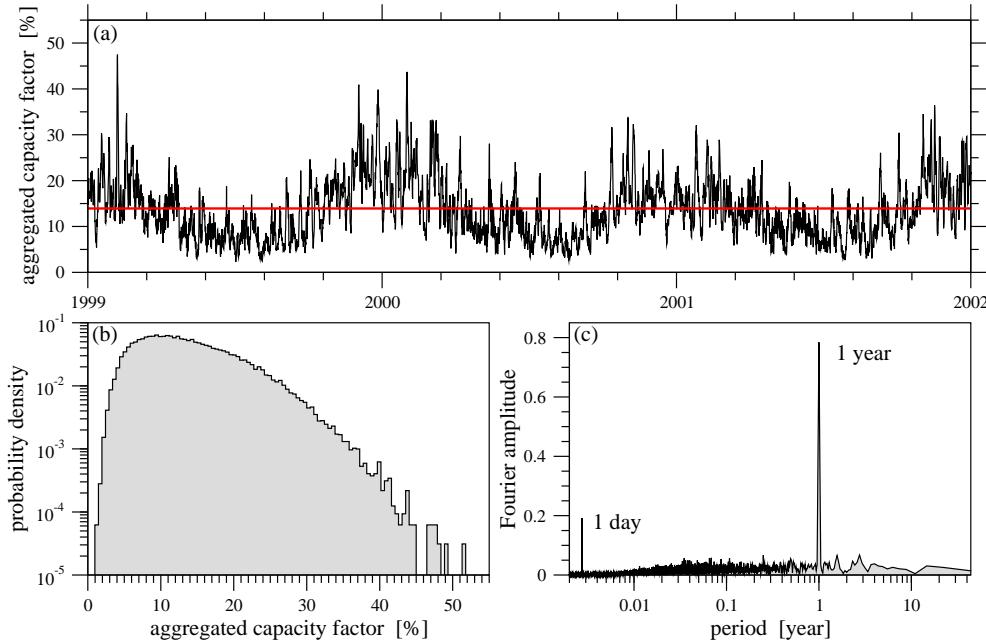


Figure 1.44: (a) Time series of aggregated wind energy production over 1325 sites (only years 1999-2001 are shown). The 44 year average capacity factor is 14.4 % (red line). (b) Normalized amplitude distribution of the time series. (The vertical scale is logarithmic.) (c) Power spectrum of the time series as a function of period. The two isolated peaks indicating a strong annual and a weak daily periodic background are marked.

Figure 1.44c might suggest that the annual periodicity is remarkably strong which could be exploited for prediction and output scheduling. Figure 1.45 illustrates, however, that this is not really feasible. Lower ($p_{tot} < 10\%$) or higher ($p_{tot} > 20\%$) than average aggregated output can happen in any calendar day, however with much higher probability during the Summer (see Figure 1.45a) or Winter months (Figure 1.45b), respectively. Note also that the frequency for most calendar days is well below the value of 44, which

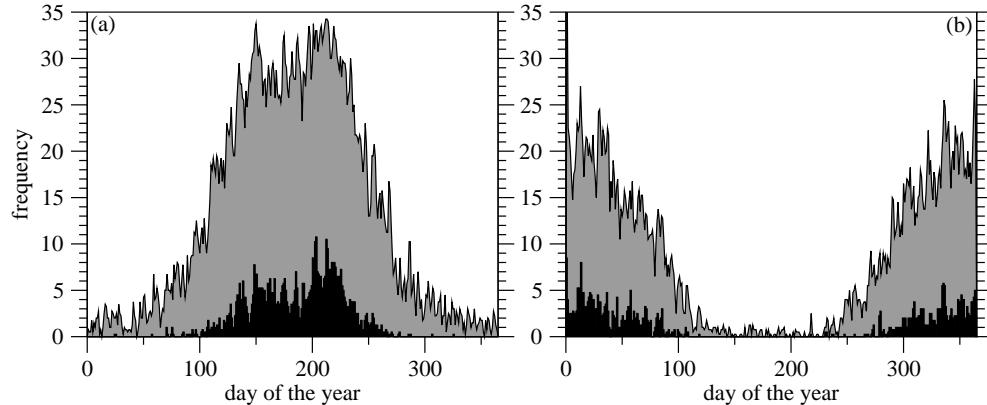


Figure 1.45: (a) 44 year frequency of calendar days with less than 10 % (grey) and less than 5 % (black) aggregated output. (b) The same as (a) for days with more than 20 % (grey) and more than 30 % (black) output.

means that any statistical forecasting has strongly limited efficiency.

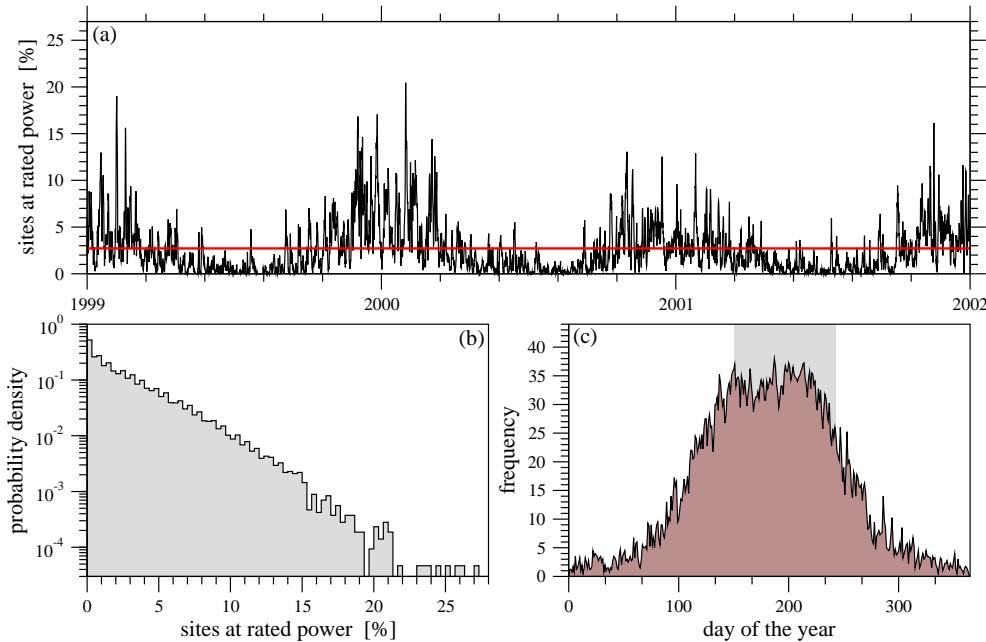


Figure 1.46: (a) Percentage ratio of sites at rated power (capacity factor $p \geq 100\%$) as a function of time (only years 1999-2001 are shown). The average value of 2.7 % is indicated with a red line. (b) Probability density distribution for (a). (c) Frequency of calendar days with less than 1 % site ratio (less than 14 cells) at rated power. Light grey shading indicates the summer period.

Essentially the same conclusions can be drawn from Figure 1.46, where the statistics of the sites at instantaneous rated power is analyzed. The rated power plateau (see Figure 1.37) is the optimal operating range for turbines especially with respect to electric grid integration. Figure 1.46a shows the time series (6 h resolution) of the number of geographic cells at peak output (normalized by the total number 1325), where hub height wind speed exceeds 12 m/s. The 44 year average value is only 2.7 % (36 cells) with the absolute maximum of 361 sites (27.2 %). Figure 1.46b illustrates the probability distribution for a given percentage ratio. Remarkable features are the very high probabilities for extremely low values and the fast, near exponential decay away from the origin. Figure 1.46c illustrates the smeared annual periodicity again, the shape of the histogram is similar to Figure 1.45a.

1.7.2 Energy production in limited area static networks

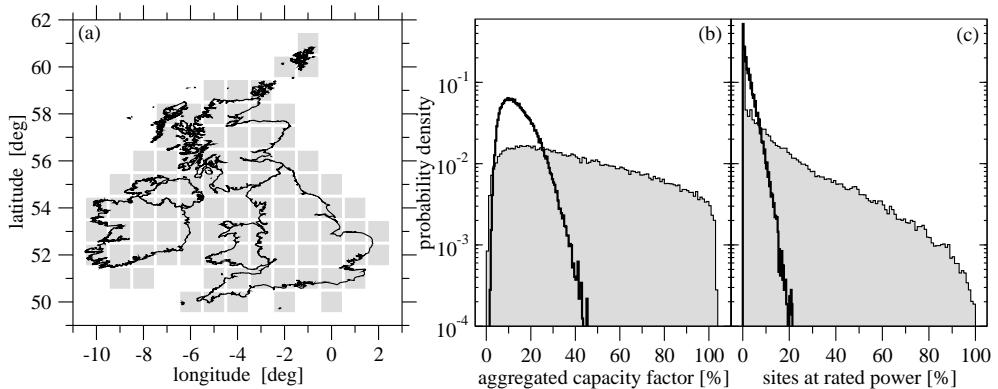


Figure 1.47: (a) Geographic setting of the analyzed area. (b) Probability density distribution for the aggregated capacity factor around Great Britain (grey histogram). Black line shows the whole-European curve of Figure 1.44b. (c) Probability density distribution for the percentage ratio of sites at rated power (grey), the black curve is the same as in Figure 1.46b.

The suboptimal nature of whole-continent static integration is clearly demonstrated by considering the time series produced by aggregation over a restricted area of high-wind locations. Figure 1.43 shows that the best sites are around Great Britain with respect to long term average values. Indeed, the integrated wind power over the shaded cells of Figure 1.47a obeys better statistics than shown for the whole continent in Figure 1.44 - 1.46: the long time aggregated average output is $41.0(\pm 25.9)$ %. Both the aggregated capacity factor (Figure 1.47b) and the percentage ratio of sites at rated power

(Figure 1.47c) have a wider probability distribution of much slower decay toward higher values.

The geographic cells can be easily ranked by individually determining the temporal ratio of operation at rated power for the period of 44 years. As an example, we show the best 50 sites around Europe in Figure 1.48b. (No wonder that the absolute majority is around Great Britain and Ireland.) We can imagine that these best cells serve as a static base network, and we can easily evaluate the aggregated output. The long time average value for such network is 53.8 % with a standard deviation of 24.3 %. The annual periodicity is quite strong, the absolute minimum value was obtained as 1.68 % in the Summer of 1966. The histogram of the capacity factor (Figure 1.48c) is much “nicer” than for Great Britain, the hypothetic connection with distant locations effectively eliminated the high probability of close to zero aggregated output.

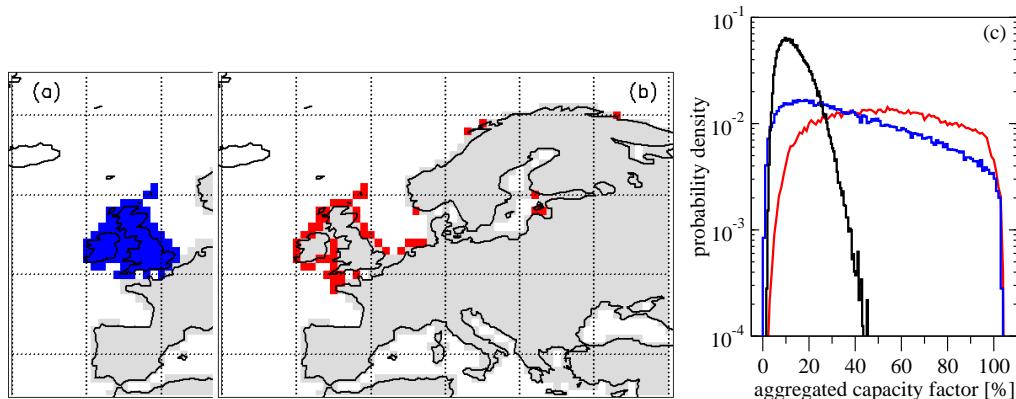


Figure 1.48: (a) The same as Figure 1.47a shown for comparison. (b) “Base network” formed by the 50 best wind sites over the continent (red cells). (c) Probability density distribution of the aggregated capacity factor for the base network shown in the map (red curve), for Great Britain (blue curve, Figure 1.47b) and for the whole continent (black curve, Figure 1.44b).

Nevertheless a marked negative consequence of the limited area spatial integration is an increased standard deviation around the mean aggregated output compared to the whole-continent integration. This means also much stronger 6-hourly fluctuations illustrated in Figure 1.49, where the empirical probability distributions for the 6-hour differences of the aggregated output are compared for Great Britain (dark gray histogram), for the optimal base network in Figure 1.48b (light gray curve) and for the whole continent (black histogram).

The large difference between the histograms in Figure 1.49 is a conse-

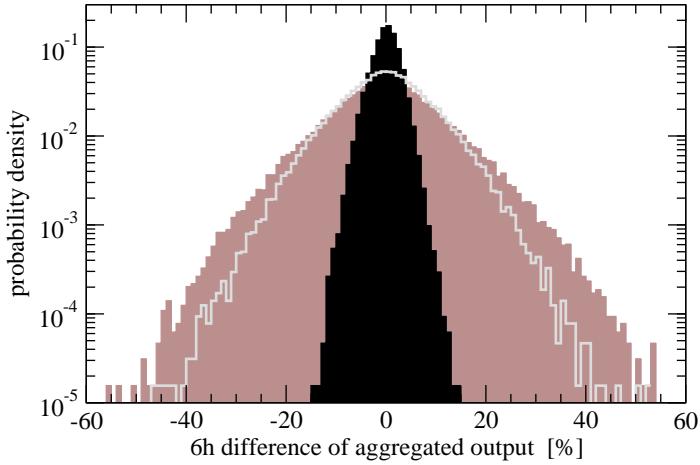


Figure 1.49: Probability density distribution for the 6-hour differences of the aggregated output over Great Britain (brown histogram), for the base network (light gray curve) and over the whole continental Europe (black).

quence of weak temporal and strong spatial correlations of wind fields. The autocorrelation function for a simulated wind power time series $p(t)$ at each geographic cell was found to exhibit an exponential decay with good approximation:

$$A(\tau) = \frac{\langle [p(t) - \bar{p}][p(t + \tau) - \bar{p}] \rangle}{\sigma^2} \propto \exp\left(-\frac{\tau}{T^*}\right), \quad (1.28)$$

where τ is the time lag, \bar{p} and σ denote the long time average value and standard deviation of $p(t)$, and T^* is the characteristic decay time. Figure 1.50b shows the geographic distribution of T^* . The characterization of spatial correlations is not so trivial, because the orography usually generates strong anisotropies depending on several other factors. The spatial resolution of ERA-40 data is rather limited, therefore we adopted the simplest possible way to visualize such correlations through an averaged length scale L^* . Firstly, the linear correlation coefficient

$$r_{i,j} = \frac{\langle [p_i(t) - \bar{p}_i][p_j(t) - \bar{p}_j] \rangle}{\sigma_i \sigma_j} \quad (1.29)$$

was determined for each pair of cells separated by maximum of $\pm 5^\circ$ in both latitude and longitude directions. The distance $d_{i,j}$, was computed by correcting curvature effects for each pair. Secondly, averaging of $r_{i,j}$ values was performed separately in bins of equal distance $d_{i,j}$, in order to minimize the effect of different sample sizes at different distances. Obviously, such averaging removes all directional anisotropies, apart from the side bands of the map where cells have less than 5 neighbors in one or two directions. The decay

of averaged correlation coefficient could be reasonably fitted by exponential functions again:

$$\overline{r}_{i,j} \propto \exp\left(-\frac{d_{i,j}}{L^*}\right). \quad (1.30)$$

The geographic distribution of the averaged correlation length scale L^* is shown in Figure 1.50a.

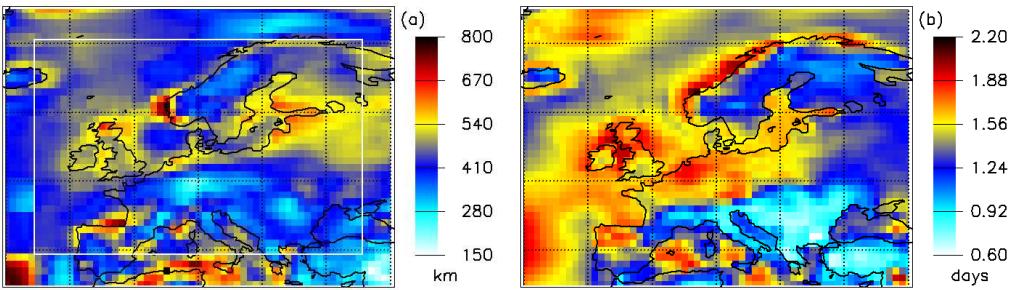


Figure 1.50: (a) Average correlation length L^ (see Equation (1.30)). A white frame indicates the area outside of which isotropic averaging was not possible due to a lack of data. (b) Autocorrelation time T^* (see Equation (1.28)) for each geographic cell obtained for the whole period of 44 years.*

Next we attempt to optimize hypothetical network architectures in order to produce better aggregated outputs than the reference case of whole-scale static integration. Practically, this means as high average capacity factor together with as low fluctuations as possible. Technical issues are out of scope of our conceptual study, we concentrate on the limitations arising from the statistical and correlation properties of the wind resource.

1.7.3 Output Variability in Growing Static Networks

As it is demonstrated in the previous Section by Figure 1.49, a larger area of wind power integration exhibits decreased fluctuations around the mean. How much improvement can we expect by gradually increasing the size of the network? One possible measure is illustrated in Figure 1.51, where the test was constructed in the following way.

The simulated wind power time series for a given number ($n = 1, \dots, 1324$) of randomly chosen cells were aggregated at each time step, and the long term average $\overline{p_{tot}}$ and its standard deviation σ were computed, as before. At each network size n , 30 different random realizations were evaluated and ensemble averages $\langle \overline{p_{tot}} \rangle$ and $\langle \sigma \rangle$ were determined. The results are plotted in Figure 1.51, where the number n was transformed to “total network area” for correcting Earth’s surface curvature. Figure 1.51a illustrates that the

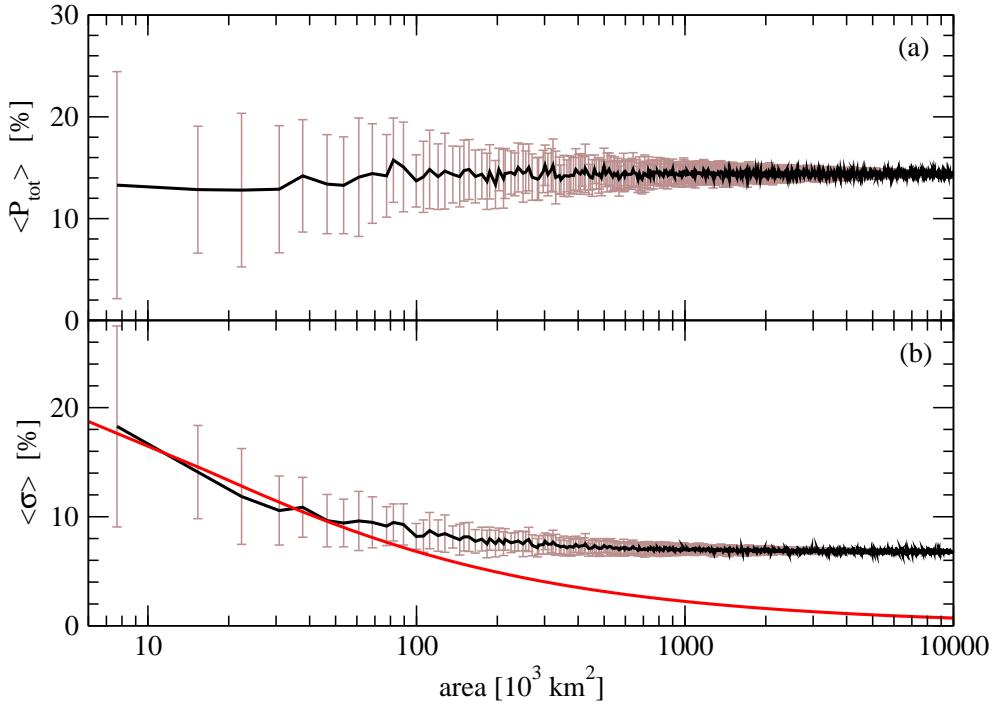


Figure 1.51: (a) Ensemble and time averaged aggregated capacity factor $\langle \bar{P}_{\text{tot}} \rangle$ for 30 samples of randomly connected cells of increasing number. The number of connected cells is transformed to total integrated area on the horizontal axes. The error bars represent standard deviation computed for the 30 realizations. (b) Averaged standard deviation $\langle \sigma \rangle$ around the mean $\langle \bar{P}_{\text{tot}} \rangle$ for the same test. The red dashed line illustrates an inverse square root function.

ensemble averaged capacity factor is essentially independent of the network size, the average for 30 random single cells is practically the same as for the whole-scale integrated value. The behavior of the ensemble averaged standard deviation shown in Figure 1.51b is more remarkable. In case the aggregated power time series represented statistically independent random variables, this parameter would decay according to an inverse square root law, illustrated by the red line in Figure 1.51b. The empirical data indicate marked deviation from this expectation for random network sizes of 15-17 cells already, which is a consequence of the rather strong spatial correlations (see Figure 1.50b).

1.7.4 Static network with different capacity limits

So far all available geographical locations were treated uniformly, that is the same capacity was assigned to each cell in the hypothetical wind power

networks. A further optimization strategy might be based on distributing available capacities unevenly to minimize the temporal variability of wind power. One motivation for testing such possibility is that the computed linear correlation coefficient $r_{i,j}$, (see Equation (1.29)) was found to have slight negative values for some distant pairs, indicating the presence of temporal anti-correlations. The number of different network configurations with different sizes and weights is practically infinite, therefore we illustrate the method only for the whole-scale connected case.

The procedure commences by assigning initial weight parameters w_i to every available grid-point i , representing the level of installed capacity at the given cell. In order to compare different configurations, normalization is needed: $\sum_i w_i = 1$, where $i = 1, \dots, 1325$ (the total capacity is fixed), consequently $0 \leq w_i \leq 1$. (The uniform setting $i = 1/1325$ reproduces the results discussed above.) The optimization task is to modify the set of weights $\{w_i\}$ in order to get less intermittent aggregated output. This problem is a linearly constrained global optimization in 1325 dimensions [85]. We implemented an iterative Monte Carlo algorithm, where in one step a randomly chosen weight changes stochastically. In order to fulfill the normalization constraint, two other randomly selected weights are modified. If the resulting configuration provides a smaller coefficient of variation for the aggregated power time series, then the set $\{w_i\}$ is updated and the iteration proceeds. This algorithm is not necessarily convergent and does not guarantee to find the global optimum, nevertheless it is able to produce moderate improvements even for the fully connected case. An example is shown in Figure 1.52. Here the initial configuration was the uniform weight distribution. The average aggregated capacity factor increased to 14.9 % (from 14.4 %) with the standard deviation of 4.2 % (6.8 % in the uniform case), which means an improvement for the coefficient of variation to 0.28 (from 0.47). In spite of this apparent success, the configuration shown in Figure 1.52 hardly represents a serious advance. A simple comparison with the map in Figure 1.43a reveals that this network is based on a few highly efficient off-shore sites at the Atlantic coast connected to huge idle capacities located mostly at extremely low-wind cells in the opposite corner of the continent, which exhibit weak temporal anti-correlations.

1.7.5 Static network based on real wind farm data

Based on real wind farm data it was possible to evaluate a network that is realistic in a sense that the wind power capacities correspond to actual

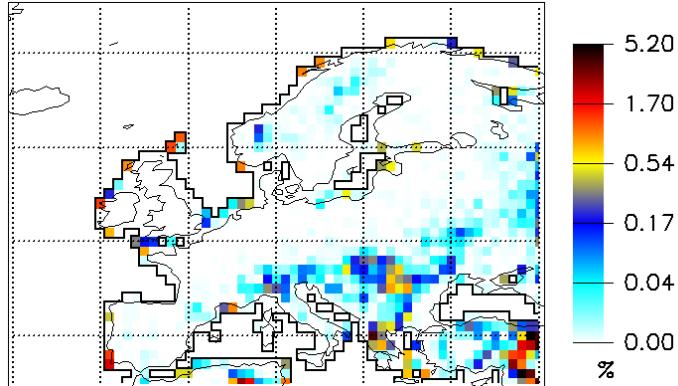


Figure 1.52: Stochastically optimized global network configuration with different weight parameters for the cells. Percentage values of the final weights $\{w_i\}$ are indicated.

data (for October 2007).⁶ The distribution of the capacities is illustrated in Figure 1.53. Here, like in all other calculations throughout this work, we did not take transportation losses into consideration. Therefore the statistics for this “real” network are somewhat optimistic. Only limitations from the wind resource are present in the results.

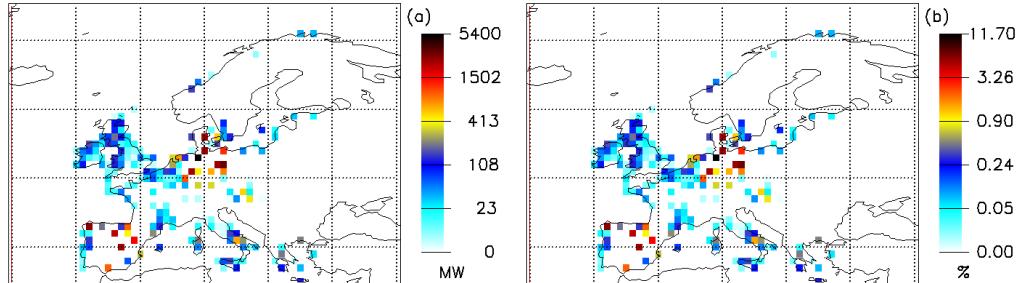


Figure 1.53: Global wind power network configuration based on real data corresponding to the situation in October 2007. (a) Operating nameplate capacities in MW, (b) weights $\{w_i\}$ in percentage of total capacity. The total capacity is 46.028 GW. Logarithmic scales are used.

The time series of the aggregated output of this “real” network can be seen in Figure 1.54. Comparing the three configurations of Figure 1.54 we can see that all have about the same average capacity factor (14.4 – 14.9 %) but the “real” network shows the biggest fluctuations (standard deviation of 12.4 %).

⁶Here I would like to thank Flavio Bono and Eugenio Gutierrez at the Joint Research Centre (JRC), Ispra, Italy to kindly provide us the data.

This might be because, as it is visible in Figure 1.53, the main capacities are concentrated in a relatively small area where wind speed values are strongly correlated. The fully connected uniform reference network and the optimized network of Figure 1.52 show smaller fluctuations (standard deviation of 6.8 % and 4.2 % respectively). However wind statistics is by far not the only aspect when designing a real wind power farm or network. Several other issues (technical, economical and political) need to be considered which are out of the scope of the present study. Here we restrict ourselves to wind power statistics.

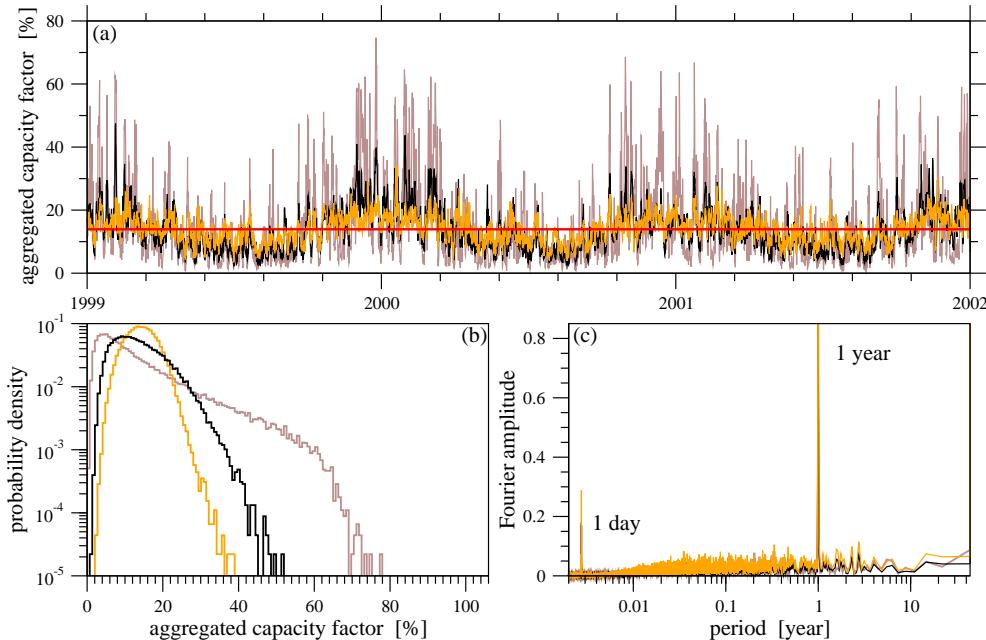


Figure 1.54: (a) Time series of aggregated wind energy production for the “real” network (brown curve). The time series corresponding to the fully connected network is shown for comparison with black color. The data for the optimal configuration of Section 1.7.4 (Figure 1.52) are shown as well with yellow color. The 44 year average capacity factor is $\sim 14\%$ (red line) for all 3 time series. Only years 1999-2001 are plotted. (b) Normalized amplitude distribution of the time series. (The vertical scale is logarithmic.) (c) Power spectrum of the time series as a function of period. The two isolated peaks indicating a strong annual and a weak daily periodic background are marked.

In the case of the realistic network (Figure 1.53) we tested the effects of changing the conversion factor η between surface and hub height wind speeds. We used 4 configurations: (1) $\eta = 1.28$, (2) $\eta = 1.50$, (3) $\eta = 1.79$ uniformly for all locations and (4) $\eta = 1.15$ for maritime locations and

$\eta = 1.79$ for onshore locations. Configuration (1) is discussed throughout Section 1.7, while configuration (3) is inspired by the Hungarian case study of Section 1.6.3. Configurations (2) and (4) are supposedly more realistic “middle paths” between the former ones.

The average and standard deviation of the power production in each configuration based on the ERA-40 data is summarized in Table 1.3, while the distributions are shown in Figure 1.55. For a comparison with real power production data see Table 1.4.

η	\bar{p} [%]	σ_p [%]
1.28	14.36	12.70
1.50	20.75	15.44
1.79	29.27	17.99
1.15 – 1.79	28.28	17.83

Table 1.3: The average \bar{p} and standard deviation σ_p of the aggregated power output time series (in capacity factor) for the realistic network depicted in Figure 1.53. ERA-40 data were used to generate the time series with different conversion factors η between surface and hub height wind speeds.

country	\bar{p} [%]
Germany	19.7
Spain	20.5
Italy	16.7
France	18.9
United Kingdom	28.2
Denmark	26.3
Portugal	21.2

Table 1.4: The average capacity factor \bar{p} for some European countries based on real measurements for the year 2007 [86, 87]. Compare the values with those of Table 1.3.

It can be seen, that not only the average capacity factor increases with increasing η , but the shape of the distributions in Figure 1.55 changes as well. This is due to the nonlinear transformation between wind speed and power (see Equation (1.26) and Figure 1.37). From the comparison with real measurements (Table 1.4) we can conclude again that the $\eta = 1.28$ assumption used throughout Section 1.7 probably underestimates wind power production and the $\eta = 1.79$ scaling derived from Hungarian data probably overestimates the production. The truly realistic configuration might be

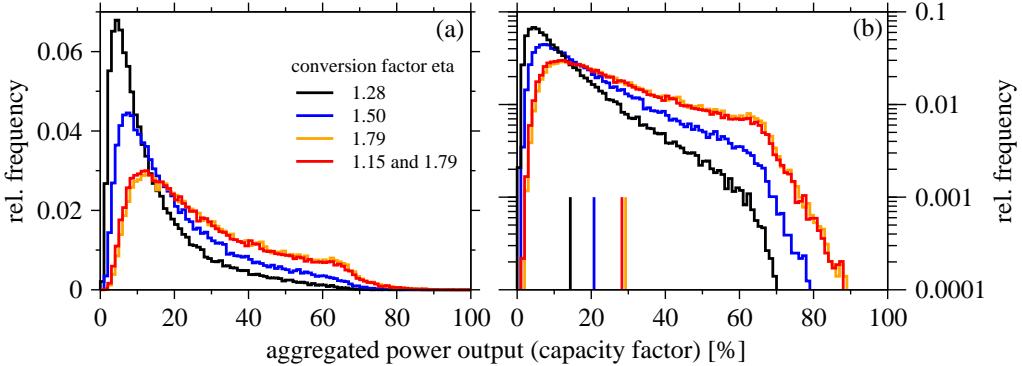


Figure 1.55: Histograms of the aggregated wind energy production for the realistic network depicted in Figure 1.53. Different conversion factors η were used to scale surface wind speeds to hub height (see the legend). (a) Linear and (b) logarithmic scales were used. The average values given in Table 1.3 are indicated with vertical lines in (b).

somewhere in between. However the use of a uniform scaling factor η is highly unrealistic.

In spite of these we believe, that our conclusions concerning the variability of the aggregated wind power production are still qualitatively valid. Of course quantitative differences are possible from a real world scenario, mainly because of the nonlinearity of the speed-to-power transformation (Equation 1.26), which amplifies the errors and biases of wind speed estimations.

1.7.6 Network “disintegration” test for wind field persistence

We demonstrated in the previous Sections that even a full-scale static integration of wind power remains suboptimal, because the global statistics is negatively influenced by many continental low-wind sites where installing turbines is strongly redundant. Before we continue the search for more optimal networking solutions, we would like to demonstrate the role of short-range temporal correlations in integrated wind energy production.

The network “disintegration” test is based on the idea that we have a fully connected grid over all the 1325 cells equipped by a control center where all the cells are dispatchable. At a given time instant t_n , the first task is to connect each site where the instantaneous output $p(t_n)$ exceeds a critical capacity factor p_c . The quantity of interest in this test is the total number of sites $N(t_n; p_c)$ where this condition holds. In the next time step t_{n+1} , the control center simply disconnects every cell, where the output

drops below the threshold p_c , and the number of remaining sites $N(t_{n+1}; p_c)$ is recorded. Since no new connections are considered in this process, the result of subsequent iterations is a monotonously decreasing time series of site numbers, until all the cells are switched off in a finite time interval. Obviously, such finite time series are very sensitive to the initial state at t_n and the later temporal evolution of the wind field, therefore an averaging is performed over the total length of ERA-40 records by setting the initial time to t_1, t_2, t_3, \dots (the total record lengths is 64240 data points). The results for different threshold values are shown in Figure 1.56.

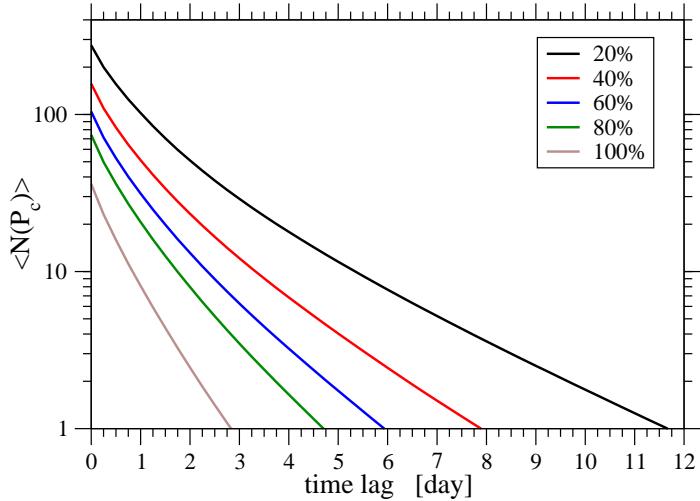


Figure 1.56: Decay of the average number of cells $\langle N(t; p_c) \rangle$ continuously exceeding a given threshold capacity factor p_c as a function of time lag from an initial instant. Results for different p_c values (see legends) are shown. The vertical scale is logarithmic.

For a given threshold level p_c , the initial value for $\langle N(t; p_c) \rangle$ is the long time average over the whole period of 44 years (e.g. on average 36 cells out of 1325 operate at $p_c \geq 100\%$, see also Figure 1.46a). Figure 1.56 indicates that the decay is very fast in the beginning, then the tails are approximately exponential functions. The surprising fact is that the expected number of “operating” cells drops very quickly to 1 even when the threshold capacity factor is only 20 % (less than 12 days for this case). This statistics clearly demonstrates how short the persistence time of the wind field is over the continent.

1.7.7 Energy production with full dynamic control

Although most of the basic statistical information about wind power availability over Europe is included in the model time series depicted in Figure 1.44 and Figure 1.45, we demonstrate its limitations in another test by using the fully connected dispatchable network architecture with dynamic control. The target is to achieve a smooth aggregated power generation at a modest level which corresponds to the total rated power of 50 cells out of the available 1325 (3.77 % of total aggregated capacity).

The best 50 sites shown in Figure 1.48b serve as the base network. The control center continuously evaluates the state of all the 1325 cells. When the output of a given operating site drops below the level of rated power, it is immediately disconnected from the grid, and replaced by another site working at rated power. When such simple replacement is not possible (global low-wind situation), the control center connects as many sites as necessary (ranked by the level of instantaneous potential output) to keep the aggregated power at the target level. This strategy is essentially equivalent to the well known “spinning reserve” mechanism, apart from the fact that here the dynamic control is required due to the fluctuating resource instead of the fluctuating demand.

The results in Figure 1.57 demonstrate that even such a low target level cannot be sustained without serious breaks. It happens regularly during the summer months that the aggregated output of all the available cells is less than the required power. Figure 1.57b illustrates that the periodically supervening global low-wind situation is characterized also by the extremely high number of cells (close to 33 %) where the wind speed remains below the cut-in value.

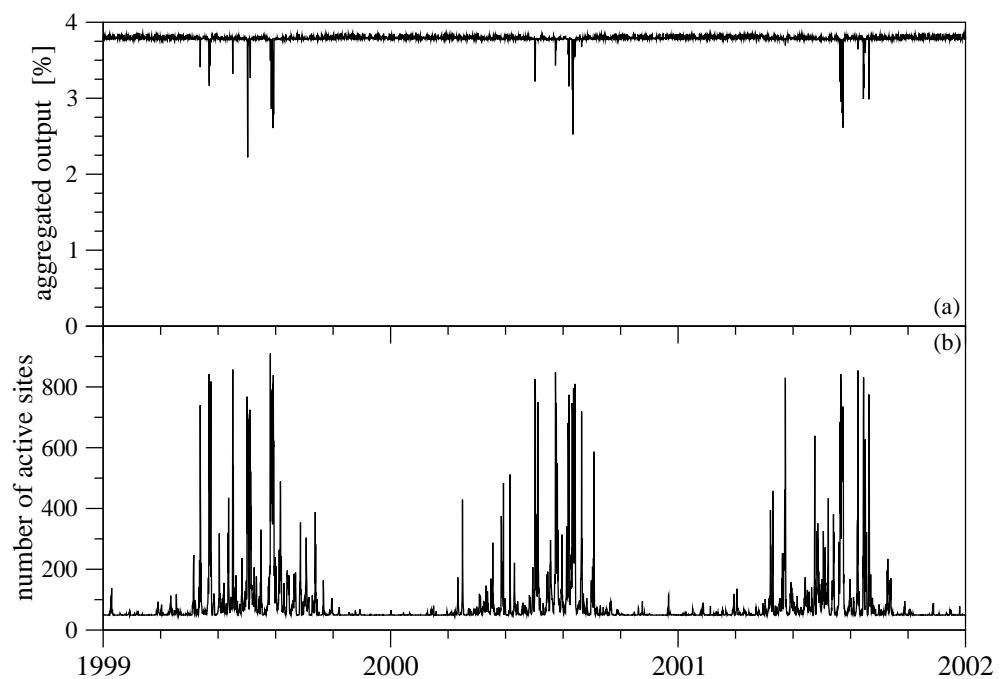


Figure 1.57: (a) Aggregated output of a dynamically controlled, fully integrated grid expressed as percentage ratio of the total installed capacity. The target level is 50 cells at rated power (3.77 %). (b) Number of simultaneously operating cells. Only years 1999–2001 are shown.

1.7.8 Matching demand and wind power supply – a case study

In this section, we compare the electricity demand of a large consumer (a big factory, time series C , see Section 1.2.2) with a hypothetical supply that is based on wind power exclusively (see also [88]). The measured consumption pattern is composed of characteristic daily, weekly and seasonal cycles following the usual rhythm of human activities (see Figure 1.5). In contrast, wind power availability is determined by the meteorological circumstances and therefore very intermittent. Our analysis is entirely based on high frequency measurements in the factory (power consumption), and at two distant locations of wind turbines (power production) in Hungary (time series M_1 , M_2 , K see Section 1.2.2). The hypothetical situation is that the factory must switch to electricity provided by wind farms at either one or the other location, furthermore we consider the case when wind electricity is integrated in a common grid.

The main findings are not very different from what one might expect, however they are quantitative. They also support the results of the previous sections. The integration from the two sources is not enough for a continuous supply, even when the wind power capacity is infinitely larger than the average consumption. This is because the area of Hungary is very small, therefore it is quite common that the wind speed nowhere exceeds the cut-in value for some period of time. We demonstrate that scaling up wind power capacities results in a slower than linear improvement in the supply at the cost of an increasing fraction of excess wind electricity. Interestingly, the distribution of length of supply and shortfall intervals exhibits a power-law (scale-free) behavior, and possible consequences are discussed. An important result is that the statistics are practically the same for cyclic and constant average loads, which makes subsequent availability studies much easier.

We used the instantaneous wind power production data of turbines M_1 , M_2 and K and the consumption data C (see Section 1.2.2). The location of the turbines and the time-line of available time series can be seen in Figure 1.58 and more generally in Figure 1.4.

Due to geographic constraints in the Carpathian basin, Hungary is not very rich in wind energy [71, 76, 77, 79, 80, 89], and electricity production is extremely intermittent at each site (see Figure 1.59). Typical capacity factors at working turbines hardly exceed 25%, see Table 1.5.

As a consequence of the small geographic distance between the K and M turbine sites (156 km), one cannot expect the independence of wind speeds. Indeed, the typical correlation length was found around 200–300 km for wind power in Section 1.7 based on ERA-40 reanalysis data. Figure 1.59c clearly

turbine	\bar{p} [%]			σ_p [%]		
K	19.92	17.42	20.89	23.71	22.61	24.98
$M1$	20.76	22.33	20.27	11.55	12.58	11.24
$M2$	21.26	23.16	21.55	11.66	12.24	11.58

Table 1.5: The average \bar{p} and standard deviation σ_p of the Hungarian turbines K , $M1$, and $M2$ are given (in capacity factor) for three consecutive years.

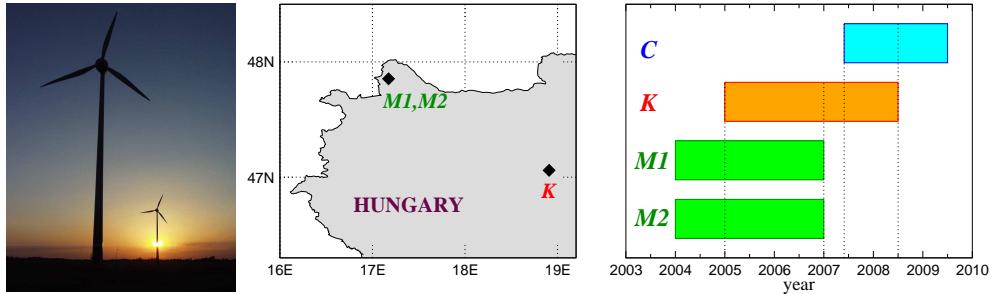


Figure 1.58: Sketch of the geographic setting of the wind turbines and timeline of the records. Heavy diamonds indicate the location of three Enercon E-40 wind turbines $M1$, $M2$ (near Mosonszolnok, $47.816^\circ N$, $17.174^\circ E$) and K (Kulcs, $47.057^\circ N$, $18.914^\circ E$). The time-line illustrates the overlapping periods: 06/01/2007–06/30/2008 for C (consumer) and K records, and 01/01/2005–12/31/2006 for K and M time series. (Photograph: Sándor Záttonyi, <http://www.panoramio.com>)

illustrates that an integration of 10 min power output in a hypothetical common grid does not result in a drastic improvement considering intermittency. Power integration in the overlapping period of 2005 and 2006 (Figure 1.58) is performed by

$$P_{\Sigma}(t) = \frac{1}{2}P_K(t) + \frac{1}{4}P_{M1}(t) + \frac{1}{4}P_{M2}(t) \quad (1.31)$$

where we assumed that the two adjacent turbines $M1$ and $M2$ operate in a “wind farm” mode, thus the two *locations* have the same weights. Sporadic missing data ($< 0.3\%$) were replaced by zeros. Substantially better aggregated output is possible when much larger distances are considered connecting climatologically separated regions [90, 91].

Further relationship between the two turbine sites is revealed by the cross-correlation functions

$$X_{i,j}(\tau) = \frac{\langle (P_i(t + \tau) - \bar{P}_i)(P_j(t) - \bar{P}_j) \rangle_t}{\sigma_i \sigma_j} \quad (1.32)$$

where $i, j \in \{M1, M2, K\}$, τ is the time-shift, $P(t)$ denotes the instantaneous power of average value \bar{P} and standard deviation σ , and $\langle \cdot \rangle_t$ indicates temporal averaging. As Figure 1.60a clearly shows, the cross-correlation between time series K and $M1$ has a maximum at $\tau = -0.09$ day (≈ -2 hours), the curve is practically the same for K and $M2$ (not shown). This means that the wind speed variations at site M often determine the changes at site K occurring a few hours later. Precisely such behavior is expected by checking the maps of the prevailing winds in the territory [92]. The cross-correlation function for the adjacent turbines (Figure 1.60a, blue line) is centered at zero, symmetric, and almost identical with the individual autocorrelation functions given by Equation (1.32) with $i \equiv j$. The power spectrum provided by standard Fourier analysis for site K (Figure 1.60b) exhibits a weak periodic component around 1 day, which is not present neither at site M (not shown) nor in the aggregated record (Figure 1.60c). This might be connected with the neighboring river basin of Danube at Kulcs.

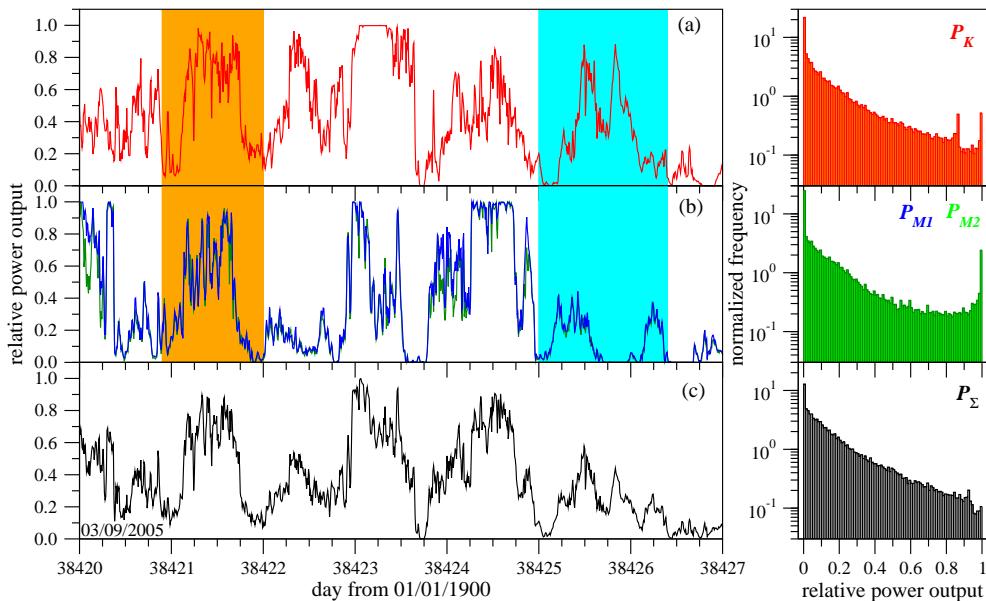


Figure 1.59: Relative power output (instantaneous measured value normalized by the measured peak power of 620 kW) for a period of one week beginning on 03/09/2005. (a) Turbine K , (b) turbines $M1$ (green) and $M2$ (blue), and (c) “integrated” power (see Equation (1.31)). Orange (cyan) shading indicates almost synchronous (counter-phase) production. The probability density distributions (normalized frequencies) are shown on the right side, for the full record lengths indicated in Figure 1.58.

Large energy consumers have the market weight to deal with electric com-

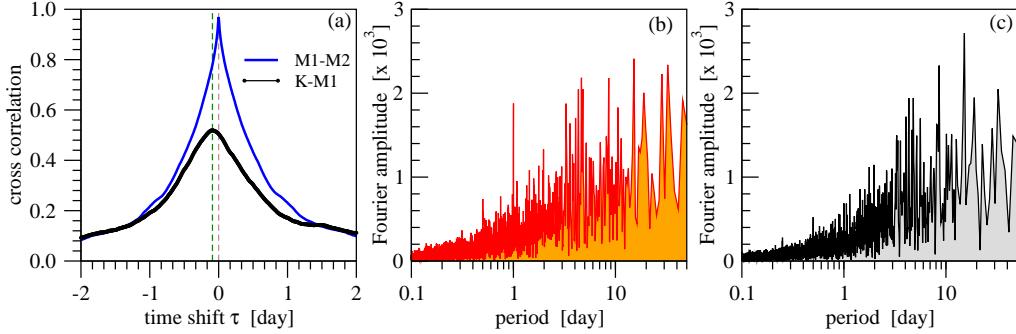


Figure 1.60: (a) Cross-correlation function (see Equation 1.32) for the power records K and $M1$ (black) and $M1$ and $M2$ (blue). Vertical dashed lines indicate the peak maxima. (b) Normalized Fourier amplitudes for the power record K . (c) The same for the integrated series shown in Figure 1.59c.

panies for customized pricing, therefore we cannot publish absolute numbers or details about the factory that kindly provided us high quality data on their power consumption. The total record length covers more than 2 years (Figure 1.4 and 1.58, record C) with temporal resolution of 15 minutes without a sole missing point. (This is because the instantaneous load is simultaneously measured by duplicated systems, as the accuracy is primary interest of the company.) The power consumption record for the calendar year 2008 is shown in Figure 1.5. The pattern is composed of highly regular daily and weekly cycles interrupted by short national holidays and two longer periods of decreased activity in the Summer and around Christmas.

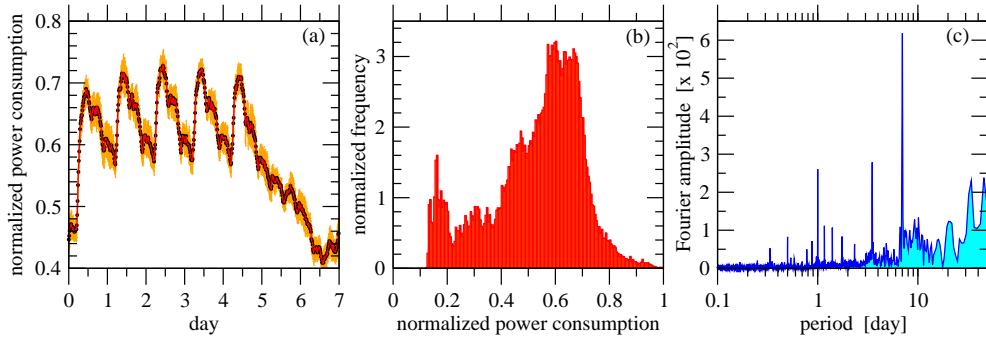


Figure 1.61: (a) Weekly power consumption pattern (black) extracted by averaging 9 regular working weeks in 2008 starting from day 39454 (Figure 1.5, middle). Red line shows the same curve with a temporal resolution of 10 min, orange is the standard deviation. (b) Histogram of the power consumption obtained from the full record. (c) Fourier analysis reveals the strong daily and weekly cycles, the other thin peaks are harmonics.

The overlap problem obviously visible in Figure 1.58 can be partially resolved by obtaining an average weekly pattern shown in Figure 1.61a, and use it repeatedly for subsequent evaluation. For the proper matching, the original curve of time resolution 15 min was resampled by linear interpolation to get 10 min data (Figure 1.61a, red line). A cursory comparison of the nontrivial peaked histogram of power consumption (Figure 1.61b) with the histograms of wind electricity production (Figure 1.59, right column) forecasts a poor fit for supply and demand. Similarly, the mismatch of Fourier spectra in Figure 1.61c and Figures 1.60b and 1.60c further decreases possible expectations.

Next we imagine a hypothetical situation where electricity is exclusively provided by wind farms installed at the test sites M and K . It is a highly unrealistic situation since nobody considers wind electricity alone as a working source of base-load supply. Nevertheless, the comparison yields quantitative information on the strength and properties of intermittency.

A key parameter of the analysis is the total wind power capacity P_{tot} . It is meaningless to “install” less capacity than it is required by the consumer in a full year. As we listed, typical capacity factors are around $f = 0.2$ for turbines in Hungary (see Table 1.5), thus the installed rated power P_{tot} must exceed either the peak or average consumption (C_{max} or C_{av}) by a factor of 5. A coefficient F characterizes the excess capacity through

$$P_{tot} = F \cdot \frac{C_{av}}{f} \quad (1.33)$$

such that $F = 1$ purports $f \cdot P_{tot} = C_{av}$.

Figure 1.62 illustrates how supply and demand fit in two windy months in 2007 for two values of factor F . Besides the intermittency, two aspects are remarkable. Firstly, when adequate supply is given by wind power, it is almost always in conjunction with considerable excess electricity production. Secondly, when the total rated power is increased by a factor of 2 (in the case of our consumer this means investing in many dozens of new turbines), the supply does not improve drastically; a number of white intervals remain which indicates electricity shortfall.

In order to characterize the improvement of supply by increasing the total rated capacity P_{tot} , we repeated the comparison for years 2005 and 2006 with the individual and aggregated wind power records and the consumption pattern formed by gluing the average weekly cycle (Figure 1.61a) as a function of factor F . The result is shown in Figure 1.63. The benefit of wind power aggregation from distant sites is clear, still the curve exhibits quick saturation for this case as well. The functional form for the total time of supply T_S :

$$T_S(F) = T_\infty [1 - a \cdot \exp(-b \cdot F^c)] \quad (1.34)$$

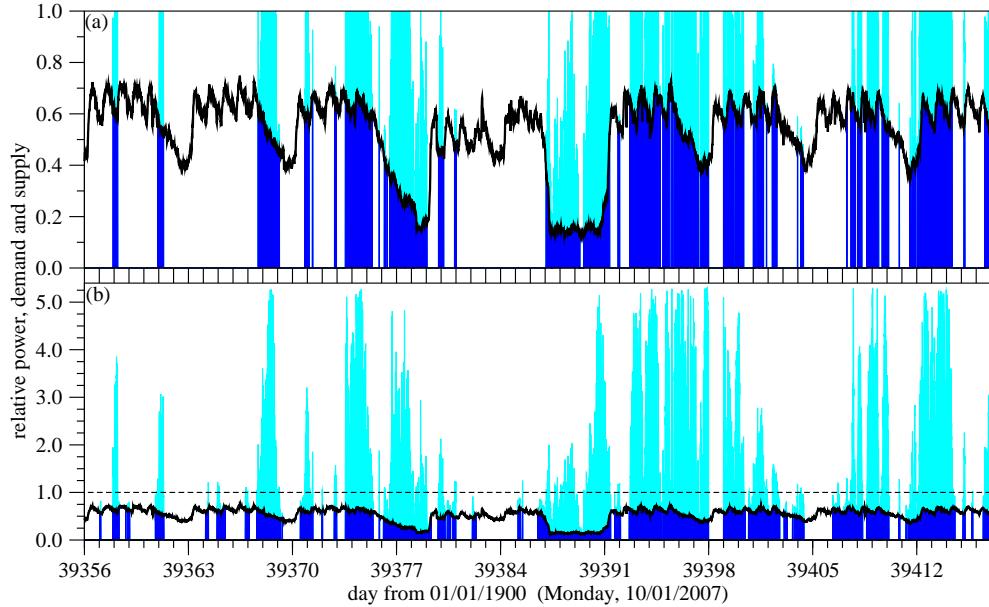


Figure 1.62: Matching of wind power supply (blue shading) with demand (black curve) in two full months (October–November, 2007) for wind power record P_K , see Figure 1.59a. Light blue shading indicates excess wind power. (a) $F = 1.0$, (b) $F = 2.0$ (see Equation (1.33)). The vertical scale is extended to illustrate excess production.

provides a good quality fit for each curve. We show only one for the aggregated wind power (Figure 1.63a, orange line) with fitted values $T_\infty = 87.57\%$, $a = 2.73$, $b = 1.49$, and $c = 0.393$. The asymptotic value $T_\infty < 100\%$ is a consequence of null periods when the wind speeds remain below cut-in value at both sites. Note that the supply level of $1/2$ ($T_S \approx 50\%$) requires around $F = 2$, which means 10 times larger installed capacity (with $f = 0.2$) than the mean consumption.

We have found that the aggregated output from two wind farms, designed to produce the same amount of electric power as the annual demand ($F = 1$), can provide adequate supply for 34% in a year. Two times larger total installed capacity ($F = 2$) increases this interval to be 52%, however any further increase has far less efficiency due to a stretched exponential saturation to the limiting value of 87%.

It is clear that a growing total rated capacity results in an increasing fraction of excess wind power E not required by the very consumer (see Figure 1.62b). Figure 1.63b illustrates this fraction as a function of factor F , the behavior is also stretched exponential according to Equation (1.34). The essential difference is that the asymptotic value (belonging to the limit

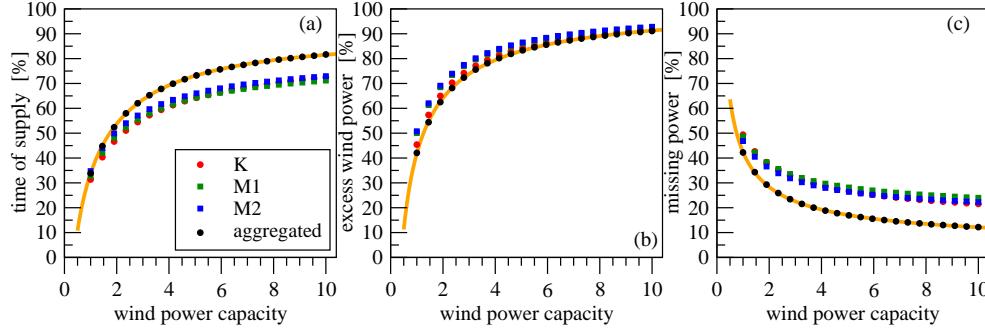


Figure 1.63: (a) Total time of supply T_S as a function of installed wind power capacity (expressed through factor F in Equation 1.33) for the individual and aggregated wind power records, see legends. 104 weeks with the pattern in Figure 1.63a are considered in the comparison. The orange line illustrates the fit by Equation (1.34). (b) Excess wind power E as a function of F . The orange line is a fit by Equation (1.34) (T is replaced by E). (c) Missing power P_{miss} in periods where wind energy production is nonzero, but remains below the demand. The orange line is a fit by Equation (1.35).

$P_{\text{tot}} \rightarrow \infty$) converges to the maximum $E_\infty = 100\%$ for each curve, the other parameters for the aggregated wind power record are $a = 15.98$, $b = 1.94$, and $c = 0.339$. This can be understood, because the consumption is finite, but nothing limits (mathematically) the installed capacity P_{tot} .

Finally, we evaluated the decreasing portion of periods when nonzero wind power—less than the demand—is produced. In such situations the missing power P_{miss} must be complemented from other sources. Limiting cases are the full supply periods with $P_{\text{miss}} \equiv 0$, and the opposite case when arbitrarily small wind power production is scaled up by F to fulfill the instantaneous requirement, thus P_{miss} is the total demand when the wind power is exactly zero. The orange line in Figure 1.63c indicates again a stretched exponential relaxation

$$P_{\text{miss}}(F) = P_{\text{miss},\infty} + a \cdot \exp(-b \cdot F^c) \quad (1.35)$$

with parameter values $P_{\text{miss},\infty} = 7.40\%$, $a = 1103.09$, $b = 3.45$, and $c = 0.197$. We emphasize that the functional forms Equations (1.34) and (1.35) serve only to estimate results out of the tested range of F , and a theoretical explanation is not intended.

Figure 1.64 exhibits curves almost identical to Figure 1.63, however we think that this is an important result. Here the consumption pattern is simply replaced by the constant average value, nevertheless the results remain very close to the above ones. To illustrate this fact, we show exactly the same fits obtained in Figure 1.63 (Figure 1.64, orange lines), the symbols

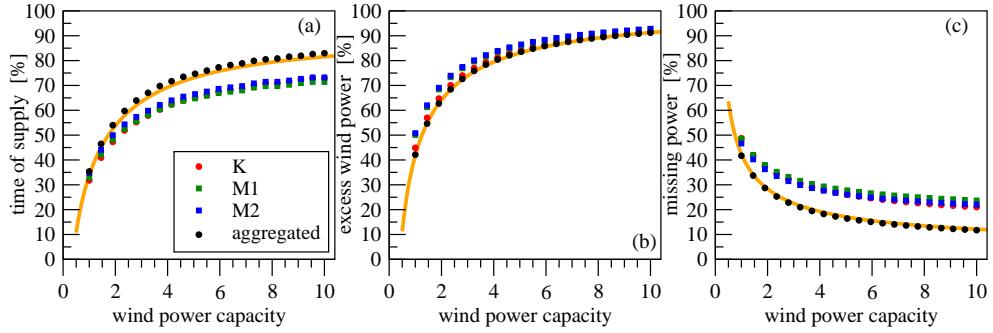


Figure 1.64: The same as Figure 1.63, by assuming a constant relative power consumption of 0.59. Note that the orange lines are not fitted here, there are identical with the ones in Figure 1.63.

belong to the repeated evaluation assuming constant consumption. Figure 1.62b demonstrates that the fluctuations of power consumption are much smaller than those of the production, therefore the former can be replaced by a constant value, indeed. Further studies can benefit a lot from this observation, because it seems that power consumption data with high resolution are not imperative to characterize intermittency of electricity production, when data of the resource are available.

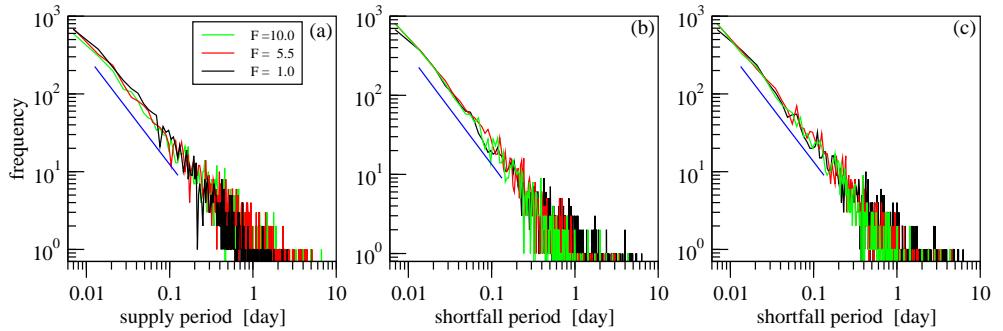


Figure 1.65: Length distribution of continuous periods of (a) supply, (b) shortfall, and (c) shortfall with constant consumption at three values of F (see legends) for wind power record P_K . (Note the double logarithmic scale.) Blue lines indicate a power-law with exponent value -1.4 .

The intermittent behavior of wind energy production is related to turbulence in the wind field. It is known for decades that the so called “level-crossing” statistics, i.e. the length distribution of time intervals above or below a given threshold value, has a power-law shape for horizontal wind speeds [93, 94]. In our analysis, a closely related statistics is the length

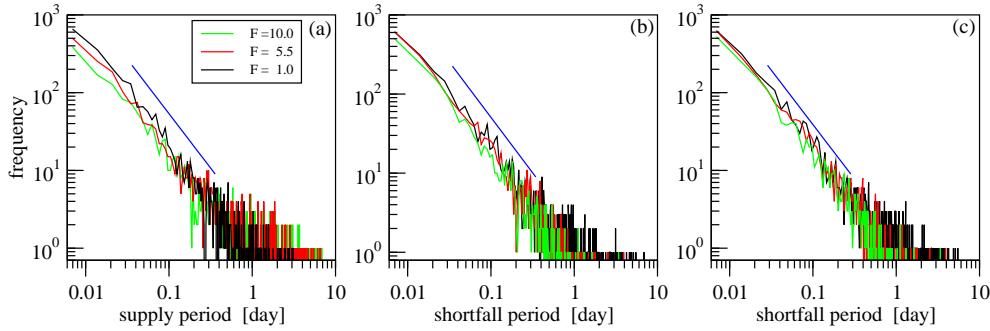


Figure 1.66: The same as Figure 1.65, for the aggregated wind power record P_Σ .

distribution of continuous time intervals of adequate supply L_{sup} or electricity shortfall L_{short} . Level-crossing statistics of wind speeds is certainly in the background, however wind electricity generation represents a nonlinear filter by the power-curve of turbines [82, 91], and the “level” to cross (power demand) is changing in time. Nevertheless we found that the empirical frequency distributions for both continuous intervals L_{sup} and L_{short} obey power law, examples for two series (P_K and P_Σ) presented in Figures 1.65 and 1.66. The frequency distributions of L_{sup} and L_{short} intervals for records $M1$ and $M2$ were found to obey power law as well (not shown).

F	$L_{sup}^*(K)$	$L_{sup}^*(\text{aggr})$	$L_{short}^*(K)$	$L_{short}^*(\text{aggr})$
1.0	2.62	3.18	6.48	7.02
5.5	5.08	6.64	4.38	4.11
10.0	6.64	6.86	4.38	3.50

Table 1.6: Maximum period lengths (in days) for continuous supply (L_{sup}^*) and shortage (L_{short}^*) at three different total power factors F . Wind power records P_K and the aggregated one P_Σ (see Equation (1.31)) are evaluated for years 2005–2006.

A closer look at the curves reveals that a simple power law cannot fit all the data, and systematic deviations are characteristic mostly at the tails. The decay gets slower for periods of supply when the total installed capacity P_{tot} is increased, while the opposite is true for periods of shortfall. The exponent values are difficult to obtain because of the apparent noise, instead we show the absolute maxima in 2005–2006 (Table 1.6) which indicate the same tendency. A saturation effect is also obvious here, as no further increase of P_{tot} can elongate supply periods when all the intervals of shortage have exactly zero wind power production.

1.8 Conclusions

As a first step we examined wind speed statistics above Europe based on the ERA-40 reanalysis dataset. We found that the wind speed probability density functions (PDFs) can be reasonably well described by the 3 parameter generalized gamma model. However the fit is not perfect at all locations, which was found to be the consequence of weak daily- and yearly periodicities. These give a superimposed peak on the “background” generalized gamma PDF, which originates from the wind speed fluctuations due to synoptic scale meteorological phenomena.

It is a difficult task to extrapolate surface (10 m) wind fields to turbine level because of the wind profile’s dependence on local orography and instantaneous meteorological conditions. Despite this fact, the available data permitted the extraction of a spatially and temporally averaged wind speed profile. It is clear from the comparison with turbine wind measurements, that the above mentioned average profile underestimates wind speeds in Moson-szolnok, Hungary, and probably in most of the onshore locations in Europe. On the other hand inland profiles seriously overestimate offshore wind speeds. For a truly realistic extrapolation, a large dataset of tower measurements would be needed, which covers the whole continent. However, in spite of the simplicity of our extrapolation, and the biases introduced by it, we believe that our conclusions concerning all-European wind power integration still hold.

After transforming hub height wind speed to power output we tested various hypothetical European wind power networks by theoretically “installing” turbines at different locations and “connecting” them with a perfect grid. The aggregated output exhibited a low average, a significant yearly periodicity and great fluctuations in the case when the whole continent was connected. The average output can be increased by deploying turbines only at high-wind sites, but this even further increases the fluctuations. An optimal siting of turbines is possible, which minimizes the aggregated power fluctuations, but this is a highly unrealistic configuration, in which huge capacities at far-away low-wind sites try to balance out the fluctuations of minor deployments on the best sites around the Atlantic coast. Furthermore the reduction of fluctuations is not that dramatic in this optimized configuration.

The strong spatial correlations present in the European wind fields determine a lower limit for the fluctuations of the aggregated wind power output, that is Europe is too small to effectively decrease the intermittency of wind power by integrating over larger and larger continental areas. It simply occurs from time to time, that lull conditions prevail over almost the whole continent.

With the large-scale deployment of intermittent resources today, backup generators that can be quickly connected to the grid, or other new technologies are needed, which increase the cost of investment and maintenance. When different intermittent energy sources are combined with each other or over large geographical regions, they are much less intermittent than at one location [91, 95]. Nevertheless a quantitative characterization of different sources is a difficult task, and as we have shown, the use of wind energy only does not give the final answer to our energy needs. The design of an optimal “energy portfolio” is far from being solved [95, 96].

With matching real consumption data with the aggregated output of a small Hungarian “wind farm” we showed a more down-to-earth example of how the wind speed fluctuations determine the power availability. Both supply and shortfall periods showed power law distributions, which did not change considerably by increasing supply capacities. Changing the actual consumption record to constant demand did not change the statistics of supply and shortfall periods either. This shows that the basic problem really lays in the intermittency of wind speed, as real world electricity consumption shows rather regular patterns and relatively small fluctuations (compared to those of wind power).

Besides the problems, wind energy seems to be a key source in the future energy portfolios. However a simple switch to wind energy from fossil fuels is not possible because of the intermittencies, which cannot be smoothed by all-European integration. The transition might still be possible, but in order to achieve a major penetration of wind power, new ideas and technologies seem to be necessary, which can be applied on a truly large scale. These can include e.g. the restructuring of the electric grid, storing excess wind energy in electric cars, hydrogen, or hydro storage plants, etc.

Chapter 2

Analysis of an assumed connection between lightning activity and cosmic rays

2.1 Introduction

One of the most problematic part of today's best climate models – the globally coupled atmosphere-ocean general circulation models – is the correct inclusion of sub-grid effects in connection with clouds. There is an ongoing debate on the role of cosmic rays in cloud formation and cloud lifetime. The nature of the problem is, that the subtle details of the atomic scale microphysical processes – some of which are supposedly only yet to be discovered – have macroscopic effects on quantities like global temperature.

Since the problem of cloud dynamics is in the lime-light of the climate community, numerous groups are doing serious research either in the form of sophisticated computer models or dedicated laboratory experiments. Fortunately the number of high quality measurements is increasing as well. Cosmic rays are known to have a major effect on atmospheric electricity, however they have not yet been linked convincingly to clouds.

Cosmic rays are furthermore thought to play a significant role in lightning activity, based on theoretical considerations, which are supported by a handful of measurements as well. The possible cosmic ray - lightning connection seems to be a little better founded theoretically, than in the case of clouds. But still the details and even the general picture are far from being clear. If the cosmic rays - lightning activity connection were revealed, it could probably help to assess questions concerning cloud dynamics as well. In this work we try to explore the possible effect of cosmic rays on global lightning

activity using data-mining techniques and the best available datasets.

2.2 Overview of terrestrial lightning activity

2.2.1 General features of lightning activity

Globally there are about 45 – 70 lightning flashes per second on Earth based on Optical Transient Detector (OTD) [97] and Cassini [98] satellite data. Most of these are concentrated over tropical land surfaces [99] in the intertropical convergence zone (ITCZ), where strong updrafts occur: vertical wind speeds can reach as high as 10 m/s over the oceans and 50 m/s over land areas [100]. The global distribution of average lightning activity based on Lightning Imaging Sensor (LIS) satellite observations can be seen in Figure 2.1 (see also Figure 2.13 and Figure 4 in [97]).

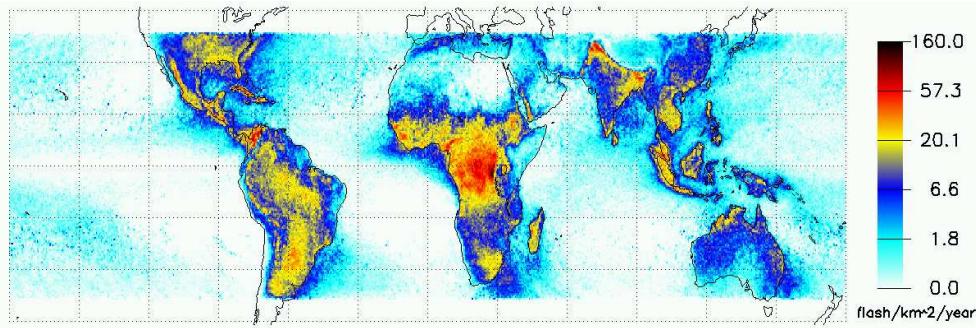


Figure 2.1: Average lightning activity over the globe based on LIS satellite data. Note the three tropical chimneys: Central Africa, South America, and the Maritime Continent, where the lightning activity is maximal due to the convergence of air masses and the large solar heating, which cause strong updrafts.

Most lightning flashes remain in the thunderclouds (IC, intracloud flashes), while only a minority of them span between the surface and the thundercloud (CG, cloud-to-ground flashes). The IC:CG ratio (also referred to as the Z value) was found to be between 2.6 – 2.9 over the continental United States [101], but it shows large variations.

Lightning activity displays a strong daily periodicity in local time due to the daily evolution of convective activity [99]. This periodicity is rather uniform all over the globe: minimal lightning activity is observed in the calm night, then activity starts to rise in the morning and peaks around 14-17 local time, then lightning activity decreases during the evening. As an illustration see Figure 2.2, where the average daily cycle in lightning activity can be seen over the Kennedy Space Center, Florida, USA.

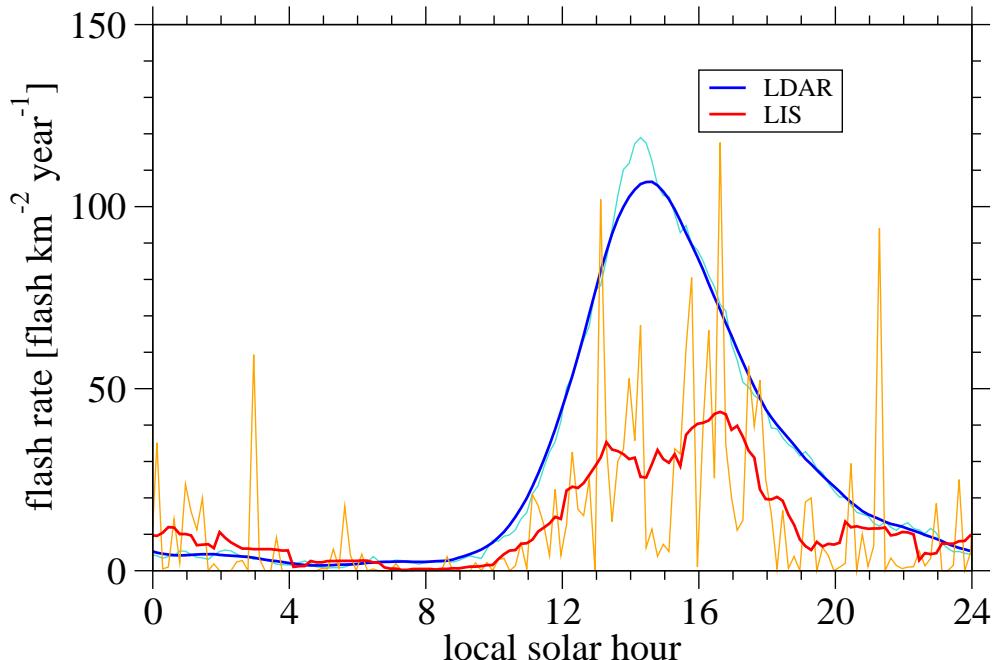


Figure 2.2: The average daily cycle of lightning activity over the Kennedy Space Center, Florida, USA. Optical satellite data (LIS, orange line) and ground-based radio data (Lightning Detection And Ranging, LDAR, cyan line) are presented. Red and blue lines correspond to 2-hour moving averages of the LIS and LDAR data respectively. For a detailed description of the datasets see Section 2.4.

As the daily periodicity in lightning activity is connected to the strength of convection which is mainly determined by the amount of solar heating, it can be expected that the annual (or semiannual in the case of the tropics) cycle is also visible in lightning activity. This is indeed the case: maximum lightning activity follows the thermal equator [99].

2.2.2 The global electrical circuit

There is a large scale electrical circuit in the Earth's atmosphere between the highly conductive ground (conductivity of land is $10^{-8} - 10^{-2}$ S/m, typically $\sim 10^{-2}$ S/m; oceans: $3 - 4$ S/m) and the ionosphere, which is also a good conductor ($\sim 10^{-7}$ S/m) [102, 103]. The ionosphere is a layer of plasma beyond $80 - 95$ km above the ground, which is created by the interaction of solar extreme UV and X-ray radiation with the atmosphere. The atmosphere between these two conductors behaves as a leaky insulator with conductivities increasing with altitude from 10^{-14} S/m at ground level till 10^{-7} S/m at the ionosphere. This weak conductivity is due to ionization from galactic

cosmic rays (GCR) and natural radioactivity (only near the ground). The ionosphere is at a potential of $V_I \sim +250$ kV relative to the ground. Far from thunderstorms there is a so-called fair weather current flowing between the ionosphere and the ground $J_c \sim 2$ pA/m² (positive charges are propagating downwards). In agreement with these there is a vertical potential gradient in the atmosphere with values ~ 130 V/m near the ground. The main “batteries” in the circuit are the thunderstorms, where charge separation occurs via various microphysical processes, out of which charging, which involves the collision or different growth of ice particles is thought to be very important. As a result of these processes large - mainly horizontally stratified - areas of accumulated charge form inside the cloud.

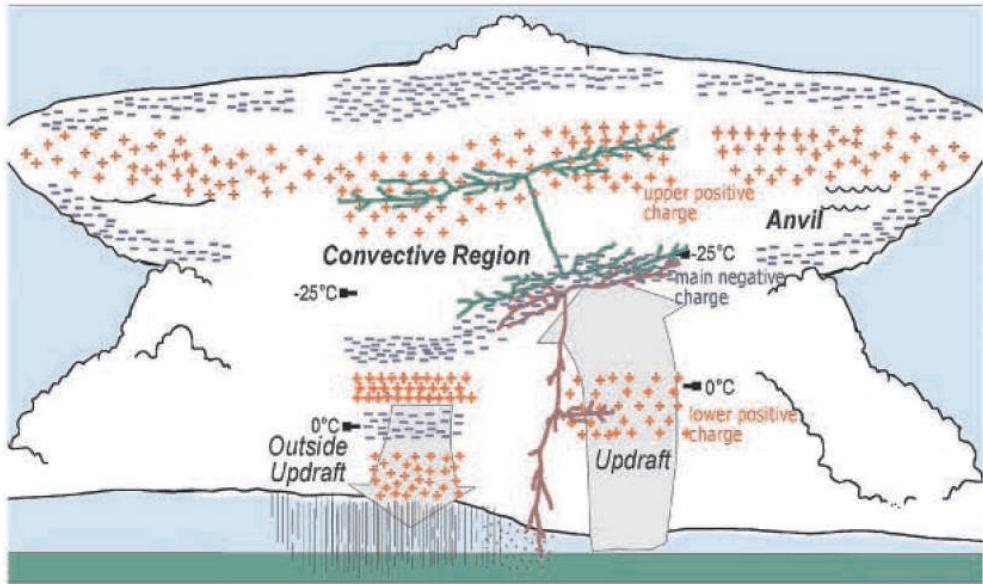


Figure 2.3: Conceptual model of the electrical structure in mature, mid-latitude convection. Four main charge regions (with red + for positive charge, blue for negative charge) are typically found in soundings through updrafts, while soundings outside updrafts have at least six charge regions in common. Schematic representations of an intracloud flash (in green) and a cloud-to-ground flash (in purple) are shown as they might appear in lightning mapping data. Figure adapted from [104].

In the early stages of the thunderstorm’s life there is usually a large negatively charged region in the middle of the cloud (6-9 km above the ground) and a positive region above it [104]. These constitute the main dipole of the cloud. At the bottom of the cloud, below the main negative charge there is a weaker lower positive charge and there is often a shallow layer of negative

charge at the top of the cloud, above the upper positive region. This is only the general simplified picture. The real situation is usually more complex with horizontal variations as well, which are associated with strong updraft / non-updraft conditions. Also, as the storm evolves and many lightning flashes occur, more charge centers are created and the charge structure becomes even more complex. A schematic drawing of the charge structure of a typical thundercloud can be seen in Figure 2.3.

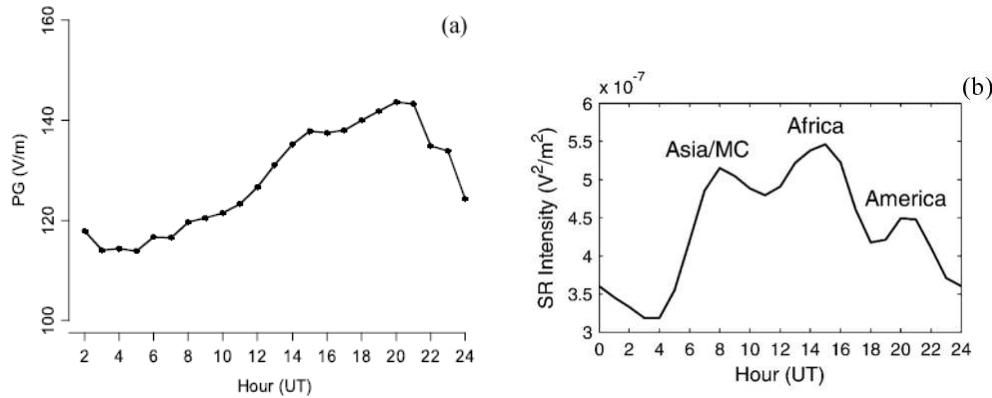


Figure 2.4: (a) Diurnal variation of the mean hourly ground level potential gradient - the so-called Carnegie curve - in fair weather conditions. Data were obtained during the Carnegie ship's cruise 7 between 1928-1929. Figure adapted from [103]. (b) Mean Fall-Winter Schumann resonance intensity at Nagycenk, Hungary. Schumann resonances can be used as a proxy for global lightning activity because the former are directly excited by the latter. The individual peaks can be associated to the three tropical chimneys, as indicated in the graph. Figure adapted from [105]. The differences of the two figures can be due to damping effects of ELF waves (Schumann resonances).

The simple picture is that the main dipole drives the current in the global circuit, and also lightning strikes act as drivers: 90 % of cloud-to-ground lightning strikes transfer negative charge to the ground from the main negative charge region.

A prominent illustration of the thunderstorm - global circuit connection is the good correlation of the so-called *Carnegie curve* (the daily cycle of the ground level potential gradient in universal time) with the global lightning activity [103]. The reader is referred to Figure 2.4. Further drivers of the global circuit can be rain (through negatively charged droplets) and the interaction of solar wind with the Earth's magnetic field can cause additional currents as well.

A simplified model of the DC (direct current) global circuit described

above can be seen in Figure 2.5

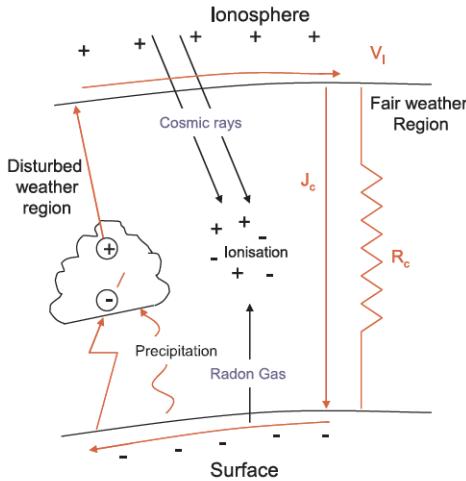


Figure 2.5: Simplified “electrical engineering” model of the global circuit, illustrating charge generation in disturbed weather regions, conduction through the ionosphere (upper layer) and surface (lower layer), and discharge through the finite conductivity of fair weather regions. In fair weather regions, the ionospheric potential V_I , conduction current density J_C and unit area columnar resistance R_C are related by Ohm’s Law. Figure adapted from [102].

To make the picture more complete, there is a wide variety of recently discovered luminous discharges in the Earth’s atmosphere between the top of the thunderclouds and the ionosphere. They are collectively called transient luminous events (TLEs), e.g. sprites, jets, elves. TLEs are believed to result from discharge processes that develop in the quasi-electrostatic fields that appear in the upper atmosphere following a cloud-to-ground lightning discharge in which large quantities of positive charge (~ 100 to several hundred Coulombs) and occasionally negative charge are transferred to ground [106]. Most TLEs are thought to be events of similar intensity as individual lightning strikes, on the other hand the *gigantic jets* are thought to have a macroscopic effect on the global circuit.

As an addition to the DC global circuit there is also a so-called AC (alternating current) global circuit: the cavity between the ground and the ionosphere behaves like a resonator. The Schumann resonances – electromagnetic standing waves – at frequencies of 8, 14, 20, . . . Hz are excited by the lightning activity, since lightning strikes radiate a very broad spectrum of electromagnetic radiation [102, 105].

2.2.3 The breakdown processes during lightning flashes

The path of the lightning discharge – the plasma channel – is created by some type of breakdown mechanism: in the presence of a strong enough electric field (\mathbf{E}), electrons are accelerated if the force $\mathbf{F} = e\mathbf{E}$ acting on them (e is the elementary charge) is larger than the “frictional force” due to inelastic collisions (translational, rotational, vibrational, and ionizing collisions) with the neutral molecules. Furthermore, in order to ionize the air along their path, the energy of some of the accelerated electrons must exceed the ionization energy of the neutral gas, and the production rate of secondary electrons by ionization must exceed their recombination rate with the positive ions. If these conditions are met, an avalanche of electrons can form in the neutral gas.

Presently two breakdown mechanisms are known: the conventional breakdown [106] and the runaway breakdown [106, 107, 108]. In the case of the conventional breakdown thermal electrons (due to cosmic ray ionization) constitute the seed for the electron avalanche, which is then driven by electrons in the high energy end (tens of eV) of the electron energy distribution. The critical electric field required for the conventional breakdown is ~ 3000 kV/m in clean air at standard atmospheric pressure, and it can decrease to ~ 350 kV/m (also at standard pressure) due to local field enhancements caused by raindrops [104].

In the case of the runaway breakdown a fast (tens of keV) seed electron is needed, and the avalanche is driven by fast electrons as well (10 keV - 10 MeV). The main idea is that the frictional force (due to ionization) acting on such fast electrons is smaller than the one acting on ~ 10 eV electrons. The reason is that in collisions with such high energies, the neutral molecules can be considered as free electrons and nuclei, and therefore the smaller Rutherford-Coulomb cross section is to be considered instead of the classical constant two-body collisional cross section. Of course the different excitational cross sections of the different molecules should be taken into account as well.

While detailed numerical calculations and some initial laboratory experiments have been carried out concerning the runaway breakdown, precise and comprehensive experimental validation is not presently available. Laboratory experiments are rather difficult because of the large avalanche scale lengths (tens of meters at atmospheric pressure).

An important difference between runaway breakdown and conventional breakdown is that when fast electrons collide with matter (in the case of runaway breakdown), then high energy γ - and X-ray radiation is produced (“Bremsstrahlung”). These are however difficult to observe because of their

attenuation in the air (X-ray attenuation length is around 100 m to 1 km at thunderstorm altitudes). Still there exist some measurements (ground based, balloon-borne and satellite) of terrestrial thunderstorm related γ - and X-ray radiation in the atmosphere [106, 107, 109].

The threshold electric field for runaway breakdown is ~ 280 kV/m at standard atmospheric pressure, which is lower than for conventional breakdown. However it must be noted that all threshold electric fields are proportional to air density, thus the values must be scaled with height.

Balloon measurement of thunderstorms show that electric fields are typically less than 150 kV/m with a few extreme values up to 400 kV/m, however in order to compare these values with the threshold electric fields, scaling to standard atmospheric pressure is needed. The few extreme \mathbf{E} values were in most cases observed just before a nearby lightning flash, which usually destroyed the instrument. As of 2008, when adjusted to standard pressure, the largest observed electric field value $E = |\mathbf{E}|$ was 626 kV/m (127 kV/m at 13.4 km) and the largest estimated E was 929 kV/m (~ 200 kV/m at an estimated altitude of 12.2 km) [104]. All these extreme values exceed the runaway breakdown threshold by a factor of 1.1-3.3 and some exceeded the hydrometeor-enhanced conventional breakdown threshold as well. However it is unlikely that liquid water exists at such high altitudes (ice particles can decrease the conventional breakdown threshold as well to 400-500 kV/m at sea level). In order to determine E , in situ measurements are needed, and it is difficult to sample a huge thunderstorm with a few balloons. Furthermore, when the situation gets interesting, the instruments might fail due to nearby lightning strikes. However there is a growing number of observations, which might hopefully shed more light on the initiating processes of lightning.

2.2.4 Description of one distinct lightning flash

One distinct lightning flash is still a series complicated processes which span altogether 0.1 – 1000 ms in time. It might be useful to review briefly what happens during the most extensively studied type of lightning, the downward negative cloud-to-ground (CG) flash. Downward means that the leader, which is the process that creates the ionized channel through some breakdown mechanism, propagates downward from the cloud charge towards the ground. Negative means that negative charge is transferred to the ground. It is believed that 90 % of CG flashes are downward negative. First of all an initial breakdown occurs in the thundercloud, however, “the exact electric field magnitude and the specific mechanism for lightning initiation remain unknown” [104]. This breakdown initiates a stepped leader, a negatively charged plasma channel, that propagates downward in discrete steps: each

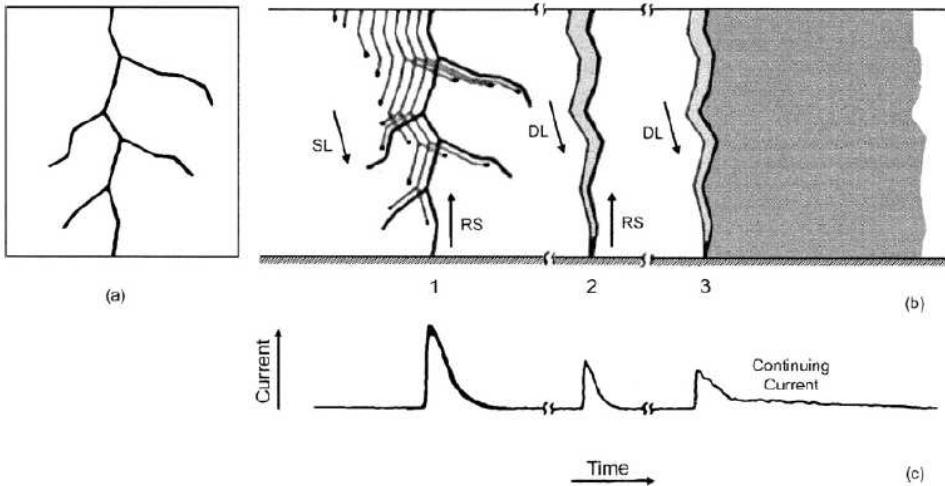


Figure 2.6: Diagram showing the luminosity of a downward negative lightning flash to ground containing three strokes and the corresponding current at the channel base: (a) still image, (b) time resolved image, and (c) channel-base current. The abbreviations stand for: (SL) stepped leader, (RS) return stroke, and (DL) dart leader. Figure adapted from [110].

step is tens of meters long, takes $\sim 1 \mu\text{s}$ with $20\text{-}50 \mu\text{s}$ intervals between steps. Several Coulombs of negative charge are deposited along the channel. As the downward leader gets close to the ground an upward connecting leader initiates from the ground (tens of meters long). As the two channels connect, a return stroke wave starts and carries positive charge upwards to neutralize the leader charge (but this neutralization is usually not complete). The return stroke takes tens of microseconds and the peak current is about 30 kA. This process emits a broad spectrum of electromagnetic radiation and a shock wave (thunder) is also initiated as the channel rapidly heats up and expands. In most of the negative CG flashes there are subsequent return strokes, which are preceded by so-called dart leaders – similar to the stepped leaders, but these propagate downward along the already existing main channel. There are typically 3–5 strokes separated by tens of milliseconds. The total negative charge lowered to the ground is several Coulombs [110]. The schematic view of the processes described above can be seen in Figure 2.6. As a real world example, the stepped leader of a lightning strike (21 March 1997 11:56:6.19 UT) above Florida can be seen in Figures 2.7 and 2.8 captured by the LDAR detector.

The various types of lightning (IC, GC, and even the upward propagating jets) are reasonably well understood [111, 112] concerning their propagation,

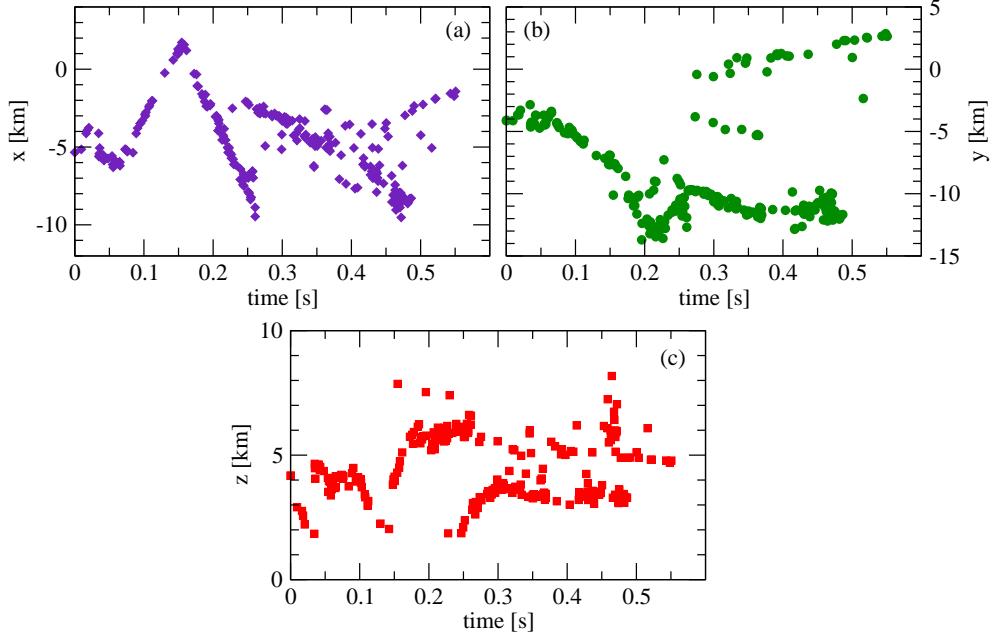


Figure 2.7: The time evolution of the stepped leader of a lightning strike, which started 21 March 1997 11:56:6.19 UT above the Kennedy Space Center, Florida: (a) x (northward) coordinate, (b) y (eastward) coordinate, and (c) z coordinate (elevation) of the leader steps is shown versus time elapsed since the first detected breakdown. Data of the LDAR instrument are presented.

given the charge structure and the initial breakdown in the thunderstorm.

2.3 Cosmic radiation – the missing link in climate?

2.3.1 General properties of cosmic radiation

Cosmic rays (CR) are energetic particles (mostly protons, $\sim 10\%$ He nuclei, $\sim 1\%$ electrons, $\sim 1\%$ other elements) which are constantly bombarding the Earth's atmosphere. Their energy varies from a few MeV to beyond 10^{20} eV. Their flux decreases rapidly with increasing energy from a few particles per square centimeter per second to less than one particle per square kilometer per century above 10^{20} eV [113, 114]. Most of the cosmic rays come from outside the Solar System (e.g. neutron stars in our galaxy). The most energetic particles were recently linked to nearby active galactic nuclei [114]. Charged particles arrive from the Sun as well with the solar wind (Solar En-

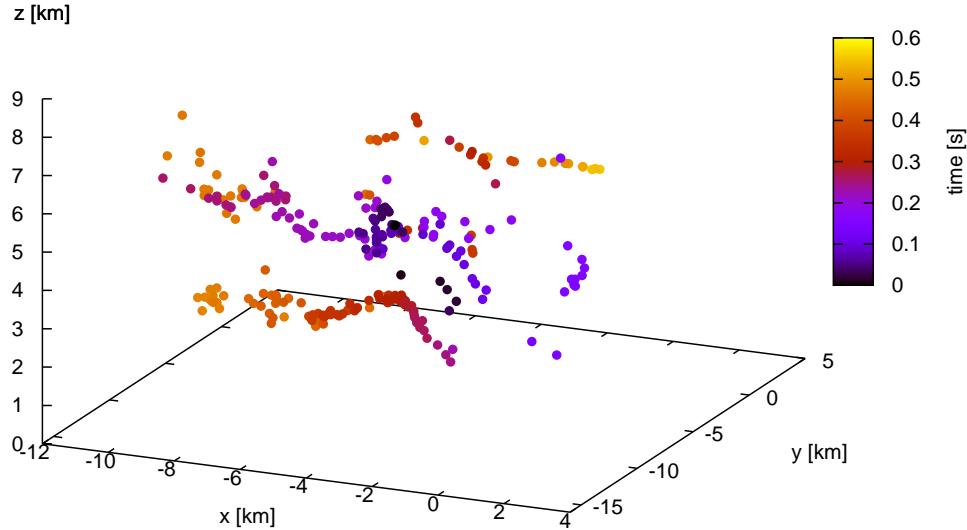


Figure 2.8: The spatial structure of the stepped leader of Figure 2.8.

ergetic Particles - SEP), but these are less energetic (tens of keV to GeV) and are absorbed in the upper atmosphere. However there are events connected to solar flares and coronal mass ejections when the solar energetic particles can reach the ground and their flux can increase by as much as 4 orders of magnitude and eventually get larger than that of galactic CRs for a few hours [115, 116]. A third source of charged particles needs to be noted as well: electrons can precipitate to the Earth from the Van Allen belts in the Earth’s magnetosphere.

Cosmic rays are responsible for the ionization of the lower atmosphere with peak values of ~ 50 ion pairs $\text{cm}^{-3} \text{ s}^{-1}$ in the stratosphere and ~ 2 ion pairs $\text{cm}^{-3} \text{ s}^{-1}$ at ground level. At low altitudes (up to 2-4 km) the Earth’s radioactivity adds to the ionization with a few ions $\text{cm}^{-3} \text{ s}^{-1}$ [117].

Measurements of cosmic radiation flux are possible either directly from satellites, balloons, or indirectly from ground level through the secondary particles (e.g. muons, neutrons), which are created when CR particles with energies > 1 GeV collide with the atomic nuclei of air.

Since the cosmic radiation consists of charged particles, the interplanetary magnetic field and the Earth’s magnetosphere deflect them. Low energy particles can arrive to the lower atmosphere only around the poles, while only particles with energies > 14 GeV (for protons) are able to reach the equatorial regions [118]. The motion of cosmic rays in the geomagnetic field is determined by the so-called particle rigidity $R = \frac{cP}{ze}$, where c is the speed

of light, P is particle momentum, z is the ionic charge number, and e is the elementary charge. For a given geographical location a vertical cutoff rigidity R_c can be determined, which is the minimal rigidity required for a particle to reach the ground vertically. The global map of R_c can be seen in Figure 2.9. The values of the cutoff rigidity are obviously in close connection with the strength of the geomagnetic field.

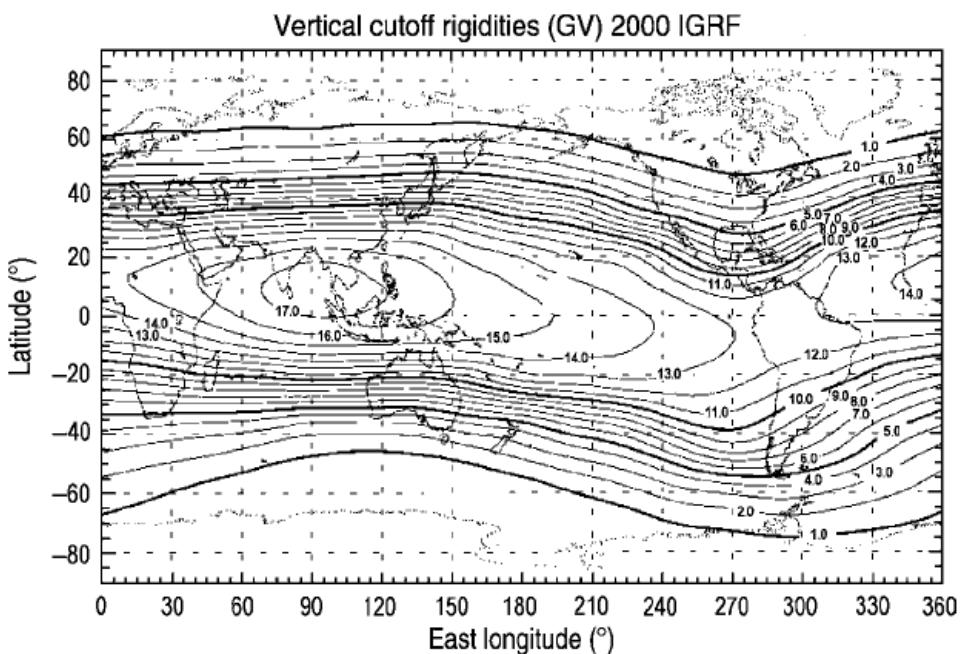


Figure 2.9: Contours for vertical geomagnetic cutoff rigidities in GV for the epoch 2000 based on the IGRF (International Geomagnetic Reference Field) model. Figure adapted from [119].

2.3.2 Modulations of cosmic radiation flux

From the above reasoning it follows, that the *variations* of the magnetic field around the Earth are reflected in the CR flux as well. Indeed, the CR flux and the ionization rate follow very well the 11-year solar cycle and show very high anticorrelation with the solar quantities (e.g. sunspot number, solar radiation): the higher magnetic fields deflect more CRs during the solar maxima. To be more precise, even the 22-year solar magnetic cycle is reflected in the CR flux records (the Sun's magnetic field reverses polarity between subsequent solar maxima). As an illustration see Figure 2.10.

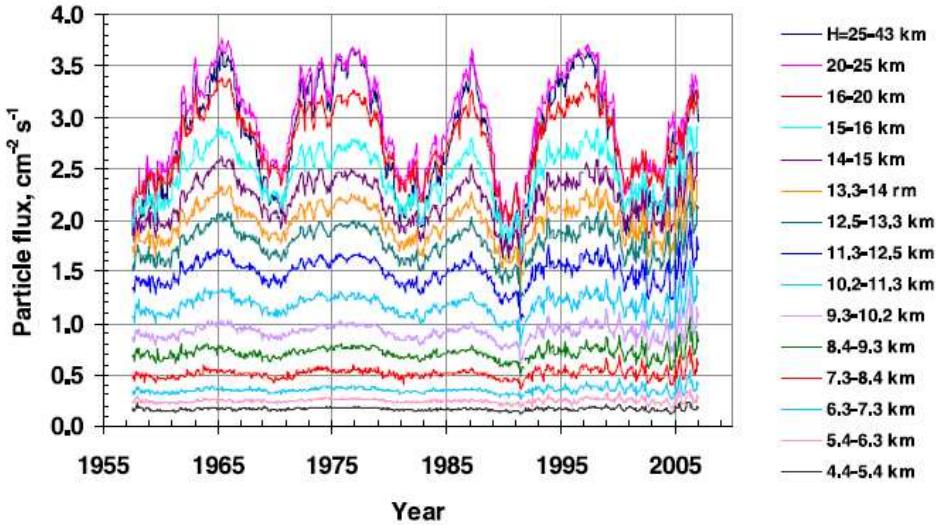


Figure 2.10: Monthly averaged fluxes of ionizing particles in the atmosphere over Murmansk region as measured by an omni-directional Geiger counter. Various colors present fluxes at various heights. Days when solar or magnetospheric particles invaded the atmosphere are excluded from averaging. Figure adapted from [118].

Besides the 11-year cycle, the Sun's rotation (with a period of ~ 27 days) can affect the CR flux as well in cases when the solar magnetic field is very inhomogeneous (due to sunspots).

There are solar modulations of the CR flux on shorter timescales as well (3-10 days) associated with the so-called *Forbush decreases*. These are events of increased magnetic field (lower CR flux) due to solar flares and coronal mass ejections. Forbush decreases occur mainly in the decreasing phase of solar maxima: the CR flux sharply decreases (1-5 days), and then gradually recovers (2-10 days). For examples of Forbush decreases see Figures 2.20 and 2.22a, b. Forbush decreases can be considered to be stochastic events, therefore they provide a great opportunity to link various atmospheric variables to cosmic radiation. Note, that in the case of the 11-year solar cycle it is barely possible to distinguish the effects of the irradiation change and the CR flux change by only examining datasets. The amplitude of cosmic ray flux modulations due to both the 11-year solar cycle and the Forbush decreases are about 5-30 % .

The changes of the Earth's magnetic field affect the CR flux as well, but these act on longer time scales.

2.3.3 Climatic connections - Clouds

The overwhelming majority of climatic processes is driven by the energy of solar radiation. This radiation is rather constant (solar constant = 1.37 kW/m²), but shows variations due to the effects mentioned above and also longer term changes in the Sun and the Earth's rotation and orbit changes play a role. Today the most pronounced effect - besides daily and yearly variations - is the ~ 0.1 % change in total solar radiation due to the 11-year cycle [120]. Most of this variation is concentrated in the UV part of the spectrum, which is absorbed in the upper atmosphere.

In contrast to these small effects, pronounced solar influences (e.g. 11-year cycle) have been found in many climatic variables or quantities connected to them starting from the wheat prices of the late 18th century [121], and later temperature, thunderstorm frequency, atmospheric circulation, etc. were associated with subtle changes of the irradiance [122]. Even the stream flow of the river Paraná was shown to be strongly correlated with the solar irradiance [123]. But to go a little more back in time, the Maunder minimum [124] – a period with almost no sunspots between 1650-1715 – is often associated with the Little Ice Age in Europe.

The fact that these variables are connected to the main driving force, the solar radiation is not surprising at all. However the magnitude of the effects mentioned above is considerably larger, than what a 0.1 % change in the forcing would imply (in some cases the effect can be measured in tens of % [123]). Therefore some mechanism is needed to explain the amplification. Two major mechanisms were proposed: (1) a direct radiation forcing effect or an indirect dynamic effect through the change of the amount of stratospheric ozone, which depends on the incoming UV radiation. But this dependence is not trivial, and far from being understood. Not to speak about the further effects [120]. (2) Modified cloud formation effects, or cloud lifetime caused by the changing cosmic ray flux, through enhanced aerosol formation due to increases in ambient ion concentration, or electric effects in already developed clouds [125].

The mechanisms involving clouds and cosmic rays might seem to be promising, because on one hand, the energy transported to Earth by the cosmic rays is marginal (comparable to starlight) [122], but on the other hand their importance is huge, since they ionize the atmosphere and thus they are a key element in creating the global electric circuit. Furthermore CR flux shows large, 5-30 % variations due to changes in solar activity, so large responses seem to be more plausible.

The above mentioned mechanisms (involving ozone and cosmic rays) were tested mainly in computer simulations, either in the largest scale general

circulation models, or in small scale microphysical cloud models, since both accurate atmospheric measurements and laboratory experiments are very difficult to carry out. A general problem is that almost any mechanism can be envisaged, but to actually quantify, or sometimes even to identify the effects is a very difficult task.

Cloud microphysics and its connections to global climate is a very vibrant field today [126], where cutting-edge research is being carried out. These include computer simulations, sophisticated laboratory experiments, like the CLOUD experiment (see <http://www.cloud-itn.uni-frankfurt.de/>) prepared at the CERN, Switzerland, etc. The leading motivation is to quantify the different processes from small to large scales and to supply physics for the climate projections.

A very useful tool to identify mechanisms and link different phenomena is to search for correlations between different environmental variables. The most suggestive arguments supporting the CR-involving mechanisms are correlations between cosmic radiation flux and cloud cover during a solar cycle as described by Henrik Svensmark and his collaborators [127, 128, 129]. In addition several other authors proposed similar CR-climate connections [130, 131], like the one described by Svensmark et al. However, the “evidences” are strongly debated because of many reasons: (1) The cloud dataset used in the various studies, the International Satellite Cloud Climatology Project (ISCCP, see <http://isccp.giss.nasa.gov/>) is based on infrared remote sensing and it is presumably not free of biases, however it is constantly improving. (2) If the direct CR-cloud relationship proves to be responsible for the observed correlations through the “ion-aerosol clean-air mechanism” and the “ion-aerosol near cloud-mechanism” (see [125]), then a complete reevaluation of global warming theories would be necessary. (3) A causal relationship does not necessarily follow from the mere correlations observed.

However, much more can be deduced from the data, than only a single number indicating high correlation. In the earlier studies, a connection between the amount of high cirrus clouds and Forbush decreases of the cosmic radiation was identified [132] using surface actinometric data. However later studies did not confirm the correlation between high clouds and CRs [129]. On the other hand, low cloud coverage (up to 3-4 km) was found to correlate with the 11-year modulation of the CR flux by Svensmark et al. in the above mentioned papers, see Figure 2.11. The problem is that on a closer look this connection already seems dubious, as neither the well known latitude dependence of the CR flux, nor the effects of the Forbush decreases are visible in the low-cloud records [122]. The question, whether the CR induced changes in atmospheric ionization significantly affect cloud cover or not, is still unanswered.

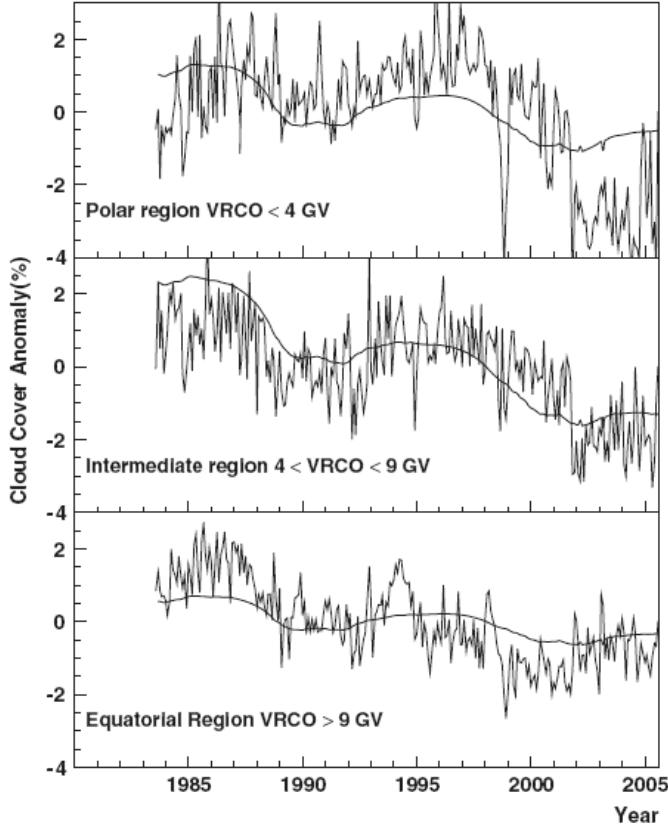


Figure 2.11: The low cloud cover anomaly as a function of time for various ranges of vertical cut-off rigidity (VRCO). The smooth curve shows a fit of the monthly mean of the daily sunspot number with an assumed linearly falling systematic change. The sunspot number is anti-correlated with the CR count rate with a lead time of some months. Figure adapted from [122].

2.3.4 Connection with lightning activity

While the hypothesized CR-cloud connection described above includes various poorly understood mechanisms, that are practically not supported by any evidence based on direct measurements, the connection between cosmic radiation and lightning activity seems to be more established.

Cosmic rays might obviously play some role in thunderstorm development (via inductive charging mechanisms) and lightning propagation by changing the properties (e.g. potential gradient [133]) of the global electric circuit as they are responsible for the weak ionization of the air, and they are a key element of the circuit. There exist some studies, in which the authors investigated a possible connection between CRs and thunderstorms [134, 135],

however the results are not convincing.

The exact way of lightning initiation is so far unknown. Perhaps the most accepted theory of runaway breakdown [107] heavily relies on cosmic rays, which – in theory – provide the fast seed electrons for the first breakdown during the initiating processes of lightning strikes (see Section 2.2.3). These considerations suggest, that there should be a connection between lightning activity and cosmic ray flux, which should be visible in the observations as a positive correlation. If confirmed, this link between cosmic rays and a tropospheric climate variable – lightning activity – could perhaps shed more light on the more general CR-climate speculations.

Our aim was to test the above mentioned hypothetical CR-lightning connection. Our analyzes are presented in the following sections (see also [136]) after a description of the data we used.

2.4 Datasets used

2.4.1 Lightning data

Local radio networks

It is possible to monitor lightning activity in various ways due to the strong electromagnetic emissions and the wide spectrum emitted. The main reasons for the observations are meteorological forecasting, scientific research and the protection of man-made structures. There exist several country scale ground-based radio lightning detection networks, which record hundreds of lightning flashes daily, however their data are mostly inaccessible. A very precise network is operating in the United States, which covers an area, that is presumably large enough to be worthwhile to search for global effects, such as the hypothesized connection with cosmic rays. However, despite extensive negotiations with the United States Precision Lightning Network (USPLN), we did not manage to obtain archive ground-based VHF observations of lightning activity over the United States.¹

There exist so-called lightning mapping arrays, which are small scale local networks of VHF antennae, which detect the electromagnetic radiation emitted by the *stepped leaders* (see Section 2.2.4) of a lightning strike, thus making it possible to examine the evolution of individual lightning strikes with high spatial and temporal resolution.

We used the level 1 data of the ground based Lightning Detection And

¹Here I would like to thank Kim Rauenzahn at the USPLN for his/her kind help and efforts to try to provide us with the USPLN data, despite the lack of agreement.

Ranging (**LDAR**) system, that is space (3D) and time coordinates of the stepped leaders.² This is a seven-station VHF (operating at 66 MHz), time of arrival network capable of detecting both cloud-to-ground and intra-cloud lightning in 3 dimensions. The LDAR is located at the NASA Kennedy Space Center (KSC) in Florida. The coordinates of its center are 28.5387° N and 80.6428° W. Since the data were practically unprocessed, extensive filtering was needed. The data were filtered for the period 1997-2007. Events with an invalid time stamp, a stationary calibration signal and the aeroplane tracks (for more information on these biases see [137]) have been removed using our own clustering algorithm. Unfortunately the aeroplanes emit 66 MHz radio signals during their flight as well. In total 8638 aeroplane tracks were identified. Unfortunately not all of the aeroplanes were removed by the algorithm, but their number has been reduced by approximately two orders of magnitude. The ones with a curved trajectory or flying close to thunderstorms were not identified by the algorithm thus adding some bias to the lightning data. Only data within an 80×80 km square (north-south and east-west sides, centered at the LDAR center) were used, where the detection efficiency is very high, close to 100 % [137]. The LDAR has very high time coverage over the almost 11 years of measurements used here, but the spatial coverage is very low with only about 0.001 % of the Earth's surface covered.

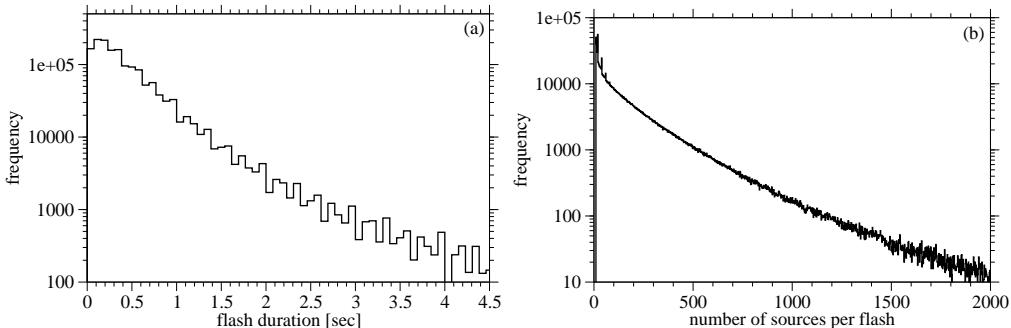


Figure 2.12: (a) The distribution of the time duration of individual LDAR flashes is shown along with (b) the number of events (leader steps) detected in the LDAR flashes.

Since the LDAR detects the stepped leaders of a lightning flash, there are numerous events recorded during one flash. Besides the LDAR event time series we extracted and analyzed the flash rate time series as well. The latter data were obtained using another clustering algorithm. Only flashes with at least 10 events were accepted. This way part of the remaining aeroplane

²Here I would like to thank Sherry Harrison at the Global Hydrology Resource Center (GHRC), Huntsville, for kindly providing us the LDAR data.

tracks and some noise were filtered out. Note that it is a very difficult task to separate individual lightning flashes in severe storms. Therefore it is possible, that more flashes were considered as one in extreme storms, thus adding some bias to the LDAR flash data. The distribution of the duration of individual flashes and the number of events (leader steps) detected in the LDAR flashes can be seen in Figure 2.12.

Global radio networks

Radio waves in the VLF range are damped very slowly in the atmosphere, therefore the detection of lightning strikes is possible from several 1000 kilometers in this frequency range. The World Wide Lightning Location Network (**WWLLN**) which is an international project between various science institutions (including the Lornd Etvs University, Budapest) started in 2003 with 11 stations. In September 2009 the network already consists of 40 stations around the globe. The detectors work in the 6-22 kHz band, where the vertical electric field from strong lightning strikes usually exceeds the background noise (due to power lines mainly).

We used the “Reloc A” data of the WWLLN which refers to the algorithm used to match the individual lightning strikes observed at the different stations.³ The data base contains lightning strikes, which were identified by at least 5 stations [138]. The location accuracy is very good, about 10 km. We used data between the beginning of 2005 and the end of January 2009.

The WWLLN data have very good spatial and temporal coverage and resolution as well. The main disadvantage of this dataset is the varying detection efficiency. It is (1) changing in time due to the increasing number of detectors installed worldwide, (2) not constant in space either due to the different density of detectors at different geographical locations. Obviously the United States and Europe are very well sampled, but Central Africa is unfortunately very underrepresented, where the lightning activity is the most intense on our planet. The global detection efficiency was estimated to be between $\sim 1.3\%$ (2005) and $\sim 3.0\%$ (2007). This is in agreement with the satellite observations (LIS). The changing detection efficiency makes global statistical analyzes very difficult if not impossible. The average lightning frequency based on the WWLLN data can be seen in Figure 2.13.

The reader is referred to the homepage <http://wwlln.net/> for further information on the WWLLN project. The growing number of scientific papers (also available on the given homepage) using the WWLLN data confirms the importance of this international network.

³Here I would like to thank Jnos Lichtenberger at the Space Research Group, Lornd Etvs University, Budapest to kindly provide us the WWLLN data.

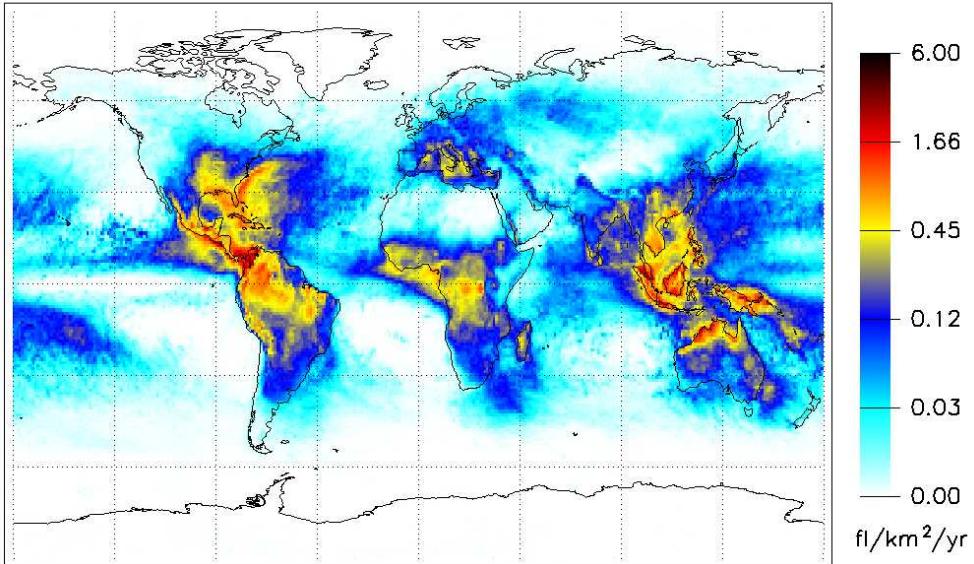


Figure 2.13: Average lightning activity over the globe based on WWLLN ground based radio data. Compare figure with Figure 2.1: the undersampling of Central Africa by the WWLLN network is clearly visible.

Satellite measurements

Optical detection of lightning strikes is possible as well, which is probably best achieved from satellites. Two major satellite datasets exist: (1) the Optical Transient Detector (OTD) was operating on board the tiny OrbView-1 (formerly MicroLab-1) spacecraft between April 1995 and March 2000. (2) the Lightning Imaging Sensor (**LIS**) is operating on board the major Tropical Rainfall Measuring Mission (TRMM) satellite since November 1997. The interested reader is referred to the TRMM homepage <http://trmm.gsfc.nasa.gov/>. Both the OTD and LIS instruments are similar, therefore only the LIS instrument is described below. The major difference between the two datasets is the spatial coverage due to the satellite orbits (both the OrbView-1 and the TRMM are low Earth orbit satellites with periods of a few hours). The OTD mapped lightning activity between 70° North and South latitudes, but the LIS data cover only the tropical region between 35° North and South latitudes.

Because of the longer temporal coverage and better detection efficiency we used only the LIS data. The LIS instrument consists of a 128x128 pixel CCD detector which takes images at a sampling rate of ~ 500 frames per second. The field of view is about 600x600 km and the spatial resolution varies between 3-6 km [139]. In front of the detector, a narrow-band filter

is used to improve signal-to-noise ratio. The filter is centered on the OI(1) neutral atomic oxygen multiplet at 777.4 nm. The TRMM satellite featuring the LIS instrument orbits the Earth at an altitude of ~ 350 km with an orbit period of 92 minutes. The equatorial region between 35° North and South latitudes is almost fully sampled in 0.96 days (15 periods). However, one overpass takes only ~ 90 seconds, that is each geographical location is observed only ~ 90 seconds a day (0.1 % of time).

LIS records transient events (sudden brightening), which is a very robust method. Furthermore, it is a relative measurement, therefore it is less prone to biases due to instrument degradation. This makes it reasonable to evaluate trends in long records as well.

We did not use any correction for detection efficiency, which is thought to be at most ~ 95 % for nighttime and ~ 75 % for daytime retrieval [140]. With neglecting the dependence of the detection efficiency on local solar hour, about a 10 % error is introduced in the data, however this is negligible compared to the effects of very low sampling. Furthermore the global averages we used to extract time series efficiently eliminate this bias. Note also that other groups found the detection efficiency to be somewhat lower [141], which is plausible, since the satellite observes the cloud top, and mainly flashes with a component in the upper part of the thunderstorm – mainly IC flashes – are detected. Our findings support this statement as well – see Figure 2.2 for a comparison with LDAR data. In the above mentioned studies concerning the detection efficiency of (LIS) satellite data, very precise lightning mapping array data were used as a comparison. Note also that the LIS and LDAR datasets were inter-compared as well [142, 143], and a good agreement was established.

The so-called LIS Science Data we used, were already processed and filtered: flashes with more return strokes spanning more pixels were grouped into different data products, out of which the “flash” product is thought to correspond to individual lightning strikes. The interested reader is referred to [144] for the details of the raw data processing algorithms. A large number of filtering algorithms is already incorporated in the *flash* product to minimize the effects of detector overflow, energetic particles, sea glint, etc. However, we used some further filters to ensure high data quality at the expense of losing some dubious data. The additional filters were: we allowed all alert flags (instrument, platform, external factors, and processing algorithms) to say warning at most (data with any of the flags failed were discarded), a minimum cluster index (probability of not being random noise based on surrounding events) of 80% and a minimum glint view angle of 5° were set. Also a minimum effective observation time of 0.1 seconds was specified for the continuous visibility of a gridcell in one overpass. Note that the LIS

Science Data are a set of flashes observed during the satellite's path and the corresponding viewtime and orbit data. Therefore further processing was necessary to obtain flash rates on a geographical grid.

We used data between 1998 and 2007. A primary gridding in time and space was performed on the data before final processing, that is we extracted 20 second resolution flash rate time series with a spatial resolution of 0.5×0.5 degrees (latitude/longitude), which were later averaged over certain areas.

The LIS Science Data are freely available at <ftp://ghrc.nsstc.nasa.gov/pub/data/lis/science>. A useful site with a list of available lightning data – including LIS – and documentation can be found at <http://thunder.msfc.nasa.gov/>.

Schumann resonance data

With special ELF antennae, it is possible to measure the Schumann resonances of the Earth-ionosphere cavity which are excited by the global lightning activity. The ELF radiation exhibits very little attenuation in the atmosphere, therefore it is possible to measure the effects of the most distant flashes as well, but the individual flashes are not distinguishable (note, that there are individual events visible in the records – so-called Q-bursts – which were connected to transient luminous events [105]). Albeit the damping of ELF waves is small, spatial effects exist, e.g. the measured intensity of the tropical chimneys depends on their distance from the location of the measurement. Several other effects complicate the interpretation of Schumann resonance observations: the properties (e.g. height) of the ionosphere are different in daytime / nighttime conditions, which affect wave propagation; the slightest motion of the antennae relative to the Earth's magnetic field add noise to the measurements, and even the global temperature seems to affect Schumann resonance intensities [145].

We used the freely available data from the Northern California Earthquake Data Center (NCEDC). We used the two horizontal magnetic component intensities of the first Schumann resonance (7.8 Hz), which were recorded around San Francisco, California at the following sites: Parkfield (PKD1, PKD) and Hollister (SAO). For further information see <http://www.ncedc.org/>.

The data records are generally very noisy and hardly comparable, thus limiting their usage.

2.4.2 Cosmic radiation data

Ground-based neutron monitors

As a proxy for global cosmic ray intensity, we used ground based hourly resolution neutron count data of three IGY neutron monitors⁴, which are operating at the following sites: **Jungfraujoch**, Switzerland (for a description and history of this instrument see [146]), and **Climax**, Colorado, USA, and **Haleakala**, Hawaii, USA. For a review of the global neutron monitor network, operating since the 1950s, the reader is referred to [147].

The data of the Jungfraujoch neutron monitor are accessible at <http://cosray.unibe.ch/>, while those of the Climax and Haleakala detectors can be downloaded at http://ulysses.sr.unh.edu/NeutronMonitor/neutron_mon.html.

Direct satellite measurements

We used direct satellite measurements of the cosmic ray flux as well, that is the data of the Medium Energy Proton and Electron Detector (MEPED) and the Total Energy Detector (TED) instruments on board the Polar Orbiting Environmental Satellites (POES) of the National Oceanic and Atmospheric Administration (NOAA). The detectors are directional: either zenith or horizon viewing, and contain filters corresponding to different particle energies. Protons and electrons are detected from 30 keV up to > 300 keV (for electrons) and > 6900 keV (for protons). Omni-directional proton detectors are used to resolve higher energies, up to > 140 MeV.

The data are available at http://www.ngdc.noaa.gov/stp/NOAA/noaa_poes.html, where further information can be found on the dataset as well.

2.4.3 Cloud data

We took a brief look at cloud data as well, in search of cosmic radiation effects. We used the “reflectivity” product of the Ozone Monitoring Instrument (OMI) on board the EOS Aura satellite. OMI reflectivity is measured at 380 nm, and high values correspond to thick cloud coverage. Data are available since October 2004 at <http://toms.gsfc.nasa.gov/>.

⁴IGY refers to the International Geophysical Year 1956/1957, when the instrument design was standardized.

2.5 Construction of time series

Because of the very poor temporal sampling (LIS dataset) and the extensive processing required (LDAR and also LIS data), it is not a trivial task to obtain lightning time series from the available data. The only variable we extracted from the lightning data is lightning flash rate, that is the number of lightning flashes per unit area per unit time. Note, that lightning flash rate is a non-negative quantity. We had no data on the polarity and the charge transferred by the flashes. Besides it is hard to see, how polarity and intensity would be coupled to cosmic rays, therefore here they might be considered to be less important.

2.5.1 Lightning statistics of Central Africa – an example

In order to understand the capabilities and limitations of the LIS dataset, let us examine the lightning flash rate statistics over a smaller geographical area, the part of the Congo Basin in Central Africa between 1.5° Southern and 1.5° Northern latitudes and 22.0° and 24.5° Eastern longitudes. Note, that this is the area where most lightning flashes occur on Earth.

The examined area consists of $30\ 0.5^{\circ}$ by 0.5° (latitude/longitude) primary gridcells, for which 20-second gridded data were extracted. Note, that during the satellite overpasses usually only part of the area was seen by the satellite for only a part of the 20 second intervals. In these cases (which give the overwhelming majority) the flash rates based on the reduced sample were assigned to the whole area. Averaging over areas of 3° by 2.5° , like the one described here, is carried out to improve statistics (increase the number of flashes by a factor of 30) without distorting it. The latter is supported by the fact that this area is still not very large, it regularly fits into the field-of-view of the satellite. And it is still smaller, than the typical sizes of tropical cyclones, or mesoscale convective systems.

Even using this improved statistic, we only have 40180 data points (20 second averages) for the 10 years of observations (0.255 % of time). In contrast to being the most thunderous area on Earth, in 86.5 % of the observations there is no lightning detected. Altogether 40865 flashes were seen by the satellite. Note, that there are certain areas of similar size over the oceans where no flashes were detected during the entire 10 years of observation.

An excerpt of the LIS time series $f(t)$ over the Central African region can be seen in Figure 2.14. The empirical probability density function (PDF) of 20 second flash rates over Central Africa can be seen in Figure 2.15. The distribution is close to a power-law PDF with an exponent -1.30 to -2.04 ,

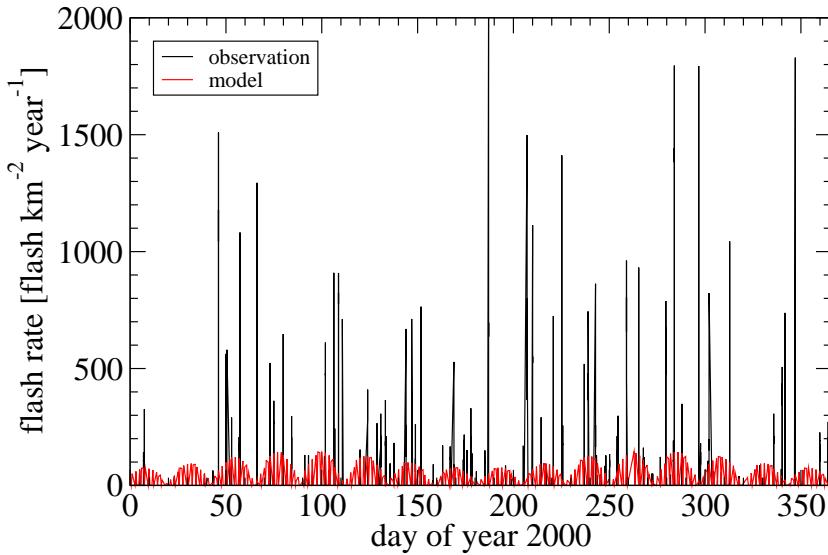


Figure 2.14: Black line: an excerpt of the LIS lightning flash rate time series $f(t)$ with 20 second time bins acquired over the area between 1.5° S and 1.5° N latitudes and 22.0° E and 24.5° E longitudes (Central Africa). Red line corresponds to the model $f_{\text{model}}(t)$ described in the text.

so that the average is barely existing, but the standard deviation is definitely not.

The high temporal resolution and the long record permits to extract the daily- and the yearly cycles as well. These can be seen in Figure 2.16.

The above mentioned properties also permit a detailed spectral analysis. However the data are by far not evenly sampled, therefore the only possible way to extract the spectrum is to use the Lomb algorithm [54]. Several peaks with detailed structure can be seen in the spectrum above a continuum, which seems to be essentially white noise, see Figure 2.17.

The origin of the various peaks in the spectrum can be revealed by a simple model, which goes as follows. The basic well understood fact is, that lightning activity is connected to solar heating. Therefore we consider only the – simplified – daily- and yearly cycles present in the lightning time series. The modelled lightning flash rate time series $f_{\text{model}}(t)$ is the product of these, that is:

$$f_{\text{model}}(t) = a_0 D(t) Y(t) , \quad (2.1)$$

where t is time, a_0 is a constant, $D(t)$ and $Y(t)$ represent the daily- and yearly cycles respectively. Multiplying the cycles instead of adding them is on the one hand practical, because $f(t)$ is a non-negative quantity, on the other hand it has a deeper reason as well. If we think about $f(t)$ as a result

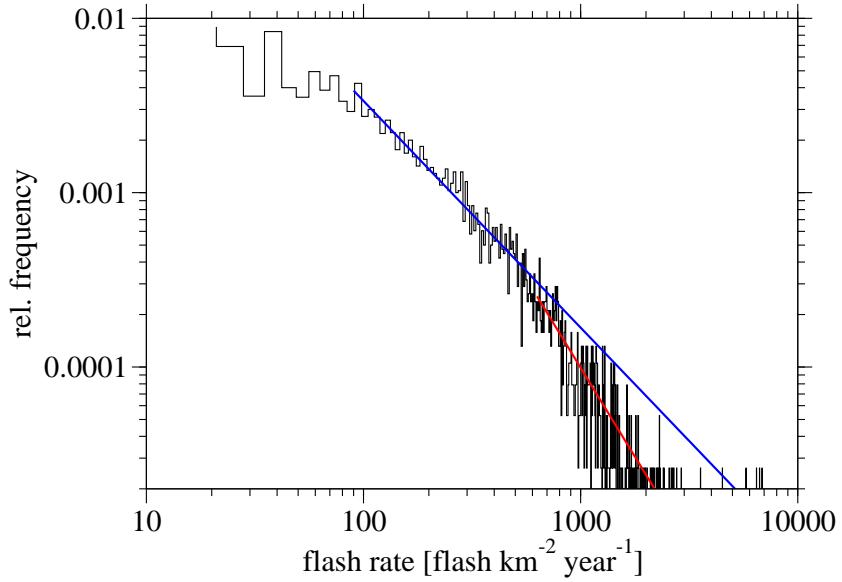


Figure 2.15: Histogram of 20 second lightning flash rate data acquired by LIS over the area between 1.5° S and 1.5° N latitudes and 22.0° E and 24.5° E longitudes (Central Africa). The data with flash rate = 0 ($f = 0$) were omitted to show the statistics. The blue line indicates a power law fit with $p(f) \sim f^{-1.30}$ and the red line a power law fit with $p(f) \sim f^{-2.04}$.

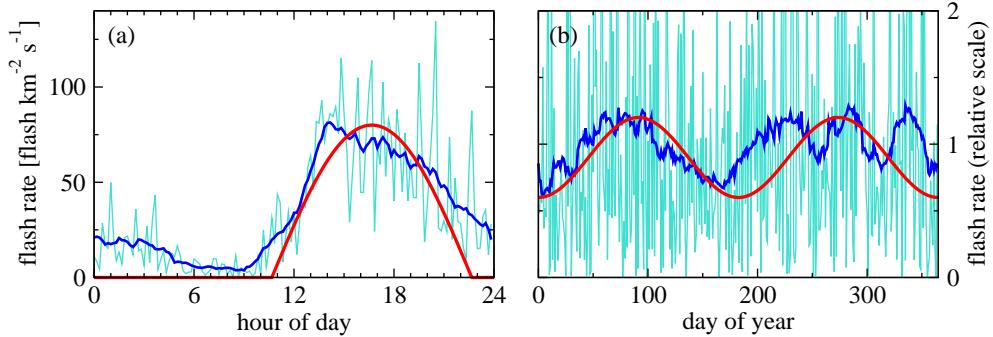


Figure 2.16: (a) Daily cycle of lightning flash rate over the area between 1.5° S and 1.5° N latitudes and 22.0° E and 24.5° E longitudes (Central Africa). The cyan curve represents 10 minute resolution data, the dark blue line corresponds to a 2-hour moving average. (b) Yearly cycle for the same area. The cyan curve represents 1 day resolution data, the dark blue line corresponds to a 31-day moving average. The red lines correspond to the model described in the text. LIS data were used.

of a deterministic variation (solar heating; let it be a constant S for a while) and a stochastic process, which result in a PDF for $f(t)$ close to the one in

Figure 2.15, then the only characteristic value is S . In this case the average flash rate $\langle f(t) \rangle_t$ is proportional to S and the fluctuations of $f(t)$ over a finite period of time are necessary determined by S as well. That is the standard deviation of $f(t)$ is expected to be proportional to S . Therefore it is plausible to divide $f(t)$ by S to isolate and “standardize” the stochastic process. Now let S be a function of time, namely let $S(t)$ be the solar heating, $S(t) = D(t)Y(t)$. In the case of the lightning data – in the analyzes described in the following sections – we used the above reasoning as well, and divided $f(t)$ by $S(t)$ in order to remove the effects due to the solar forcing. Note, that our calculations show that the standard deviations of LIS lightning flash rate time series are indeed proportional to their averages (figure not shown).

In the case of the Central African time series we used a simple half-sine function to describe the daily cycle:

$$D(t) = \begin{cases} \sin\left(\frac{2\pi(t-t_{\text{start},1})}{1 \text{ day}}\right) & , \text{ if } > 0 \\ 0 & , \text{ otherwise,} \end{cases} \quad (2.2)$$

where $t_{\text{start},1}$ is an appropriate time offset.

The yearly cycle is actually a half-yearly cycle at the equator with maximum irradiation during the equinoxes and minimum during the solstices. We modelled this behavior with a simple sine-square function:

$$Y(t) = a_1 + a_2 \sin^2\left(\frac{2\pi(t-t_{\text{start},2})}{1 \text{ year}}\right), \quad (2.3)$$

where a_1 and a_2 are constants, and $t_{\text{start},2}$ is another time offset. The used functions $D(t)$ and $Y(t)$ can be seen as red lines in Figures 2.16a and b respectively.

The resulting $f_{\text{model}}(t)$ model function was sampled in exactly the same time instants as the Central African LIS time series. A sample of both can be seen in Figure 2.14. The spectrum of the original LIS time series and that of the described simple model can be seen in Figure 2.17. It is striking that even this simple model combined with the uneven sampling reproduces the overwhelming majority of the peaks in the original spectrum. From this it follows, that the real information, the effects other than the solar heating are hidden in the white noise, the continuum part of the spectrum.

2.5.2 Global time series of lightning activity

The LIS dataset covers a considerable part of the globe, the equatorial regions between 35° Southern and Northern latitudes, where most lightning activity occurs on Earth, therefore it is possible to extract a “global” lightning time

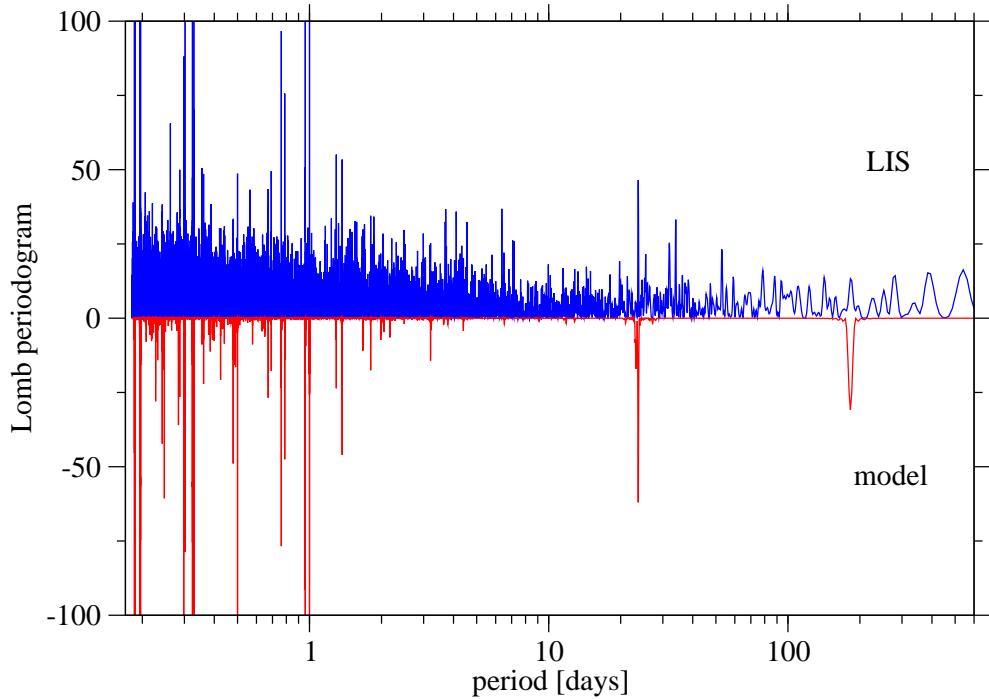


Figure 2.17: Blue curve: Lomb periodogram of LIS lightning flash rate data over the area between 1.5° S and 1.5° N latitudes and 22.0° E and 24.5° E longitudes (Central Africa). The red curve correspond to the Lomb periodogram of the model described in the text. The Lomb periodogram of the model was multiplied by -1 for a better comparison.

series. However, because of the poor but systematic sampling of the satellite, it is not meaningful to average the original data, because this would lead to serious biases.⁵ Therefore we used the following procedure. We extracted the average daily and yearly cycles for $3^{\circ} \times 3^{\circ}$ (latitude/longitude) patches for the whole available area of LIS observations. We applied 2-hour and 1-month running means to the daily and yearly cycles respectively to reduce the effects of low sampling. Then we extracted the flash rate time series for all available $0.5^{\circ} \times 0.5^{\circ}$ (latitude/longitude) primary gridcells and “normalized” them with the corresponding cycles. That is we divided each time series by the corresponding smoothed daily and then yearly cycle. Then we averaged and binned the individual normalized time series. We used 1-hour time bins, and a weighted average with the original average flash rates as weights. This

⁵There exist freely downloadable and highly processed gridded LIS products, but in the case of the time series they involve a ~ 100 day moving average in order to smooth out sampling effects.

way we obtained a LIS “global” lightning flash rate anomaly (multiplicative) time series, which is determined mainly by the areas with high lightning activity. The resulting time series can be seen in Figure 2.18.

We also tried averaging with equal weights, but in this case the resulting “global” time series is dominated by areas with little lightning activity, and single sporadic storms in these areas get overrepresented in the final average. However, in the analyzes carried out in the coming sections we also tested the equally weighted global average. The final conclusions were identical in both cases.

The LDAR data cover only an 80x80 km portion of the globe, however we extracted a lightning flash rate anomaly time series as well. This record obviously shows much larger fluctuations, as the individual storms are still present in the time series because of the lack of real spatial averaging. The LDAR anomaly time series is visible in Figure 2.18 as well.

Because of the good spatial and temporal coverage of the WWLLN data, it is not necessary to use such sophisticated averaging methods, as in the case of the LIS data. Therefore we extracted a simple global average, a global daily and yearly cycle. Again we used the cycles to “normalize” the time series. The result can be seen in Figure 2.18.

Unfortunately the WWLLN has some characteristics, which make the extracted global anomaly time series hardly meaningful. The spatially and temporally changing detection efficiency affects the resulting global long-term series to a great extent, therefore limiting its usage.

The intensity of Schumann resonances is closely related to the energy radiated by the global lightning activity, therefore it might be used as a proxy, and it is certainly worthwhile comparing it to the curves of Figure 2.18. A sample of the Schumann resonance time series averaged for the PKD and PKD1 sites and the LIS anomaly time series can be seen in Figure 2.19. Note that in the case of the LIS anomaly data we used 30-day moving average in Figure 2.19. In the case of the Schumann resonance data we plotted the difference between the 30-day and the 200-day moving averages of the horizontal magnetic field intensities of the first Schumann resonance. This way part of the non-stationarities of the signal were removed.

The comparison yields the strange result, that for certain periods a strikingly good correlation can be observed between the lightning data and the Schumann resonance records, while for other periods there seems to be no connection. We tried to link the periods of good / bad correlation to seasons and various oscillations, like the El Niño-Southern Oscillation (ENSO), but no agreement was found. The periods of alternating good-bad correlation can be due to several reasons including: (1) errors or biases arising from the inaccuracy or low statistical sampling of the measurements, (2) “local” ef-

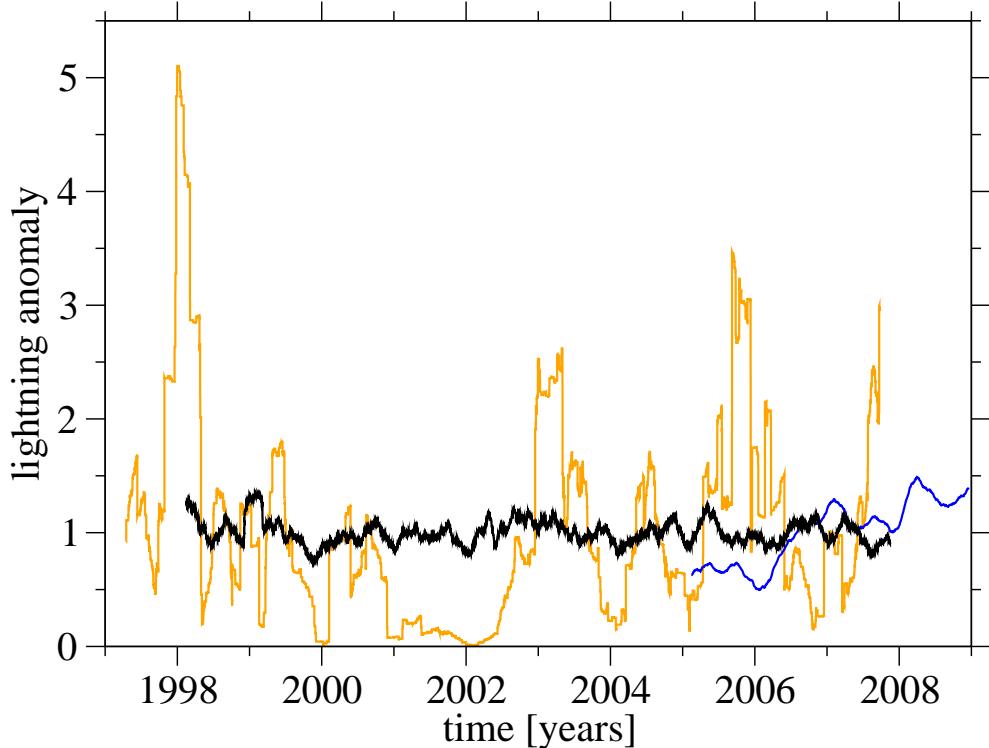


Figure 2.18: Lightning flash rate anomalies based on LIS data (black curve, global between 35° Southern and Northern latitudes), LDAR data (orange curve, only around KSC, Florida, USA) and WWLLN data (blue curve, global). The daily and yearly cycles were removed from the original data by division. Therefore the curves represent multiplicative anomalies. All curves were smoothed using an 89-day moving average.

fects, e.g. the global lightning activity increases due to increased storminess in the Maritime Continent, but the American Schumann resonance detector is more sensitive to the decreasing lightning activity in the South American chimney. However, these are only speculations. More data on Schumann resonance and global lightning activity would be needed to identify the true reasons.

2.5.3 Cosmic radiation time series

Accurate cosmic radiation data are much easier to find, thanks to a global network of neutron monitors initiated around the International Geophysical Year of 1956/1957. The data of all the stations we examined are very highly correlated, showing that cosmic radiation changes are a truly global effect.

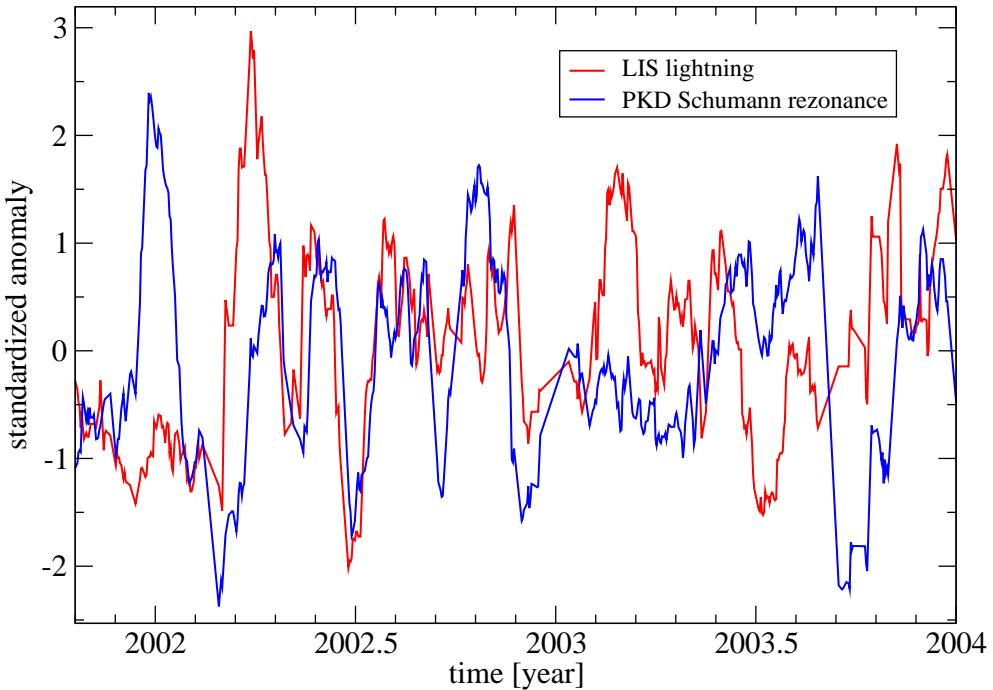


Figure 2.19: A 30-day moving average of the “global” lightning flash rate anomalies based on LIS data (red curve) are compared with Schumann resonance intensities averaged for the PKD and PKD1 sites (blue curve). In the case of the Schumann resonances the difference of the 30-day moving average and the 200-day moving average is shown. The anomalies were further “standardized” by subtracting the average and dividing by the standard deviation.

Of course there are differences in the absolute fluxes due to the different cut-off rigidities and the different elevations of the various stations, but the fluctuations are almost identical. A very good agreement is found also with satellite measurements of proton fluxes. For the cosmic radiation time series see Figure 2.20. The most obvious feature of the CR fluxes is a slow periodic component caused by the 11-year solar cycle. The minimum between 2001-2004 is due to the maximum of solar cycle 23. The superimposed minima are the Forbush decreases also indicated in the graph. Some solar proton events are visible in the proton fluxes, and the event of 20 January 2005 is visible also in the Climax neutron record. The satellite records are clearly non-stationary, showing different periodicities appearing and disappearing. This is due to the changing number of satellites, which provided particle flux data.

The times of the Forbush decreases were extracted based on a commonly used definition: we considered the periods, when the neutron flux decreased

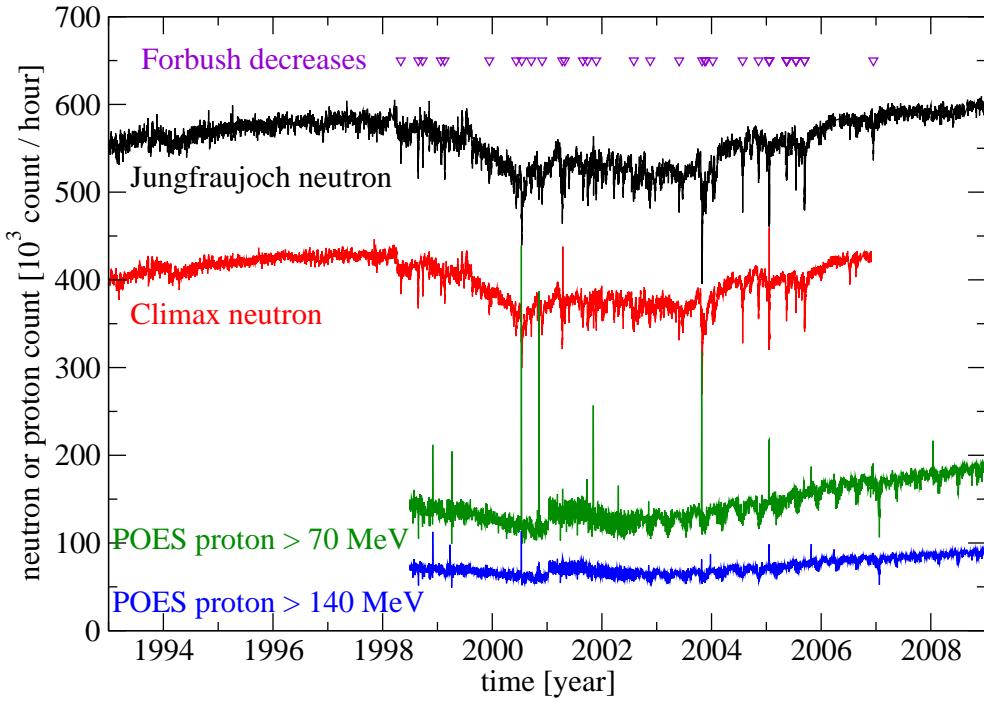


Figure 2.20: Ground based neutron monitor counts are shown from the Jungfraujoch, Switzerland (black curve) and the Climax, Colorado, USA (red curve) stations along with space measurements of the POES satellite network. Omnidirectional proton counts of two channels are shown: protons with energies > 70 MeV (green curve), and protons with energies > 140 MeV (blue curve). Violet triangles indicate the Forbush decreases. The ground based neutron counts are hourly resolution data, the satellite proton fluxes are 2-day moving averages of 16-second resolution data. No correction for the area of the detectors was applied.

more than 5 % below its 89-day running mean value. We used the Climax and the Jungfraujoch datasets as well. Since precise timing was needed, we assigned the time label of a Forbush decrease to the deepest minimum of neutron flux in all cases. The neutron flux anomaly averaged over the identified Forbush decreases can be seen in Figure 2.21a. Anomaly in this case means deviation from average pre and post Forbush decrease levels of neutron flux.

2.6 Lightning activity during Forbush decreases

Forbush decreases are stochastic events from the point of view of the Earth's climate, therefore they provide a very useful tool to search for connections with cosmic rays. In case the hypothesized cosmic ray - lightning connection described in Section 2.3.4 existed, one would expect a decrease of lightning activity during Forbush decreases.

We used the superposed epoch method [148], that is we averaged the lightning time series for the period of Forbush decreases. The results can be seen in Figure 2.21. We did not manage to identify a statistically significant negative peak at the Forbush decreases, which could possibly support the hypothesis of cosmic ray induced lightning initiation.

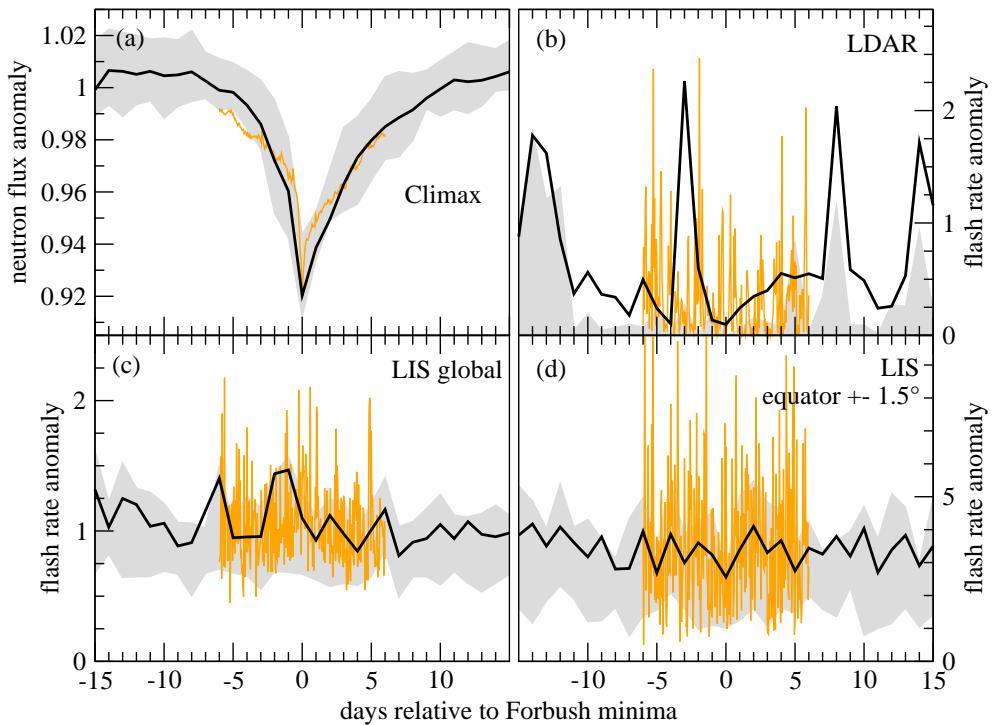


Figure 2.21: Superposed epoch averages of (a) Climax neutron flux anomaly and (b-d) lightning activity over Forbush minima are shown. The lightning flash rate anomaly time series were used, which were extracted from (b) the LDAR dataset, (c) the LIS dataset with no spatial constraint, and (d) the LIS dataset between 1.5° S and 1.5° N latitudes. Black and orange lines correspond to 1-day and 1-hour binned time series respectively. The grey area corresponds to the interquartile range of the 1-day binned data.

By constraining the LIS data over various regions (mainly near the equator), a – non-significant – negative peak can be observed during the Forbush decreases. However by increasing the time resolution from 1 day to 1 hour, this peak gets lost in the noise instead of showing more detail.

By supposing that CRs have an effect on lightning initiation, it might be plausible, that this effect is best observable around the equator. The possible argument supporting this would say, that because of the greatest cutoff rigidities, there is a relatively small flux of cosmic rays in the equatorial region, and a further decreasing flux (due to the Forbush decreases) would already make the presence of CRs a critical ingredient in lightning initiation. This way it is possible, that equatorial regions are hypothetically more sensitive to CR flux variations.

However, the equator where the effect is expected to be strongest based on the above thoughts is the Geomagnetic, not the Geographical Equator. We tested the $\pm 1.5^\circ$ (latitude) region of the geomagnetic equator with the superposed epoch method as well, but the negative peak at the Forbush decreases is even less pronounced, than at the geographical equator (figure not shown).

Furthermore, the space-borne detectors might directly be affected by the changes in CR flux as well, which would produce the same bias as the hypothetical CR-lightning effect. We tried to minimize this possible bias by filtering the data (see Section 2.4.1).

Unfortunately the WWLLN data are less useful in such a study, (1) because of the non-stationarity of the data, and (2) because there were only a few Forbush decreases after 2005 (the start of the WWLLN records), and thus the superposed epoch method gives very poor statistics. Nevertheless we examined the global WWLLN lightning anomaly during the Forbush decreases, but we found no significant correlation (figure not shown).

2.7 Lightning activity during a great Solar Proton Event

Sudden increases of solar cosmic ray flux can sometimes be observed due to solar flares or coronal mass ejections. The CR flux can increase by as much as 4 orders of magnitude in these solar proton events, like in the event of 20 January 2005, which was the second largest solar proton event ever observed [115, 116]. The CR flux increase affects ground levels in these cases as well causing a so-called ground level enhancement, which lasts a few hours. However the effect is not uniform globally, it depends on the relative position

of the Sun and Earth and the interplanetary- and geomagnetic fields. In the case of the 20 January 2005 event, the largest flux increase was observed above Antarctica, but a tenfold increase in neutron flux was measured even at the Jungfraujoch station in Switzerland [115]. Note that the event of 20 January 2005 happened during a Forbush decrease.

In the case of drastically increased CR flux around 20 January 2005, one would expect an increase in global lightning activity as well based on the hypothesis described in Section 2.3.4. The CR flux and lightning time series around 20 January 2005 can be seen in Figure 2.22.

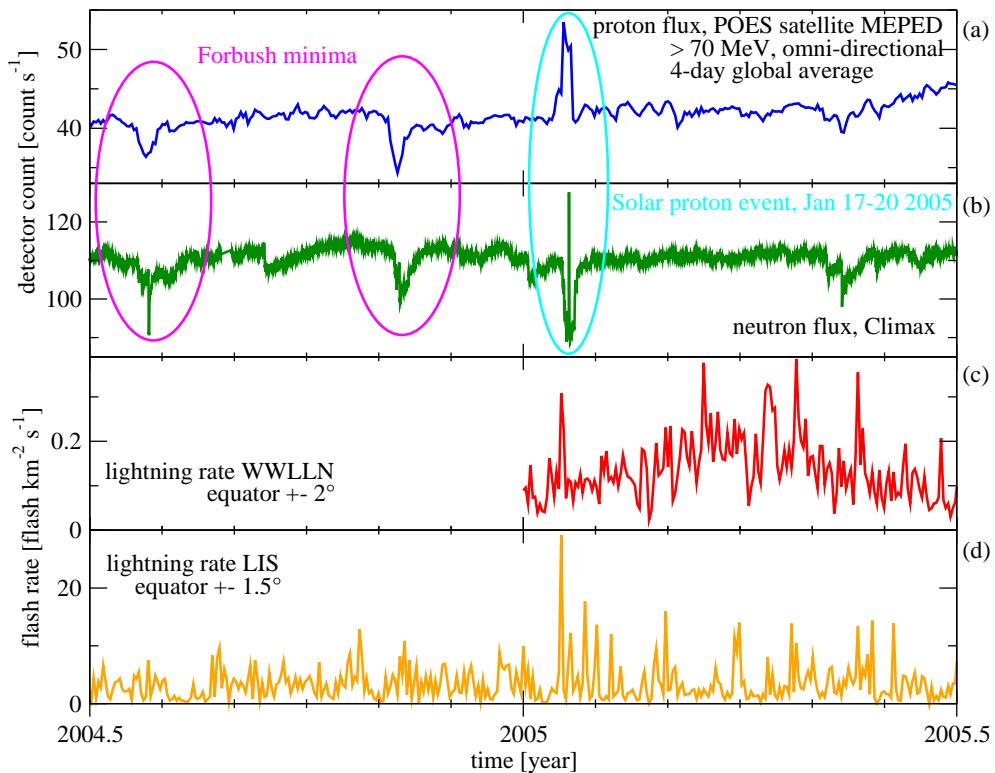


Figure 2.22: Cosmic ray and lightning time series around the solar proton event of 20 January 2005. (a) 4-day running mean of the global average of POES MEPED omni-directional proton count is shown for energies > 70 MeV along with (b) Climax neutron count. Equatorial lightning flash rate anomaly time series are shown based on (c) WWLN data for a $\pm 2^\circ$ region around the equator and (d) LIS data for a $\pm 1.5^\circ$ region around the equator. The largest peak in (d) is on 17 January 2005. Two Forbush decreases are indicated in the upper two panels along with the solar proton event of 20 January 2005.

No increase in global lightning activity can be seen around 20 January 2005 (figure not shown), however a peak in equatorial lightning activity is

visible in both the WWLLN and the LIS records (see Figure 2.22). The problem is that lightning activity peaks on 17 January, thus *preceding* the proton event. Note that there was a smaller solar proton event on 17 January 2005 as well.

The WWLLN data enable us to examine the spatial pattern of the lightning activity increase as well. The 17 January enhancement was due to an increased lightning activity of the South American and the African chimneys, see Figure 2.23. The simple coincidence of two local storms cannot be excluded.

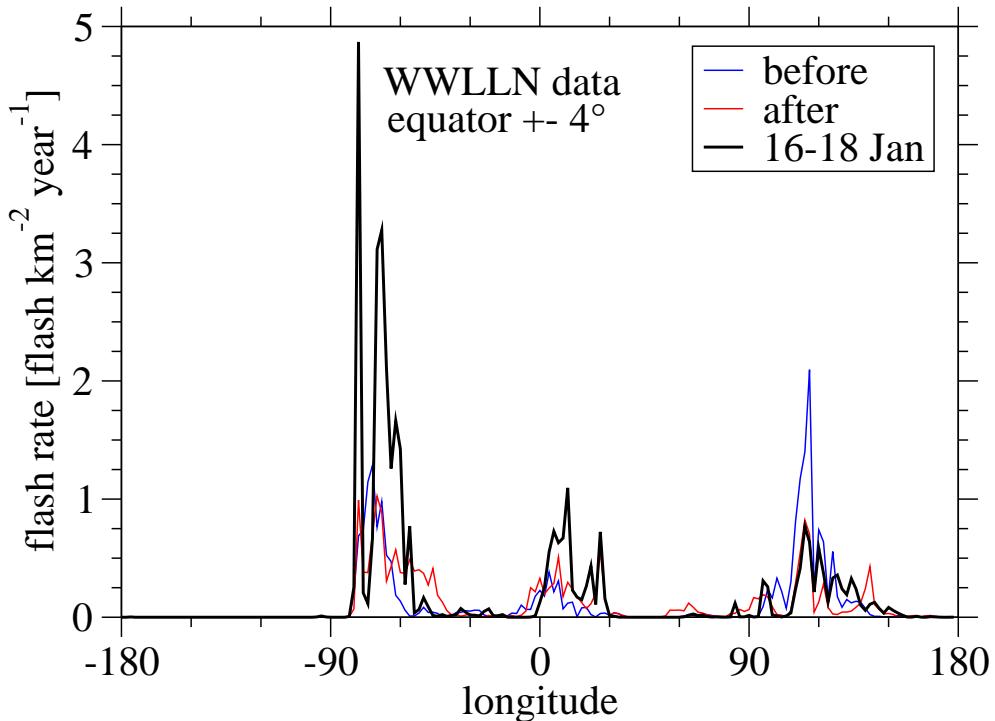


Figure 2.23: Average WWLLN lightning flash rate for a region $\pm 4^\circ$ around the equator for three time periods. Blue curve correspond to 6 days before the solar proton event of 20 January 2005 (10-15 Jan 2005), red curve correspond to 6 days after the event (22-27 Jan 2005), and the heavy black curve corresponds to the period 16-18 Jan 2005.

If the 17 January increase in lightning activity is due to cosmic ray effects, it is plausible to expect, that the much bigger 20 January event should affect lightning activity as well. However this was not observed.

2.8 Further attempts to link cosmic rays and climatic variables

2.8.1 Lightning activity in hurricanes

The lightning activity of hurricanes changes in a wide range showing differences as large as 2 orders of magnitude in flash rate. Recently this lightning activity was proposed to predict wind speeds inside the hurricane a few days ahead [149].

We tried to link the huge differences in lightning activity of similar storms, the category 4-5 hurricanes on the Saffir-Simpson scale to changes in cosmic ray flux, but we found no correlation.

Note that in another study the lightning activity of the storms observed by the OTD and LIS detectors was found to be rather constant [150].

The connection of lightning activity and “storm strength” (e.g. precipitation rate) is far from trivial. Analyzes of the Hungarian Met Office (OMSZ) showed, that lightning activity and precipitation rate are sometimes correlated in thunderstorms above Hungary, but in some cases there seems to be no connection (Ákos Horváth, OMSZ, Siófok; personal communication, and [151]).

2.8.2 Analysis of reflectivity data

All recent studies attempting to capture the effects of CRs on cloud cover (described in Section 2.3.3) used the ISCCP cloud dataset. We instead chose to use the OMI “reflectivity” product, which is based on UV remote sensing. We could not find significant effects of either the Forbush decreases or the solar proton event of 20 January 2005 in the reflectivity time series.

2.9 Summary

Just like our climate in general, the lightning activity in particular seems to be influenced by several factors through a variety of mechanisms, the subtle details of which can seriously affect the macroscopic processes. Though there is serious ongoing research, there are still many unknown details.

We tried to shed more light on the possible global cosmic radiation forcing of lightning activity. Perhaps the most accepted theory of lightning initiation – via the runaway breakdown process – heavily relies on cosmic rays as seed particles for the electron avalanche. Therefore it seems to be plausible to suppose that the variations of the cosmic radiation flux reaching our

atmosphere should be reflected in the global lightning activity as well.

There are plenty of high quality cosmic radiation datasets available and a few global lightning data bases, which raise one's hopes to find a connection between cosmic rays and lightning activity using statistical tools. However none of the lightning datasets satisfies all needs for such studies, therefore we needed to implement various filtering and averaging algorithms. The optical satellite data of the LIS instrument proved to be the most useful for our studies. However, we believe that the WWLLN network has enormous potential, and after further development and several years of continuous operation it would be able to answer numerous questions about lightning activity.

The Forbush decreases of the cosmic radiation flux provide a unique opportunity to link various environmental processes to cosmic radiation, as they are independent from the Earth's climate. However, even by using the superposed epoch method to improve statistics, we did not manage to identify convincing correlation between lightning flash rate anomaly and cosmic rays. We found weak positive correlation between equatorial lightning flash rate anomaly and cosmic ray flux, but the details (geomagnetic effects, fine time resolution data) do not support a causal relation.

During the second greatest Solar Proton Event ever observed, already all lightning detection instruments were in operation, the data of which we analyzed. There is an indication that the equatorial lightning frequency might have responded to the event, however the details do not support this hypothesis either. We also tried to link the large variability of lightning frequency in strong hurricanes to changes in the cosmic ray flux, but we found no correlation.

None of our results support that the changes of the cosmic radiation flux would significantly affect the terrestrial lightning activity. Therefore we can conclude that the large variability observed in lightning activity seems to be mainly determined by the charge separation processes and other local in-cloud effects. Changes in the global "external" cosmic ray forcing can play only a minor role. It is of course possible that cosmic rays are very important in lightning initiation, but the changes in their flux do not seem to have major effects on lightning frequency.

Acknowledgments

I would first of all like to thank my supervisor, Dr. Imre Jánosi for all the invaluable help. He turned my attention to the barely understood physical phenomena of our climate. He showed the way along my research activities by helping to pose the right questions and by sketching the way to answer them. He also guided me through the various difficulties during my research.

I would like to thank Dr. István Csabai for drawing our attention to the so far unanswered questions of lightning activity and for the close cooperation in order to reveal some.

I would like to express my thanks to Dr. Gábor Vattay, Dr. Gábor Papp, Dr. Imre Kondor, Dr. Andrew Sors, Dr. Marc von Hobe and Dr. András Horányi for their generous financial support and for providing me a place to sit down at the Department of Physics of Complex Systems, Loránd Eötvös University, Budapest. I would like to thank that I could spend more than a year in the wonderful institution of the Collegium Budapest. I am also grateful for the kindness and sincere helpfulness of the staff.

This type of research cannot be imagined without the use of high quality data, therefore I would like to thank the Hungarian Meteorological Office, the European Centre for Medium Range Weather Forecasts, Dr. László Varga, Balázs Stelczer, Dr. Flavio Bono and Dr. Eugenio Gutierrez to provide us wind data and Kim Rauenzahn, Sherry Harrison and others at the Global Hydrology Resource Center, Dr. János Lichtenberger, the Northern California Earthquake Data Center, the National Oceanic and Atmospheric Administration, the University of Bern, and the University of New Hampshire for providing us lightning and cosmic radiation data and some information on the raw data.

Extensive data mining is impossible without proper computer support, therefore I would like to thank Norbert Purger and József Stéger for managing my accounts and enduring my requests and questions with patience.

I owe my thanks to my colleagues for the inspiring lectures, talks, debates and for their patience to listen to my problems and their help in solving them. I must not forget about the participants and lecturers of the two summer

schools I visited. I gained a lot of extra motivation and some useful ideas in Konstanz, Germany.

I have to express my big thanks to my family who supported me all the way through my studies.

Appendix A

Maximum likelihood estimation of Weibull and generalized gamma parameters

In the case of the Weibull distribution (see Equation 1.8) the maximum likelihood method leads to the following equations [55]:

$$s_0 = \left(\frac{1}{N} \sum_{i=1}^N x_i^k \right)^{1/k} \quad (\text{A.1})$$

$$\frac{N}{k} + \sum_{i=1}^N \ln \frac{x_i}{s_0} - \sum_{i=1}^N \left(\frac{x_i}{s_0} \right)^k \ln \left(\frac{x_i}{s_0} \right) = 0, \quad (\text{A.2})$$

where $\{x_i\}, i = 1, \dots, N$ is the sample, and s_0 and k are the parameters of the distribution. Equation A.1 can be substituted into Equation A.2 which yields a transcendent equation for parameter k . This can be solved using Newton iterations [54], and substituting k back to Equation A.1 gives parameter s_0 .

In the case of a generalized gamma distributed random variable (wind speed), the maximum likelihood method means maximizing the function L over the parameters s_0, k, ϵ , given a sample of wind speed data $\{x_i\}, i = 1, \dots, N$:

$$\begin{aligned} L(s_0, k, \epsilon) &= \sum_{i=1}^N \ln P_{GG}(x_i; s_0, k, \epsilon) = \\ &= N \ln k - N \ln s_0 - N \ln \Gamma(\epsilon) + (\epsilon k - 1) \sum_{i=1}^N \ln x_i - \\ &- (\epsilon k - 1) N \ln s_0 - \sum_{i=1}^N \left(\frac{x_i}{s_0} \right)^k = \max \end{aligned} \quad (\text{A.3})$$

In the maximum, the following conditions hold: $\frac{\partial L}{\partial s_0} = 0$, $\frac{\partial L}{\partial k} = 0$, $\frac{\partial L}{\partial \epsilon} = 0$, which lead to 3 algebraic equations for the parameters s_0, k, ϵ :

$$s_0 = \left(\frac{1}{\epsilon N} \sum_{i=1}^N x_i^k \right)^{1/k} \quad (\text{A.4})$$

$$\frac{N}{k} + \epsilon \sum_{i=1}^N \ln x_i - \epsilon N \ln s_0 - \sum_{i=1}^N \left(\frac{x_i}{s_0} \right)^k \ln \left(\frac{x_i}{s_0} \right) = 0 \quad (\text{A.5})$$

$$-N \frac{1}{\Gamma(\epsilon)} \frac{d\Gamma(z)}{dz} \Big|_{z=\epsilon} + k \sum_{i=1}^N \ln x_i - k N \ln s_0 = 0 \quad (\text{A.6})$$

It is worth noting, that s_0 can be expressed as a function of the other 2 parameters k and ϵ (Equation A.4), and it can be substituted into Equations A.5 and A.6. Therefore s_0 can be decoupled, and only 2 coupled nonlinear algebraic equations remain to be solved:

$$\begin{aligned} \frac{N}{k} &+ \epsilon \sum_{i=1}^N \ln x_i - \frac{\epsilon N}{k} \left[\ln \left(\sum_{i=1}^N x_i^k \right) - \ln \epsilon - \ln N \right] - \\ &- \epsilon N \left(\sum_{i=1}^N x_i^k \right)^{-1} \sum_{i=1}^N \left\{ x_i^k \left[\ln x_i + \frac{\ln(\epsilon N)}{k} - \frac{\ln(\sum_{i=1}^N x_i^k)}{k} \right] \right\} \end{aligned} \quad (\text{A.7})$$

$$-N \frac{1}{\Gamma(\epsilon)} \frac{d\Gamma(z)}{dz} \Big|_{z=\epsilon} + k \sum_{i=1}^N \ln x_i - N \left[\ln \left(\sum_{i=1}^N x_i^k \right) - \ln \epsilon - \ln N \right] = 0 \quad (\text{A.8})$$

The above set of equations (Eqs. A.7 and A.8) was solved numerically using the Newton-Raphson algorithm [54], with the initial guess obtained from the Weibull fit. The derivative of the gamma function was evaluated numerically. Note, that in the actual calculations a slightly different set of parameters was extracted: $\epsilon k - 1, k, s_0^{-k}$.

Appendix B

Acronyms

AC	alternating current
CCD	Charge-coupled Device
CERN	European Organization for Nuclear Research (Conseil Européen pour la Recherche Nucléaire, in French)
CG	cloud-to-ground lightning flash
CLOUD	Cosmics Leaving OUtdoor Droplets, upcoming experiment at CERN
CR	cosmic rays or cosmic radiation
DC	direct current
ELF	Extremely Low Frequency (3 Hz - 30 Hz) electromagnetic radiation
ELTE	Loránd Eötvös University (Eötvös Loránd Tudományegyetem, in Hungarian), Budapest, Hungary
ENSO	El Niño-Southern Oscillation
EOS	Earth Observing System, a NASA project including several satellites
FFT	Fast Fourier Transformation
GCR	Galactic Cosmic Rays
GG	generalized gamma (distribution)
GHRC	Global Hydrology Resource Center, Huntsville, USA
GN	generalized normal (distribution)
IC	intracloud lightning flash
IGY	International Geophysical Year 1956/1957
IID	independent identically distributed (random variable)

ISCCP	International Satellite Cloud Climatology Project
ITCZ	intertropical convergence zone
KSC	Kennedy Space Center, Florida, USA
LDAR	Lightning Detection And Ranging, a lightning mapping array at the KSC, Florida
LIS	Lightning Imaging Sensor, lightning detection instrument on board the TRMM satellite
MEPED	Medium Energy Proton and Electron Detector, instrument on board the POES satellites
NASA	National Aeronautics and Space Administration, Washington, DC, USA
NCEDC	Northern California Earthquake Data Center, ..., USA
NOAA	National Oceanic and Atmospheric Administration, Washington, DC, USA
OMI	Ozone Monitoring Instrument on board the EOS Aura satellite
OMSZ	Hungarian Met Office (Országos Meteorológiai Szolgálat, in Hungarian), Budapest, Hungary
OTD	Optical Transient Detector, a lightning detection instrument on board the OrbView-1 (formerly MicroLab-1) satellite
PDF	probability density function
POES	Polar Orbiting Environmental Satellites of the NASA
SEP	Solar Energetic Particles
TED	Total Energy Detector, instrument on board the POES satellites
TLE	Transient Luminous Events (e.g. Jets, Sprites, Elves)
TRMM	Tropical Rainfall Measuring Mission satellite (NASA)
ULF	Ultra Low Frequency (300 Hz - 3 kHz) electromagnetic radiation
USA	United States of America
USPLN	United States Precision Lightning Network, Andover, USA
UT	Universal Time
UV	ultraviolet
VHF	Very High Frequency (30 MHz - 300 MHz) electromagnetic radiation
VLF	Very Low Frequency (3 kHz - 30 kHz) electromagnetic radiation
WWLLN	World Wide Lightning Location Network: global VLF ground network for location of lightning strikes

Bibliography

- [1] World Wind Energy Association. World Wind Energy Report 2008. 2009 http://www.wwindea.org/home/images/stories/worldwindenergyreport2008_s.pdf.
- [2] <http://www.ewea.org> (The European Wind Energy Association).
- [3] Goldemberg J. Pumping Renewables. *Nature* 2008;456:26.
- [4] Archer CL, Jacobson MZ. Evaluation of global wind power. *J Geophys Res* 2005;110:D12110.
- [5] Lu X, McElroy MB, Kiviluoma J. Global potential for wind-generated electricity. *Proc Nat Acad Sci* 2009; <http://www.pnas.org/cgi/doi/10.1073/pnas.0904101106>
- [6] Landberg L, Myllerup L, Rathmann O, Petersen EL, Jørgensen BH, Badger J, Mortensen NG. Wind Resource Estimation – An Overview. *Wind Energy* 2003;6:261.
- [7] Kalnay E, et al. The NCEP/NCAR 40-Year Reanalysis Project. *Bull Am Meteorol Soc* 1996;77:437.
- [8] Kanamitsu M, Ebisuzaki W, Woollen J, Yang S-K, Hnilo JJ, Fiorino M, Potter GL. NCEP–DOE AMIP-II Reanalysis (R-2). *Bull Am Meteorol Soc* 2002;83:1631.
- [9] Uppala SM, Källberg PW, Simmons AJ, Andrae U, Da costa Bechtold U, Fiorino M, et al. The ERA-40 reanalysis. *Q J Roy Meteorol Soc* 2005;131:2961-3012.
- [10] <http://www.ecmwf.int/research/era/do/get/era-interim>
- [11] Francis JA. Validation of reanalysis upper-level winds in the Arctic with independent rawinsonde data. *Geophys Res Lett* 2002;29:1315.

- [12] Goswami BN, Sengupta D. A note on the deficiency of NCEP/NCAR reanalysis surface winds over the equatorial Indian Ocean. *J Geophys Res* 2003;108:C3124.
- [13] Kumar MS, Anandan VK. Comparison of the NCEP Reanalysis II winds with those observed over a complex terrain in lower atmospheric boundary layer. *Geophys Res Lett* 2009;36:L01805. doi:10.1029/2008GL036246.
- [14] Landberg L. The availability and variability of the European wind resource. *Int J Sust Energy* 1997;18:313-320.
- [15] Ackermann T, Söder L. An overview of wind energy-status 2002. *Renew Sust Energy Rev* 2002;6:67-127.
- [16] Holttinen H. Hourly wind power variations in the Nordic countries. *Wind Energy* 2005;8:173-195.
- [17] Giebel G. A variance analysis of the capacity displaced by wind energy in Europe. *Wind Energy* 2007;10:69-79.
- [18] Lund H. Large-scale integration of wind power into different energy systems. *Energy* 2005;30:2402-2412.
- [19] Marris E. Upgrading the grid. *Nature* 2008;454:570-573.
- [20] Papaefthymiou G, Schavemaker PH, van der Sluis L, Kling WL, Kurowicka D, Cooke RM. Integration of stochastic generation in power systems. *Int J Elec Power* 2006;28:655-667.
- [21] Katsaprakakis DA, Papadakis N, Christakis DG, Zervos A. On the wind power rejection in the islands of Crete and Rhodes. *Wind Energy* 2007;10:415-434.
- [22] Dondi P, Bayoumib D, Haederlic C, Julian D, Suter M. Network integration of distributed power generation. *J Power Sources* 2002;106:1-9.
- [23] Pepermans G, Driesen J, Haeseldonckx D, Belmans R, Dhaeseleer W. Distributed generation: definition, benefits and issues. *Energy Policy* 2005;33:787-798.
- [24] Burton T, Sharpe D, Jenkins N, Bossanyi E. *Wind energy handbook*. Chichester: John Wiley & Sons; 2001.

- [25] Patel MR. Wind and solar power systems: Design, analysis, and operation. 2nd ed. Oxford: Taylor & Francis; 2005.
- [26] Hau E. Wind turbines: Fundamentals, technologies, application, economics. 2nd ed. Berlin: Springer; 2006.
- [27] Conradsen K, Nielsen LB. Review of Weibull statistics for estimation of wind speed distributions. *J Appl Meteor* 1984;23:1173-83.
- [28] Seguro JV, Lambert TW. Modern estimation of the parameters of the Weibull wind speed distribution for wind energy analysis. *J Wind Eng Ind Aerodyn* 2000;85:75-84.
- [29] Ramírez P, Carta JA. Influence of the data sampling interval in the estimation of the parameters of the Weibull wind speed probability density distribution. A case study. *Energy Convers Manage* 2005;46:2419-38.
- [30] Monahan AH. The probability distribution of sea surface wind speeds. Part I: Theory and SeaWinds observations. *J Climate* 2006;19:497-520.
- [31] Van der Auwera L, De Meyer F, Malet LM. The use of the Weibull three-parameter model for estimating mean wind power densities. *J Appl Meteor* 1980;19:819-25.
- [32] Luna RE, Church HW. Estimation of long-term concentrations using a "universal" wind speed distribution. *J Appl Meteor* 1974;13:910-16.
- [33] Justus CG, Hargraves WR, Mikhail A, Graber D. Methods for estimating wind speed frequency distributions. *J Appl Meteor* 1978;17:350-3.
- [34] Brown BG, Katz RW, Murphy AH. Time series models to simulate and forecast wind speed and wind power. *J Appl Meteor* 1984;23:1184-95.
- [35] Bardsley WE. Note on the use of the inverse Gaussian distribution for wind energy applications. *J Appl Meteor* 1980;19:1126-30.
- [36] Bauer E. Characteristic frequency distributions of remotely sensed in situ and modelled wind speeds. *Int J Climatol* 1996;16:1087-102.
- [37] Li M, Li X. MEP type distribution function: a better alternative to Weibull function for wind speed distributions. *Renew Energy* 2005;30:1221-40.
- [38] Akpinar S, Akpinar EK. Wind energy analysis based on maximum entropy principle (MEP)-type distribution function. *Energy Convers Manage* 2007;48:1140-49.

- [39] Kiss P, Jánosi IM. Comprehensive empirical analysis of ERA-40 surface wind speed distribution over Europe. *Energy Convers Manage* 2008;49:2142-2151.
- [40] Cohen AC, Whitten BJ. Parameter estimation in reliability and life span models. Boca Raton: CRC Press; 1988.
- [41] Hantel M, editor. Observed global climate. Berlin: Springer; 2005.
- [42] http://www.ecmwf.int/research/era/ERA-40_Atlas/docs/section_B/parameter_10mvwwi.html (ERA-40 Atlas)
- [43] Monahan AH. The probability distribution of sea surface wind speeds. Part II: Dataset intercomparison and seasonal variability. *J Climate* 2006;19:521-34.
- [44] Dodson B. The Weibull analysis handbook. 2nd ed. Milwaukee: ASQ Quality Press; 2006.
- [45] Coles S. An introduction to statistical modeling of extreme values. New York: Springer; 2001.
- [46] Justus CG, Hargraves WR, Yalcin A. Nationwide assessment of potential output from wind-powered generators. *J Appl Meteor* 1976;15:673-8.
- [47] Takle ES, Brown JM. Note on the use of Weibull statistics to characterize wind-speed data. *J Appl Meteor* 1978;17:556-9.
- [48] Stewart DA, Essenwanger OM. Frequency distribution of wind speed near the surface. *J Appl Meteor* 1978;17:1633-42.
- [49] Tuller SE, Brett AC. The characteristics of wind velocity that favour the fitting of Weibull distribution in wind speed analysis. *J Appl Meteor* 1984;23:124-34.
- [50] Garcia A, Torres JL, Prieto E, De Francisco A. Fitting wind speed distributions: a case study. *Sol Energy* 1998;62:139-44.
- [51] Celik AN. Assessing the suitability of wind speed probability distribution functions based on wind power density. *Ren Energy* 2003;28:1563-74.
- [52] Chang TJ, Wu YT, Hsua HY, Chub CR, Liao CM. Assessment of wind characteristics and wind turbine characteristics in Taiwan. *Ren Energy* 2003;28:851-71.

- [53] Erickson DJ, Taylor JA. Non-Weibull behavior observed in a model-generated global surface wind field frequency distribution. *J Geophys Res C* 1989;94:12693-8.
- [54] Press WH, Flannery BP, Teukolsky SA, Vetterling WT. Numerical recipes in C. 2nd ed. Cambridge: Cambridge University Press; 1992.
- [55] Ghosh A. A FORTRAN program for fitting Weibull distribution and generating samples. *Computers Geosci* 1999;25:729-38.
- [56] Dadpay A, Soofi ES, Soyer R. Information measures for generalized gamma family. *J Economet* 2007;138:568-85.
- [57] Archer CL, Jacobson MZ. Spatial and temporal distributions of U.S. winds and wind power at 80 m derived from measurements. *J Geophys Res* 2003;108:D4289.
- [58] Lazarus SM, Bewley J. Evaluation of a wind power parameterization using tower observations. *J Geophys Res* 2005;110:D07102.
- [59] Emeis S, Türk M. in Wind Energy. Proceedings of the Euromech Colloquium (Springer-Verlag, Berlin, 2007), p. 61.
- [60] Dhanju A, Whitaker P, Kempton W. Assessing offshore wind resources: An accessible methodology. *Renew Energy* 2008;33:55-64.
- [61] Mengelkamp H-T. Wind Climate Simulation over Complex Terrain and Wind Turbine Energy Output Estimation. *Theor Appl Climatol* 1999;63:129.
- [62] Pryor SC, Schoof JT, Barthelmie RJ. Empirical downscaling of wind speed probability distributions. *J Geophys Res* 2005;110:D19109.
- [63] Hsu SA, Meindl EA, Gilhousen DB. Determining the power-law wind-profile exponent under near-neutral stability conditions at sea. *Journal of Applied Meteorology* 1994; 33:757-765.
- [64] Farrugia RN. The wind shear exponent in a Mediterranean island climate. *Renewable Energy* 2003; 28: 647-653. DOI: 10.1016/S0960-1481(02)00066-6
- [65] Sedefian L. On the vertical extrapolation of mean wind power density. *Journal of Applied Meteorology* 1980; 19: 488-493.

- [66] Petersen EL, Mortensen NG, Landberg L, Højstrup J, Frank HP. Wind power meteorology. Part I: climate and turbulence. *Wind Energy* 1998;1:2-22.
- [67] Gryning SE, Jørgensen H, Larsen S, Batchvarova E. The wind profile up to 300 meters over flat terrain. *J Phys: Conf Ser* 2007;75:012066.
- [68] Garratt JR. *The Atmospheric Boundary Layer*. Cambridge University Press: Cambridge, 1994; 1-334.
- [69] Tennekes H. The logarithmic wind profile. *Journal of the Atmospheric Sciences* 1973; 30: 234-238.
- [70] Carl DM, Tarbell TC, Panofsky HA. Profiles of wind and temperature from towers over homogeneous terrain. *Journal of the Atmospheric Sciences* 1973; 30: 788-794.
- [71] Kiss P, Varga L, Jánosi IM. Comparison of wind power estimates from ECMWF reanalyses with direct turbine measurements. *J Renew Sust Energy* 2009;1:033105. doi: 10.1063/1.3153903
- [72] Liszka T. An interpolation method for an irregular net of nodes. *Int J Numer Methods Eng* 1984;20:1599.
- [73] Pryor SC; Barthelmie RJ. Comparison of potential power production at on- and offshore sites. *Wind Energy* 2001; 4: 173-181. DOI: 10.1002/we.54
- [74] Henderson AR, Morgan C, Smith B, Srensen HC, Barthelmie RJ, Boesmans B. Offshore wind energy in Europe - A review of the state-of-the-art. *Wind Energy* 2003; 6: 35-52. DOI: 10.1002/we.82
- [75] Barthelmie R, Hansen OF, Enevoldsen K, Højstrup J, Frandsen S, Pryor S, Larsen S, Motta M, Sanderhoff P. Ten Years of Meteorological Measurements for Offshore Wind Farms. *J. Sol. Energy Eng.* 2005;127:170-176.
- [76] Tar K, Makra L, Horváth S, Kircsi A. Temporal change of some statistical characteristics of wind speed over the Great Hungarian Plain. *Theor Appl Climatol* 2001;69:69-79.
- [77] Bartholy J, Radics K, Bohoczky F. Present state of wind energy utilisation in Hungary: policy, wind climate, and modelling studies. *Renew Sust Energy Rev* 2003;7:175-186.

- [78] Varga L, Korényi Z, Hirsch T. Balancing Energy Planning in Wind Generation using Probabilistic Weather Prediction. WSEAS Trans Power Syst 2006;1:1243.
- [79] Weidinger T, Kiss A, Gyöngyösi AZ, Krassován K, Papp B. Uncertainty of wind energy estimation. In Peinke J, Schaumann P, Barth S, Eds. Wind Energy. Springer-Verlag: Berlin, Germany, 2007; pp. 167-171.
- [80] Radics K, Bartholy J. Estimating and modelling the wind resource of Hungary. Renew Sust Energy Rev 2008;12:874-882.
- [81] Kiss P, Varga L, Jánosi IM. Wind power availability over Europe. Acta Pericemonologica Debrecina 2008;3:106-116.
- [82] Kiss P, Jánosi IM. Limitations of wind power availability over Europe: a conceptual study. Nonlin Proc Geophys 2008;15:803-813.
- [83] European Wind Energy Association: Wind Energy – The Facts. <http://www.ewea.org> 2004.
- [84] Troen I, Petersen EL. European Wind Atlas. Risoe National Laboratory, Roskilde, 1989.
- [85] Pardalos PM, Romeijn HE, Tuy H. Recent developments and trends in global optimization. Journal of Computational and Applied Mathematics 2000; 124: 209-228. DOI: 10.1016/S0377-0427(00)00425-8
- [86] Wikipedia – Wind Power http://en.wikipedia.org/wiki/Wind_power
- [87] Energy Information Administration - International Energy Statistics <http://www.eia.doe.gov/emeu/international/contents.html>
- [88] Kiss P, Varga L, Jánosi IM. Contrasting Electricity Demand with Wind Power Supply: Case Study in Hungary. Energies 2009;2(4):839-850. doi:10.3390/en20400839 (“Wind Energy” special issue).
- [89] Makra L, Tar K, Horváth S. Some statistical characteristics of the wind energy over the Great Hungarian plain. Int J Ambient Energy 2000;21:85-96.
- [90] Lu X, McElroy MB, Kiviluoma J. Global potential for wind-generated electricity. Proc Natl Acad Sci 2009;106:10933-10938.
- [91] Archer CL, Jacobson MZ. Supplying baseload power and reducing transmission requirements by interconnecting wind farms. J Appl Meteorol Climatol 2007;46:1701-1717.

- [92] http://www.met.hu/omsz.php?almenu_id=climate&pid=climate_Hw&pri=1&szt=Szel (accessed on 13 August 2009)(in Hungarian).
- [93] Edwards PJ. Comments on "Run duration analysis of surface wind speeds for wind energy applications". *J Appl Meteorol* 1980;19:757-758.
- [94] Edwards PJ, Hurst RB. Level-crossing statistics of the horizontal wind speed in the planetary surface boundary layer. *Chaos* 2001;11:611-618.
- [95] Jacobson MJ. Review of solutions to global warming, air pollution, and energy security. *Energy Environ Sci* 2009;2:148-173.
- [96] Sovacool BK. The problem with the "portfolio approach" in American energy policy. *Policy Sci* 2008;41:245-261.
- [97] Christian HJ, et al. Global frequency and distribution of lightning as observed from space by the Optical Transient Detector. *J Geophys Res* 2003;108:D1 4005, doi:10.1029/2002JD002347.
- [98] Gurnett DA, Zarka P, Manning R, Kurth WS, Hospodarsky GB, Averkamp TF, Kaiser ML, Farrell WM. Non-detection at Venus of high-frequency radio signals characteristic of terrestrial lightning *Nature* 2001;409:313-315, doi:10.1038/35053009.
- [99] Yair Y, Fischer G, Simões F, Renno N, Zarka P. Updated Review of Planetary Atmospheric Electricity. in Leblanc F, Aplin KL, Yair Y, Harrison RG, Lebreton JP, Blanc M. *Planetary Atmospheric Electricity*. Springer; 2008.
- [100] Price C. Global Thunderstorm Activity. in Füllekrug M, Mareev EA, Rycroft MJ. *Sprites, Elves and Intense Lightning Discharges*. Springer, Dordrecht; 2006.
- [101] Boccippio DJ, Cummins KL, Christian HJ, Goodman SJ. Combined satellite- and surface based estimation of the intracloud-cloud-to-ground lightning ratio over the continental United States. *Mon Weather Rev* 2001;129:108-122
- [102] Aplin KL, Harrison RG, Rycroft MJ. Investigating Earths Atmospheric Electricity: a Role Model for Planetary Studies. in Leblanc F, Aplin KL, Yair Y, Harrison RG, Lebreton JP, Blanc M. *Planetary Atmospheric Electricity*. Springer; 2008.

- [103] Rycroft MJ, Harrison RG, Nicoll KA, Mareev EA. An Overview of Earths Global Electric Circuit and Atmospheric Conductivity. in Leblanc F, Aplin KL, Yair Y, Harrison RG, Lebreton JP, Blanc M. Planetary Atmospheric Electricity. Springer; 2008.
- [104] Stolzenburg M, Marshall TC. Electric Field and Charge Structure in Lightning-Producing Clouds. in Betz HD, Schumann U, Laroche P. Lightning: Principles, Instruments and Applications. Review of Modern Lightning Research. Springer; 2009.
- [105] Sátori G, Mushtak V, Williams E. Schumann Resonance Signatures of Global Lightning Activity. in Betz HD, Schumann U, Laroche P. Lightning: Principles, Instruments and Applications. Review of Modern Lightning Research. Springer; 2009.
- [106] Roussel-Dupré R, Colman JJ, Symbalisty E, Sentman D, Pasko VP. Physical Processes Related to Discharges in Planetary Atmospheres. in Leblanc F, Aplin KL, Yair Y, Harrison RG, Lebreton JP, Blanc M. Planetary Atmospheric Electricity. Springer; 2008.
- [107] Gurevich AV, Zybin KP. Runaway Breakdown and the Mysteries of Lightning. Phys Today 2005;58:37-43
- [108] Gurevich AV, Zybin KP, Roussel-Dupré RA. Lightning initiation by simultaneous effect of runaway breakdown and cosmic ray showers. Phys Lett A 1999;254:79-87
- [109] Treumann RA, Kłos Z, Parrot M. Physics of Electric Discharges in Atmospheric Gases: An Informal Introduction. in Leblanc F, Aplin KL, Yair Y, Harrison RG, Lebreton JP, Blanc M. Planetary Atmospheric Electricity. Springer; 2008.
- [110] Baba Y, Rakov VA. Present Understanding of the Lightning Return Stroke. in Betz HD, Schumann U, Laroche P. Lightning: Principles, Instruments and Applications. Review of Modern Lightning Research. Springer; 2009.
- [111] Williams ER. Predictable lightning paths? Nature Geo 2008;1:216-217
- [112] Krehbiel PR, Riousset JA, Pasko VP, Thomas RJ, Rison W, Stanley MA, Edens HE. Upward electrical discharges from thunderstorms. Nature Geo 2008;1:233-237

- [113] Bazilevskaya GA, Krainev MB, Makhmutov VS. Effects of cosmic rays on Earth's environment. *J Atm Sol-Terr Phys* 2000;62:1577-1586
- [114] The Pierre Auger Collaboration. Correlation of the Highest-Energy Cosmic Rays with Nearby Extragalactic Objects. *Science* 9 2007;318:938-943. DOI: 10.1126/science.1151124
- [115] Bütkofer R, Flückiger EO, Desorgher L, Moser MR, Pirard B. The solar cosmic ray ground-level enhancements on 20 January 2005 and 13 December 2006. *Adv Spc Res* 2009;43:499-503
- [116] Vashenyuk EV, Balabin GAYuV, Bazilevskaya GA, Makhmutov VS, Stozhkov YuI, Svirzhevsky NS. Solar Particle Event 20 January, 2005 on stratosphere and ground level observations. 29th International Cosmic Ray Conf Proc 2005;00:101-104
- [117] Harrison RG, Tammet H. Ions in the Terrestrial Atmosphere and Other Solar System Atmospheres in Leblanc F, Aplin KL, Yair Y, Harrison RG, Lebreton JP, Blanc M. Planetary Atmospheric Electricity. Springer; 2008.
- [118] Bazilevskaya GA, Usoskin IG, Flückiger EO, Harrison RG, Desorgher L, Bütkofer R, Krainev MB, Makhmutov VS, Stozhkov YI, Svirzhevskaya AK, Svirzhevsky NS, Kovaltsov GA. Cosmic Ray Induced Ion Production in the Atmosphere in Leblanc F, Aplin KL, Yair Y, Harrison RG, Lebreton JP, Blanc M. Planetary Atmospheric Electricity. Springer; 2008.
- [119] Dorman L. Cosmic Rays in Magnetosphere of the Earth and other Planets. Springer; 2009.
- [120] Intergovernmental Panel on Climate Change. Third Assessment Report. 2001. Section 6.11 http://www.grida.no/publications/other/ipcc_tar/
- [121] Herschel W. Observations Tending to Investigate the Nature of the Sun, in Order to Find the Causes and Symptoms of its Variable Emission of Light and Heat... *Philos Trans R Soc* 1801;91:261-331
- [122] Sloan T, Wolfendale AW. Testing the proposed causal link between cosmic rays and cloud cover. *Environ Res Lett* 2008;3:024001(6pp)
- [123] Mauas PJD, Flamenco E, Buccino AP. Solar Forcing of the Stream Flow of a Continental Scale South American River. *Phys Rev Lett* 2008;101:168501

- [124] Eddy JA. The Maunder Minimum Sci 1976;192:1189-1202
- [125] Carslaw KS, Harrison RG, Kirkby J. Cosmic Rays, Clouds, and Climate. Science 2002;298:1732-1737
- [126] Carslaw K. Cosmic rays, clouds and climate. Nature 2009;460:332-333
- [127] Svensmark H, Friis-Christensen E. Variation of cosmic ray flux and global cloud coverage - a missing link in solar-climate relationships. J Atmos Sol-Terr Phys 1997;59:1225-1232
- [128] Marsh ND, Svensmark H. Low Cloud Properties Influenced by Cosmic Rays. Phys Rev Lett 2000;85:5004-5007
- [129] Marsh ND, Svensmark H. Cosmic rays, Clouds, and Climate. Space Science Reviews 2000;94:215-230
- [130] Tinsley BA, Burns GB, Zhou L. The role of the global electric circuit in solar and internal forcing of clouds and climate. Adv Spc Res 2007;40:1126-1139
- [131] Perry CA. Evidence for a physical linkage between galactic cosmic rays and regional climate time series. Adv Spc Res 2007;40:353-364
- [132] Pudovkin MI, Veretenenko SV. Cloudiness decreases associated with Forbush-decreases of galactic cosmic rays. J Atm Sol-Terr Phys 1995;57:1349-1355
- [133] Márcz F. Short-term changes in atmospheric electricity associated with Forbush decreases. J Atm Sol-Terr Phys 1997;59:975-982
- [134] Lethbridge MD. Cosmic Rays and Thunderstorm Frequency. Geophys Res Lett 1981;8:521-522
- [135] Lethbridge MD. Thunderstorms, Cosmic Rays, and Solar-Lunar Influences. J Geophys Res 1990;95:13645-13649
- [136] Kiss P, Csabai I, Lichtenberger J, Jánosi IM. Kozmikus sugárzás, időjárás, klíma: hol a hiányzó láncszem? Fizikai Szemle 2009;59:238-243
- [137] Boccippio DJ, Heckman S, Goodman SJ. A diagnostic analysis of the Kennedy Space Center LDAR network 1. Data characteristics. J Geophys Res 2000;106:4769-4786

- [138] Rodger CJ, Brundell JB, Holzworth RH, Lay EH. Growing Detection Efficiency of the World Wide Lightning Location Network. COUPLING OF THUNDERSTORMS AND LIGHTNING DISCHARGES TO NEAR-EARTH SPACE: Proceedings of the Workshop. AIP Conf Proc 2009;1118:15-20
- [139] Christian HJ, Blakeslee RJ, Goodman SJ, Mach DA, Stewart MF, Buechler DE, Koshak WJ, Hall JM, Boeck WL, Driscoll KT, Boccippio DJ. The Lightning Imaging Sensor. ICAE conference 1999; http://thunder.msfc.nasa.gov/bookshelf/pubs/LIS_ICAE99_Print.pdf
- [140] Boccippio DJ, Koshak, WJ, Blakeslee, RJ. Performance Assessment of the Optical Transient Detector and Lightning Imaging Sensor. Part I: Predicted Diurnal Variability. J Atmos Ocean Tech 2002;19:1318-1332
- [141] Thomas RJ, Krehbiel PR, Rison W, Hamlin T, Boccippio DJ, Goodman SJ, Christian HJ. Comparison of ground-based 3-dimensional lightning mapping observations with satellite-based LIS observations in Oklahoma Geophys Res Lett 2000;27:1703-1706
- [142] Koshak WJ, Krider EP, Murray N, Boccippio DJ. Lightning Charge Retrievals: Dimensional Reduction, LDAR Constraints, and a First Comparison with LIS Satellite Data J Atmos Ocean Tech 2007;24:1817-1838
- [143] Krider EP, Koshak WJ. LIS Validation Studies Using Lightning at the KSC-ER. NRA-97-MTPE-03 ANNUAL REPORT 2001;
- [144] Christian HJ, Blakeslee RJ, Goodman SJ, Mach DM. Algorithm Theoretical Basis Document (ATBD) for the Lightning Imaging Sensor (LIS). NASA 2000; http://eospso.gsfc.nasa.gov/eos_homepage/for_scientists/atbd/docs/LIS/atbd-lis-01.pdf
- [145] Sekiguchi M, Hayakawa M, Nickolaenko AP, Hobara Y. Evidence on a link between the intensity of Schumann resonance and global surface temperature. Ann Geophys 2006;24:1809-1817
- [146] Flückiger EO, Bütkofer R. Swiss neutron monitors and cosmic ray research at Jungfraujoch. Adv Space Res 2009; in press
- [147] Moraal H, Belov A, Clem JM. Design and co-Ordination of Multi-Station International Neutron Monitor Networks. Space Sci Rev 2000;93:285-303

- [148] Singh YP, Badruddin. Statistical considerations in superposed epoch analysis and its applications in space research. *J Atm Sol-Terr Phys* 2006;68:803-813.
- [149] Price C, Asfur M, Yair Y. Maximum hurricane intensity preceded by increase in lightning frequency. *Nature Geo* 2009;2:329-332
- [150] Williams E, Rothkin K, Stevenson D, Boccippio D. Global Lightning Variations Caused by Changes in Thunderstorm Flash Rate and by Changes in the Number of Thunderstorms. *J Appl Met* 2000;39:2223-2230
- [151] Dombai F. Hazai villámlás lokalizációs és radar adatok összehasonlító elemzése. PhD thesis, 2007; ELTE, Budapest.