# Superstatistical Approach to Complex Systems

**G.Cigdem Yalcin**[a] **and Christian Beck**[b]

[a]*İstanbul University, Faculty of Science, Department of Physics, Vezneciler, İstanbul, Turkey*

[b]*School of Mathematical Sciences, Queen Mary, University of London, London E1 4NS, UK*

Complex Systems Summer School 2017 – Santa Fe Institute
June 29, 2017

1 What is superstatistics?

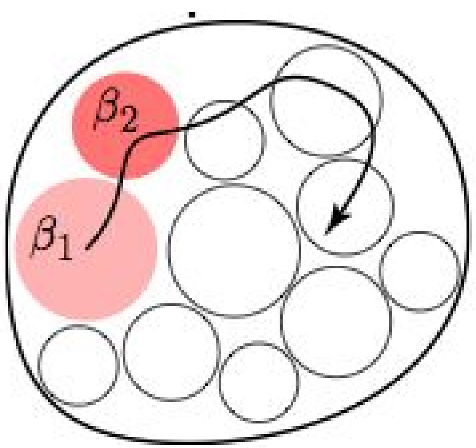
2 Typical universality classes of superstatistics

3Applications

- Environmental superstatistics (rainfall, temperature)
- Cosmic rays

Many complex systems are effectively described by a mixture of different statistics on different time scales.

Simple example:
Consider Brownian particle moving through spatio-temporal inhomogeneous environment with temperature fluctuations on a large scale.




Mixture of two statistics:
- locally ordinary Brownian motion at inverse temperature βi
- superimposed to this is the stochastic process of inverse temperatures βi

(Christian Beck, Phys. Rev. Lett., 2001)

One can also construct dynamical realizations of superstatistics in terms of Langevin equations with parameters that fluctuate on large time scales.

These local Langevin equations decribe the mesoscopic system under consideration.

The simplest example would be locally a linear Langevin equation

$$\dot{v} = -\gamma v + \sigma L(t)$$

with slowly fluctuating parameters g,s. Here, L(t) denotes Gaussian white noise.

This describes the velocity v of a Brownian particle that moves through spatial 'cells' with different local $\beta = \frac{\gamma}{2\sigma^2}$ in each cell (a non-equilibrium situation).

(Christian Beck, Phys. Rev. Lett., 2001)

If some probability distribution f(β) of the inverse temperature β for the various cells is given,

the conditional probability given some fixed β in a given cell is Gaussian

$$p(v|\beta)) \sim e^{-\frac{1}{2}\beta v^2}$$

the joint probability is $p(v, \beta) = f(\beta)p(v|\beta)$

the marginal probability is $p(v) = \int_0^\infty f(\beta)p(v|\beta)d\beta$

Integration over β effectively yields Boltzmann factors that are more general than Gaussian distributions, which depend on the specific properties of f(β).

The principal idea of superstatistics is to generalize this example to much broader systems. For example: β need not be an inverse temperature but can in principle be any intensive parameter.

Most importantly, one can generalize to general probability densities f(β) and general Hamiltonians.

In all cases, one obtains a superposition of two different statistics: that of β and that of ordinary statistical mechanics.
Short name: Superstatistics (C.Beck and E.G.D Cohen, Physica A (2003))

Superstatistics hence describes complex non-equilibrium systems with spatio-temporal fluctuations of an intensive parameter on a large scale. The effective Boltzmann factors B(E) for such systems are given by

$$B(E) = \int_0^\infty f(\beta) e^{-\beta E} d\beta$$

f(β): probability distribution of β

Many results can be proved for general f(β).

Some recent theoretical developments of the superstatistics concept:

- Superstatistics in high energy physics: Differential cross sections well described by q-exponentials and near-q-exponentials, but what really causes these power laws? Curado, Wilk, Biro, Beck, Deppman, .... (2000-2015)
- Study general symmetry group properties of superstatistics (Gell-Mann et al, PNAS 2012)
- Superstatistical path integrals (Jizba & Kleinert 2008-2012)
- Can consider superstatistical random matrix theory (Abul-Magd 2006-2012)
- Can study various theoretical extensions of the superstatistics concept (Chavanis (2005), Vignat, Plastino (2005), Grigolini et al. (2005), Crooks (2006), Naudts (2007), Abe (2007))
- Can develop variational principle for large-energy asymptotics of general superstatistics (H. Touchette and C.Beck, Phys. Rev. E (2005))
- Can prove superstatistical generalization of fluctuation theorems (C.Beck and E.G.D. Cohen, Physica A (2004))
- Can formally define generalized entropies for general superstatistics (Tsallis and Souza, Phys. Rev. E (2003))
- Can apply superstatistical techniques to networks (Abe & Thurner 2005)

…and some more practical applications:

• Universality of market superstatistics (Denys et al. Phys. Rev. E (2016)
• Transitions from one superstatistics to another (Xu et al. 2015)
• Can develop extreme value theory for general superstatistics (Rabassa et al., Entropy (2014))
• Can apply it to rocks/porous media (Correira et al., JMR 2014)
• Can apply it to cosmic ray and high energy scattering statistics (C.B. 2004, 2009, Wilk 2012)
• Superstatistics of labour productivity in manufacturing and nonmanufacturing sectors (Aoyama et al. Economics, 2009)
• Medical applications (Chen et al. (2008), Metzner et al (2015))
• Can apply it to train delay statistics (Briggs et al. 2007)
• Can apply it to hydroclimatic fluctuations (Porporato et al. 2006)
• Can apply superstatistical methods to analyse statistics of 3d hydrodynamic turbulence (C.B., PRL 2001, 2007, A. Reynolds, PRL 2003, C.B., Europhysics Lett. 2003, E. Bodenschatz 2004, C.B., E.G.D. Cohen and H.L. Swinney, Phys. Rev. E 2005, ...)
• Can apply superstatistical methods to finance (Bouchard 2003, Ausloos 2003, Duarte Queiros 2005, Anteneodo 2008, 2012)
• Can apply it to atmospheric turbulence (wind velocity fluctuations at Florence airport, Rizzo & Rapisarda (2004))

# Basic idea of superstatistics

In general, the superstatistics parameter Beta need not be an inverse temperature but can be an effective parameter in a stochastic differential equation, a volatility in finance or just a local variance parameter extracted from some experimental time series.

Consider the following well-know formula:

$$\int_0^\infty \mathrm{d}\beta f(\beta) \mathrm{e}^{-\beta E} = \frac{1}{(1+(q-1)\beta_0 E)^{1/(q-1)}},$$

where

$$f(\beta) = \frac{1}{\Gamma\left(\frac{1}{q-1}\right)} \left\{\frac{1}{(q-1)\beta_0}\right\}^{1/(q-1)} \beta^{(1/(q-1)-1)} \exp\left\{-\frac{\beta}{(q-1)\beta_0}\right\}$$

is the χ2 (or Γ) probability distribution, and $\beta_0$ and q are parameters (q>1).

We see that averaged ordinary Boltzmann factors $\mathrm{e}^{-\beta E}$ with χ2 distributed β yield an effective Boltzmann factor of q-exponential form, given by the right side of equation.

Beck C and Cohen E G D 2003 Superstatistics, Physica A 322 267–75.

## Typical classes of superstatistics

In experiments, one often observes 3 physically relevant universality classes (C.Beck, E.G.D. Cohen, H.L.Swinney, PRE 2005)

***χ2 s*****uperstatistics**, (=Tsallis statistics)

$$f(\beta)=\frac{1}{\Gamma(\frac{n}{2})}\left(\frac{n}{2\beta_0}\right)^{n/2}\beta^{n/2-1}e^{-\frac{n\beta}{2\beta_0}}$$

As it is well-known, Superstatistics based on χ2 distributions leads to q-statistics, whereas other distributions lead to something more complicated.

**Inverse *χ2* superstatistics**,

$$f(\beta)=\frac{\beta_0}{\Gamma(\frac{n}{2})}\left(\frac{n\beta_0}{2}\right)^{n/2}\beta^{-n/2-2}e^{-\frac{n\beta_0}{2\beta}}$$

**Log-normal superstatistics**,

$$f(\beta)=\frac{1}{\sqrt{2\pi}s\beta}\exp\left\{\frac{-(\ln\frac{\beta}{\mu})^2}{2s^2}\right\}$$

More recently we have shown that superstatistical techniques could be also successfully applied to environmental aspects of surface temperature and rainfall amounts distributions.

In this talk, we will discuss that superstatistical distributions $f(\beta)$ of surface temperature and rainfall amounts are very different at different geographic locations on the Earth. For surface temperature, they typically exhibit a double-peak structure for long-term data. For some of our data sets we also find a systematic drift due to global warming. On the other hand, we will discuss the extreme value statistics for extreme daily rainfall, which can potentially lead to flooding, and the waiting time distribution between rainfall events.

**G.C.Y, Pau Rabassa, Christian Beck**, Extreme event statistics of daily rainfall: dynamical systems approach, J.Phys.A: Math. Theor.,49, 15, (2016)

**G.C.Y, Christian Beck**, Environmental Superstatistics, Physica A, 392, 21, 5431 (2013)

# OUTLINES

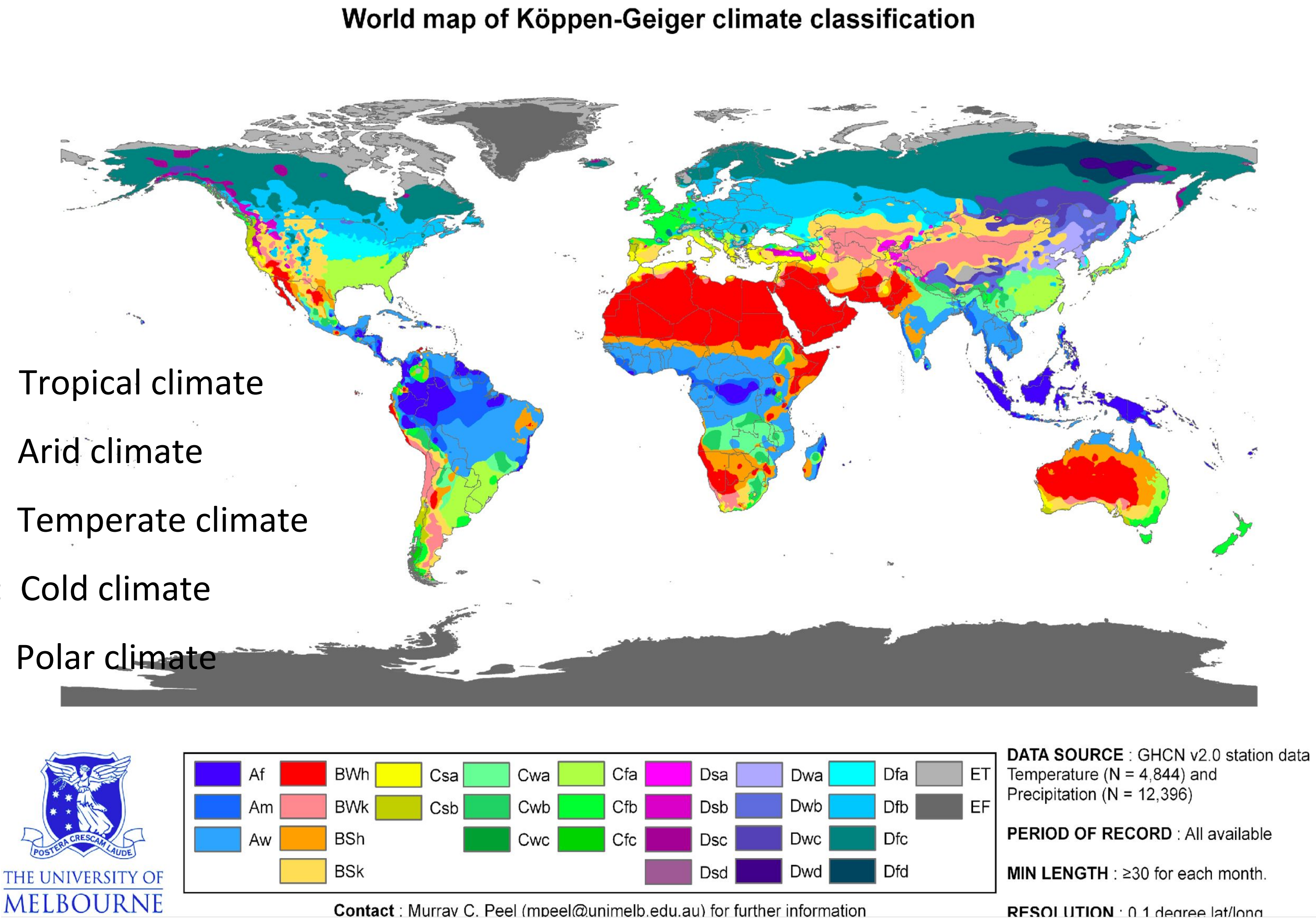
Part- I

- Observed inverse (hourly and daily measured) temperature distributions
  – monthly data – yearly data
- Interpretation of results in terms of Köppen-Geiger climate classification system
- Superstatistics with double-peaked distributions
- Global warming
- Conclusions

Part-II

- Daily rainfall amount distributions at various locations
- Waiting time distributions between rainy days
- Extreme event statistics for exponential and q-exponential distributions
- Conclusions

Our superstatistical distributions are consistent with the Köppen-Geiger climate classification system which is one of the most widely used climate classification systems.




Group A: Tropical climate

Group B: Arid climate

Group C: Temperate climate

Group D: Cold climate

Group E: Polar climate

**We have investigated  time series data for 8 different location in different climatic zones.**

**Aw (Tropical-Savannah):** Darwin (Northern Territory, Australia) (1975-2011)

**BSk (Arid-Steppe-Cold):** Santa Fe (New Mexico, USA) (1998-2011)

**BWh (Arid-Desert-Hot):** Dubai (United Arab Emirates) (1974-2011)

**Cfa (Temperate-Without dry season-Hot summer):** Sydney (New South Wales, Australia) (1910-2011)

**Cfb (Temperate-Without dry season-Warm summer):**
Central England (London-Bristol-Lancashire) (1910-2011)
Vancouver (British Columbia, Canada) (1937-2011)

**Cwa (Temperate-Dry winter-Hot summer):** Hong Kong (PRC) (1997-2011)

**Dfb (Cold-Without dry season-Warm summer):** Ottawa (Ontario, Canada) (1939-2011)

**ET(Polar-Tundra):** Eureka (Nunavut, Canada) (1951-2011)

We analyse in detail inverse temperature distributions at various geographic locations. These environmentally important distributions are different from standard examples of distribution functions discussed so far in the literature, such as the $\chi 2$, inverse $\chi 2$ or lognormal distribution.

A major difference is that the environmentally observed probability densities of inverse temperature typically exhibit a double-peak structure, thus requiring a different type of superstatistics than what has been done so far.

The original superstatistics concept was for sharply peaked distributions $f(\beta)$ with a single maximum (C.Beck, E.G.D. Cohen, 2003). But in an environmental context this concept needs to broadened: Typically observed distributions at various locations on Planet Earth are double-peaked (due to seasonal variations), not single-peaked.

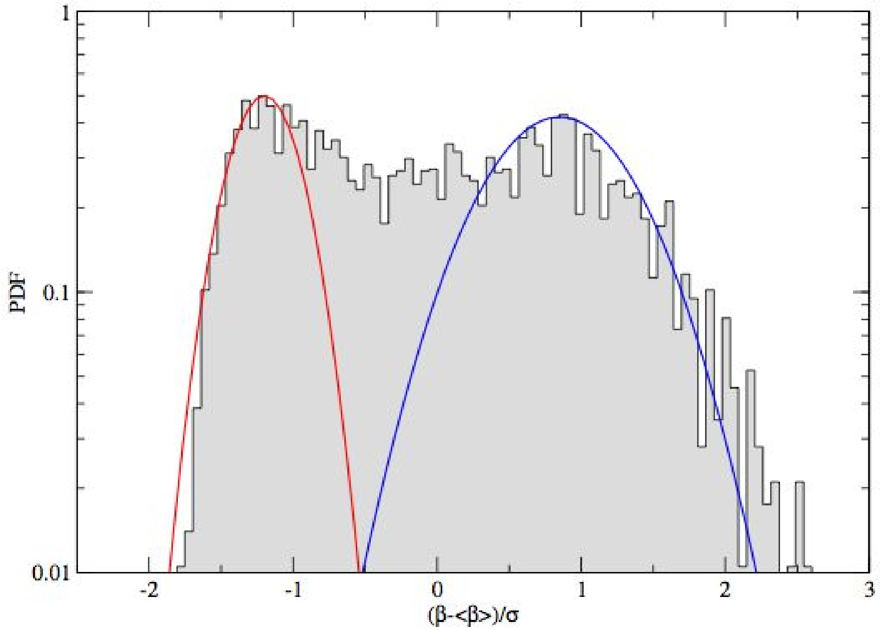
We look at long-term data including seasonal variations, which induce double-peaked distributions, but with specific differences at different geographical locations, depending on local climate.

Apparently environmental superstatistics does not have sharply peaked distributions f(β), but we will see there are broad distributions that often have two maxima. Hence the effective Boltzmann factors

$$B(E) = \int_0^\infty \mathrm{d}\beta f(\beta) \mathrm{e}^{-\beta E}$$

can only be evaluated numerically.

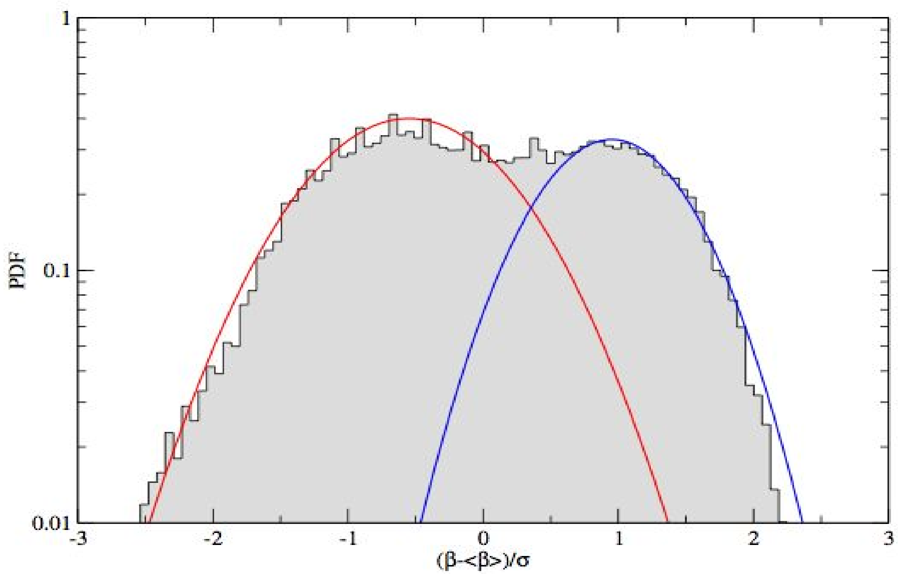
One idea would be effectively separate the two maxima and to do a superposition of two single peaked superstatistics, one for the summer and one for the winter.




Distribution of the daily measured inverse mean temperature in Santa Fe for 1998-2011

Broadly, the left peak corresponds to summer and the right peak to winter.

The entire distribution can be very roughly regarded as a superposition of two Gaussians, with intermediate behaviour between the peaks.




Distribution of the daily measured inverse mean temperature in Sydney for 1910-2011

For example, Sydney seems special since it is the only example where the summer-Gaussian has significantly higher variance than the winter-Gaussian.

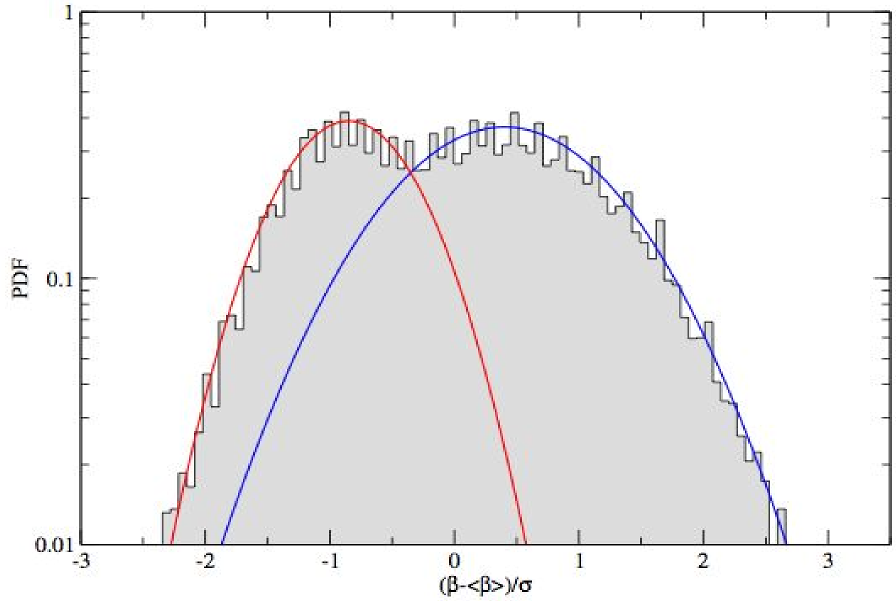



FIG. 26: Distribution of the daily measured inverse mean temperature in Central England (Lancashire, London and Bristol) for 1910-2011

The plots of Central England and Vancouver are very similar.

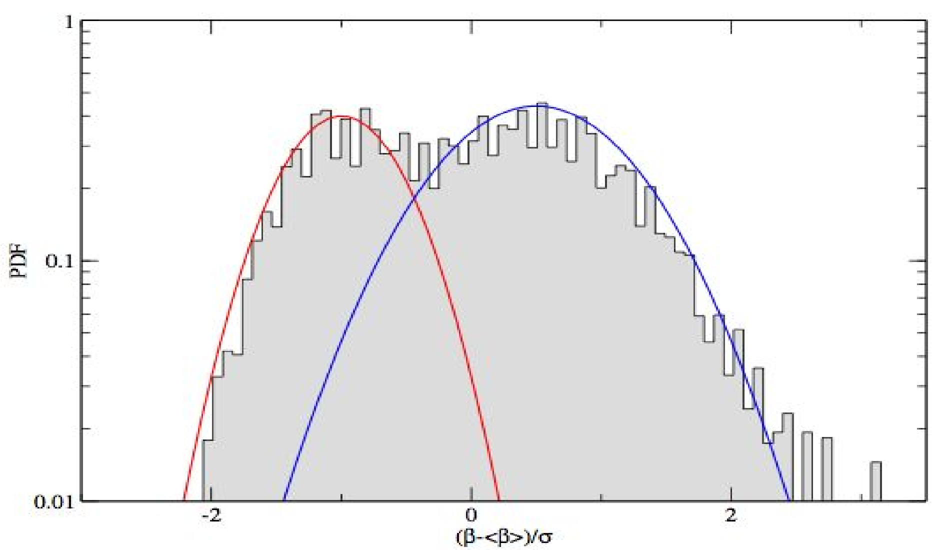
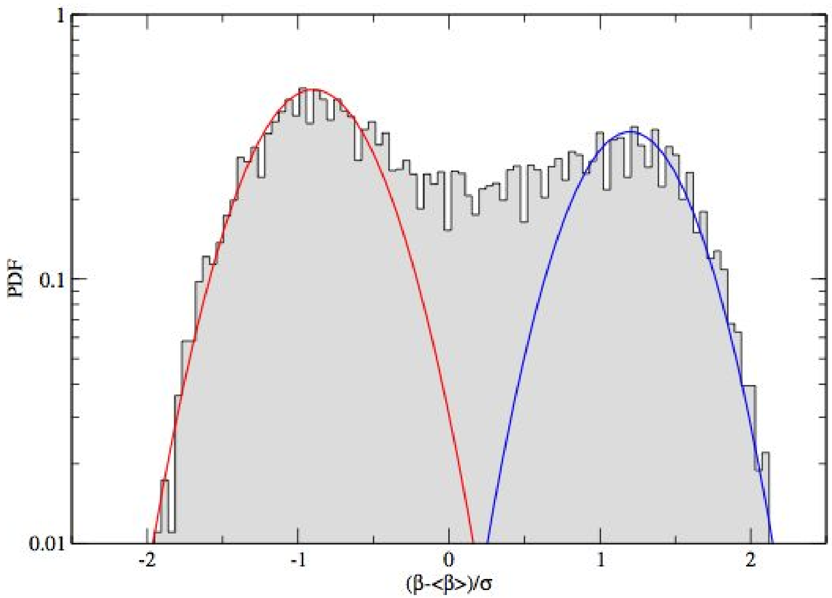



FIG. 27: Distribution of the daily measured inverse mean temperature in Vancouver for 1937-2011

This is to be expected, since both locations fall into the same climate type Cfb




Distribution of the daily measured inverse mean temperature in Dubai for 1974-2011

Dubai seems special as well, since both winter and summer-Gaussian seem to have roughly the same variance. This allows for an alternative fit with the exponential of a double-well potential.

One may try to fit the data by other functional forms than a superposition of two Gaussians.

For example, the histogram of daily measured inverse temperature in Dubai (1974-2011) is well fitted by an exponential $e^{-V(\beta)}$ of a double-well potential $V(x) \sim (C_2x^2+C_3x^{3+}C_4x^4)$.

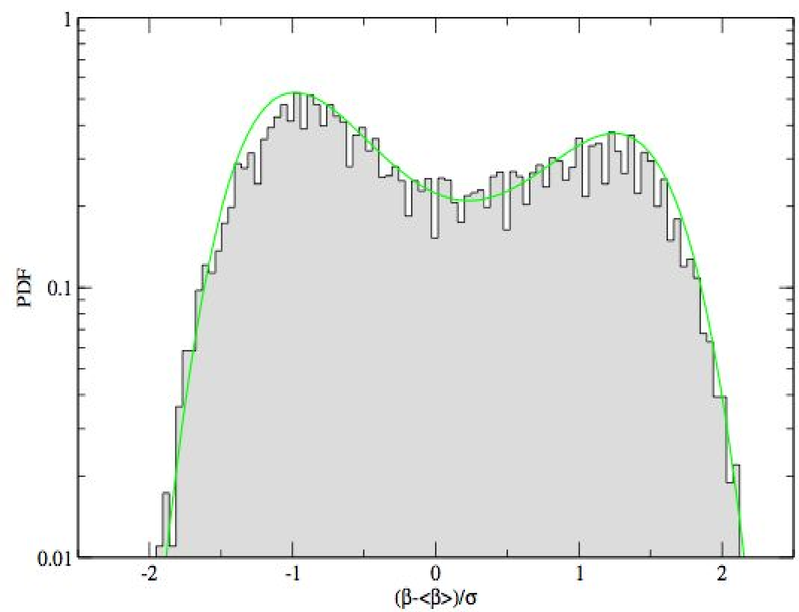
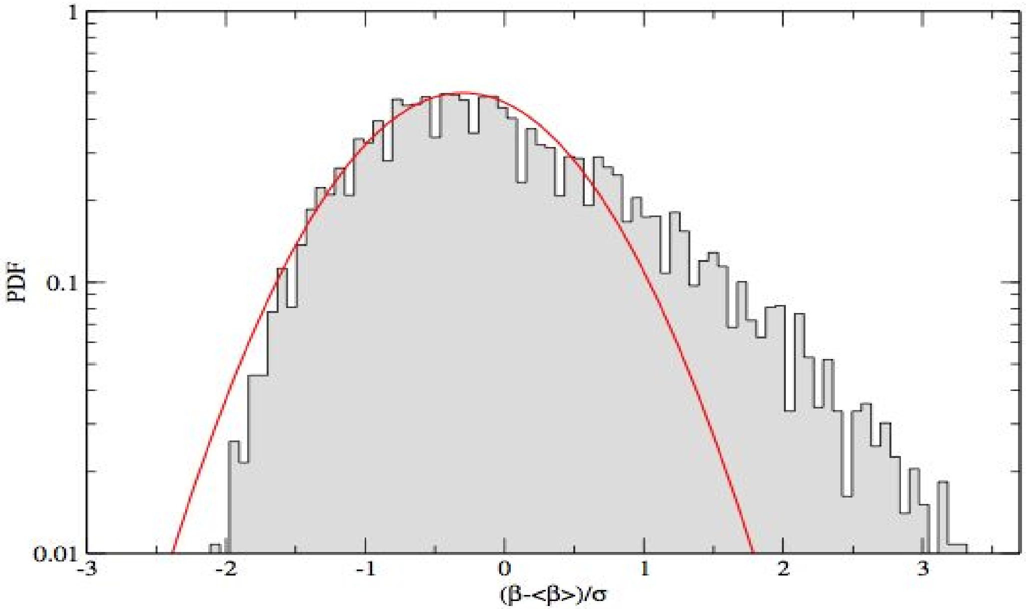



FIG. 31: Fit with the exponential of a double-well potential to the histogram of the daily measured inverse mean temperature in Dubai 1974-2011




Distribution of the daily measured inverse mean temperature in Darwin for 1975-2011

For tropical locations,

such as Darwin ,

the two peaks merge in a single peak,

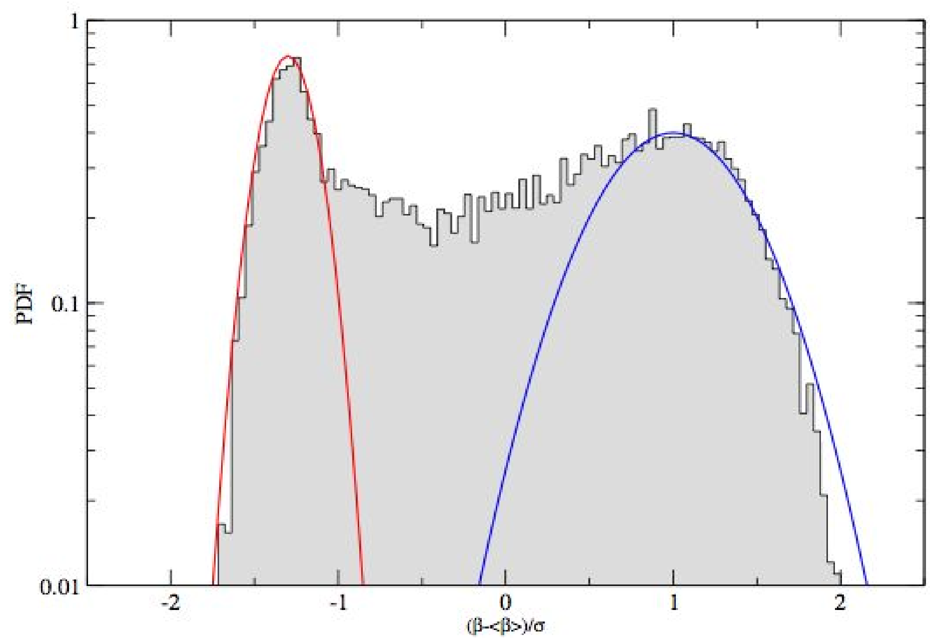
as expected for regions where there

is hardly any difference between

summer and winter temperatures.

A typical observation is now that the relevant distributions have a double-peak structure, at least for non-tropical locations.

The intermediate behaviour between the peak is more pronounced for geographical locations that have a big differences between summer and winter temperatures.




Distribution of the daily measured inverse mean temperature in Eureka for 1951-2011

| *Location and time period* | *Summer(Celcius)* | *Winter(Celcius)* | $\beta_1$ | $\beta_2$ |
|---|---|---|---|---|
| Darwin (1975-2011) | 28.32 | - | 1.8 | - |
| Santa Fe (1998-2011) | 22.06 | 3.31 | 18 | 4 |
| Dubai (1974-2011) | 33.03 | 20.39 | 7 | 8 |
| Sydney (1910-2011) | 20.22 | 13.78 | 2 | 3.5 |
| Central England (1910-2011) | 14.26 | 7.12 | 3.6 | 1.4 |
| Vancouver (1937-2011) | 15.80 | 7.05 | 5 | 2 |
| Hong Kong (1997-2011) | 27.99 | 19.31 | 10 | 2.7 |
| Ottawa (1939-2011) | 18.49 | -1.04 | 12 | 2.1 |
| Eureka (1951-2011) | 4.21 | -36.3 | 43 | 5.5 |

TABLE I: Maxima (in degree Celsius) and variance parameters $\beta_1$, $\beta_2$ of the two Gaussians $\sim e^{-\beta_i(\beta-\bar{\beta})^2}$ used in the fits.

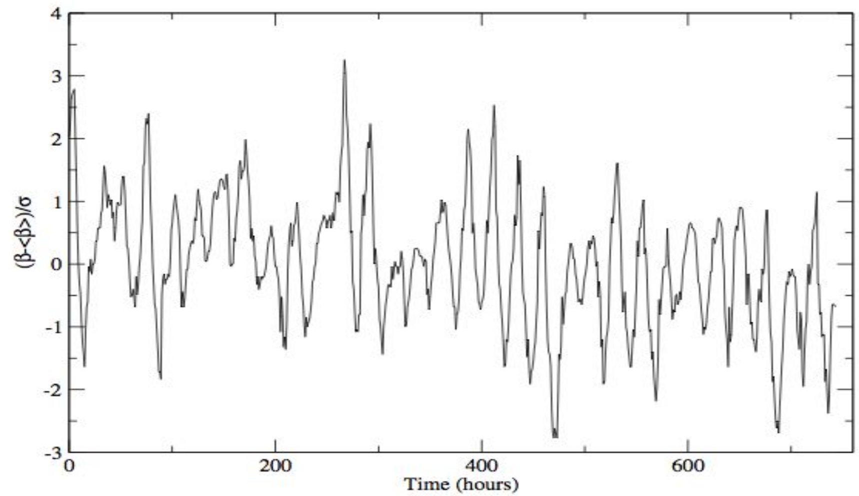
The table lists the two temperatures where the two maxima in histogram occur.

They are consistent with an average temperature observed during a couple of months corresponding to summer and winter, respectively, at different geographical locations.

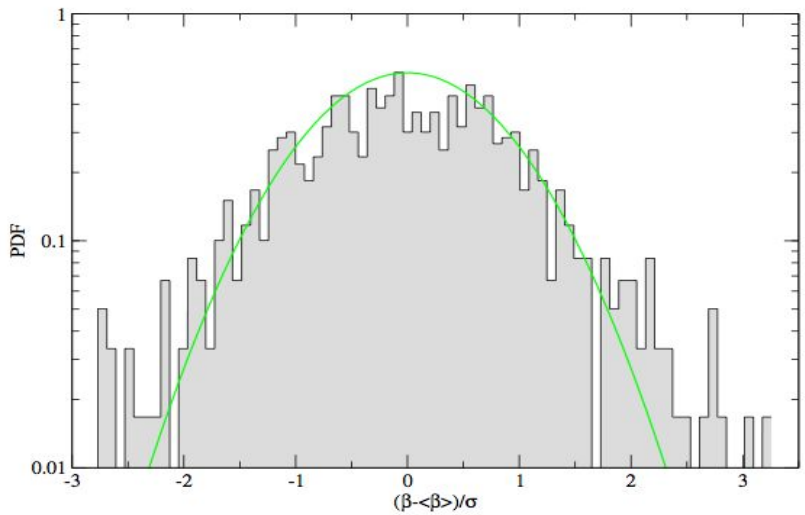
We look at monthly data (essentially eliminating seasonal variations) at various geographical locations, and check how well the data are described by Gaussian distributions.

Short-term data temperature distributions (dominated by daily fluctuations) are very different from long-term data (dominated by seasonal variations)

This figure show as an example a time series of hourly measured surface inverse temperature in Vancouver during May 2011.

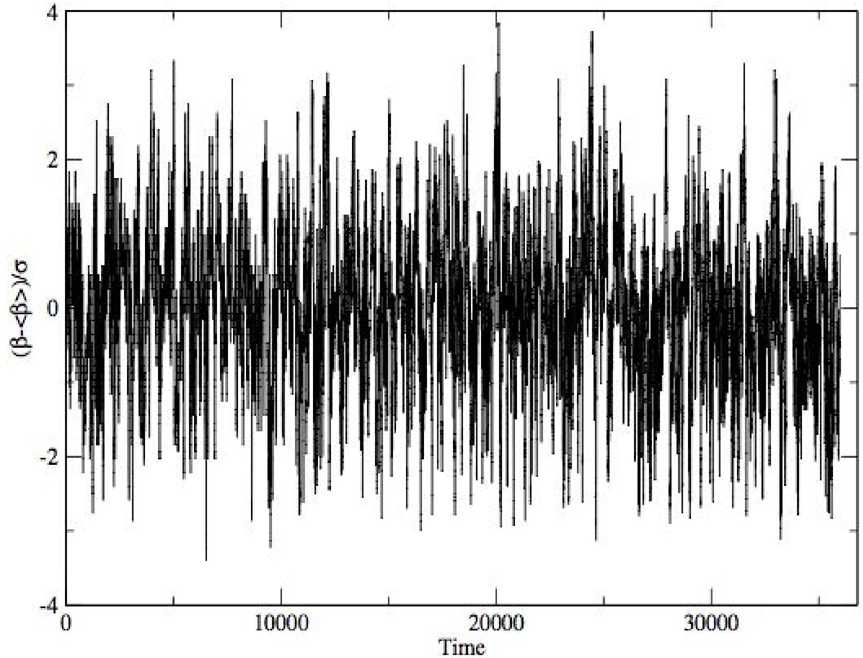



: Time series of hourly measured inverse temperature in Vancouver during May 2011




Gaussian fit to the histogram of the hourly measured inverse temperature in Vancouver during May 2011

For monthly distributions as displayed in this fig., f(β) is indeed a single-humped probability around the mean inverse temperature that month, in good approximation given by a Gaussian distribution.

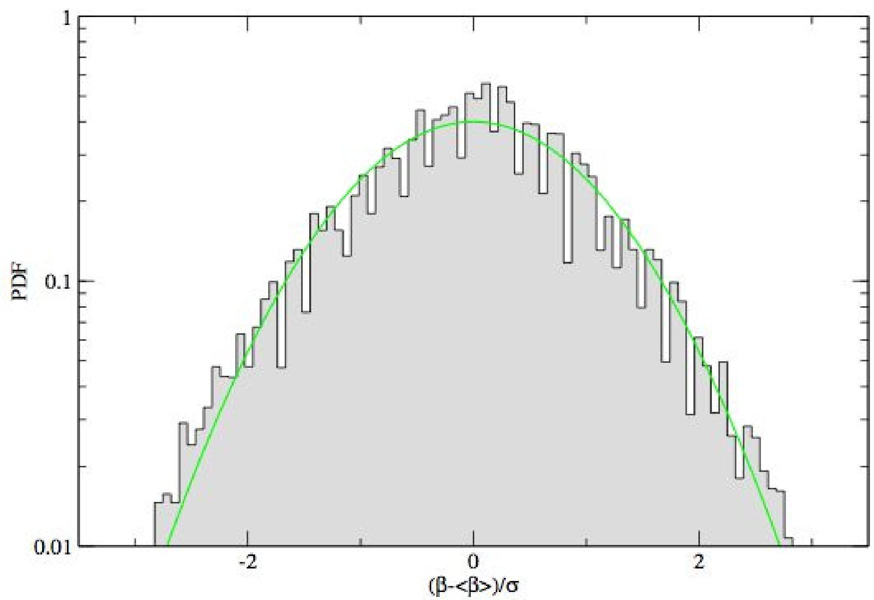



Time series of hourly measured inverse temperature in Ottawa for November, sampled from 1962 to 2011 for 50 years

Better statistics is obtained if we sample the data of a given month over many years.

We sampled hourly measured inverse temperature time series, restricted to the month of November, over the period 1966-2011 for Ottawa.

This figure shows this time series (restricted to November months and joined together).




Gaussian fit to the histogram of the hourly measured inverse temperature in Ottawa for 50 November months

The corresponding histogram is well-fitted by a Gaussian.

In some of our data one sees a systematic trend which can be associated with global warming. As it is well-known, global warming is the rise in the average temperature of the Earth's atmosphere and of the oceans since the late 19th century and its projected continuation.

Quantitatively, the Earth's average surface temperature rose by $0.74 \pm 0.18\ ^\circ C$ over the period 1906-2005. The rate of warming over the last half of that period was almost double that of the period as a whole ($0.13 \pm 0.03\ ^\circ C$ per decade, versus $0.07 \pm 0.02\ ^\circ C$ per decade).

Arctic regions are especially vulnerable to the effects of global warming, as has become apparent in the melting sea ice in recent years. Climate models predict much greater warming in the Arctic than the global average.

If sampled over many decades, the precision of some of our data is good enough to reveal the effects of global warming. As an example we show that the average of daily measured mean temperature of every single year in Eureka, Nunavut, Canada from 1951-2011. This is an arctic region and our plot shows that the average temperature grows by 0.92 ◦C per decade. So in this polar region the warming occurs at a much higher rate than global warming rate that is averaged over the entire earth.

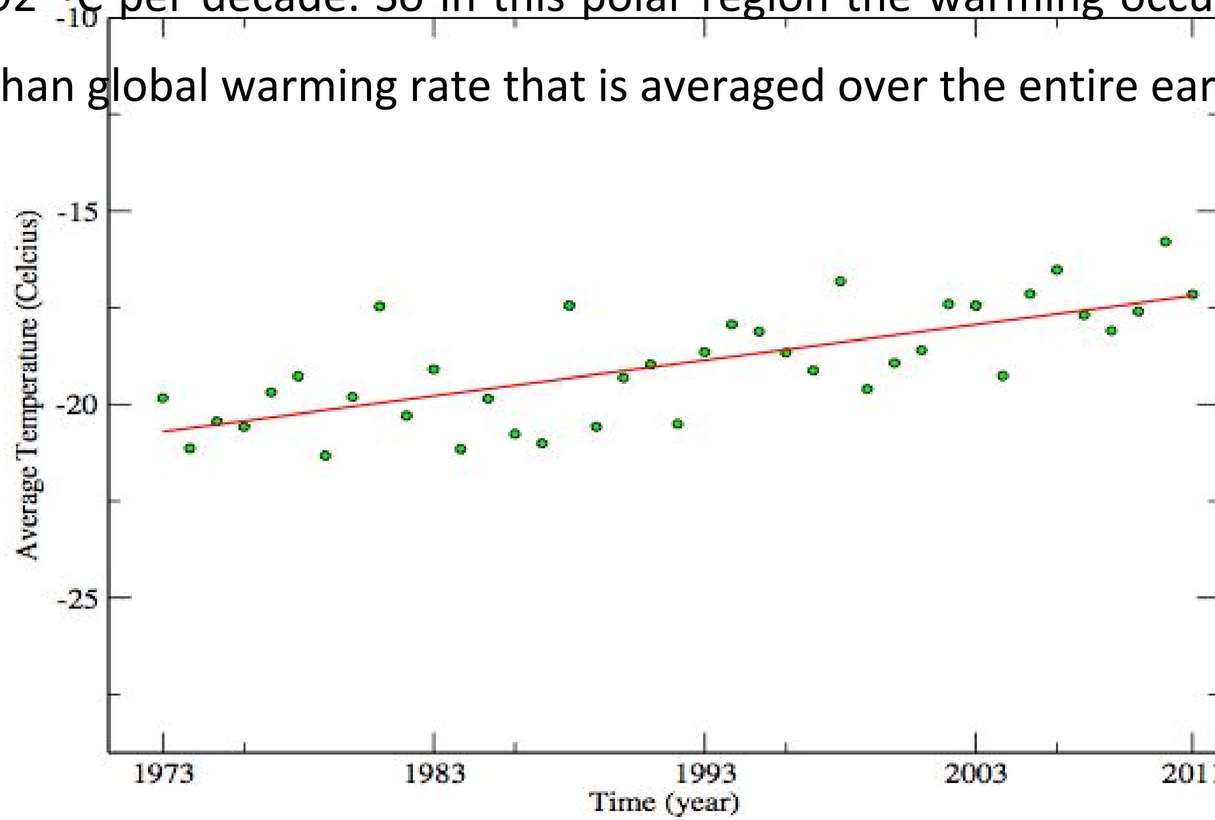



FIG. 33: Average of daily measured mean temperature of every single year in Eureka 1973-2011

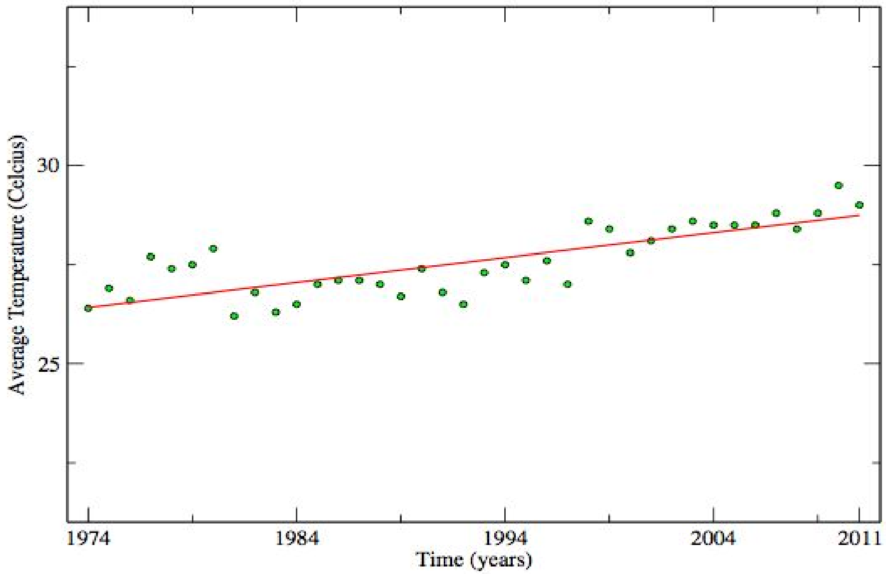



FIG. 34: Average of daily measured mean temperature of every single year in Dubai 1974-2011

For some of the other locations we also observe a systematic increase in average temperature . For Dubai the rate is also rather high, 0.63 ◦C per decade.

For Sydney , the average global warming rate 0.13 ◦C per decade.

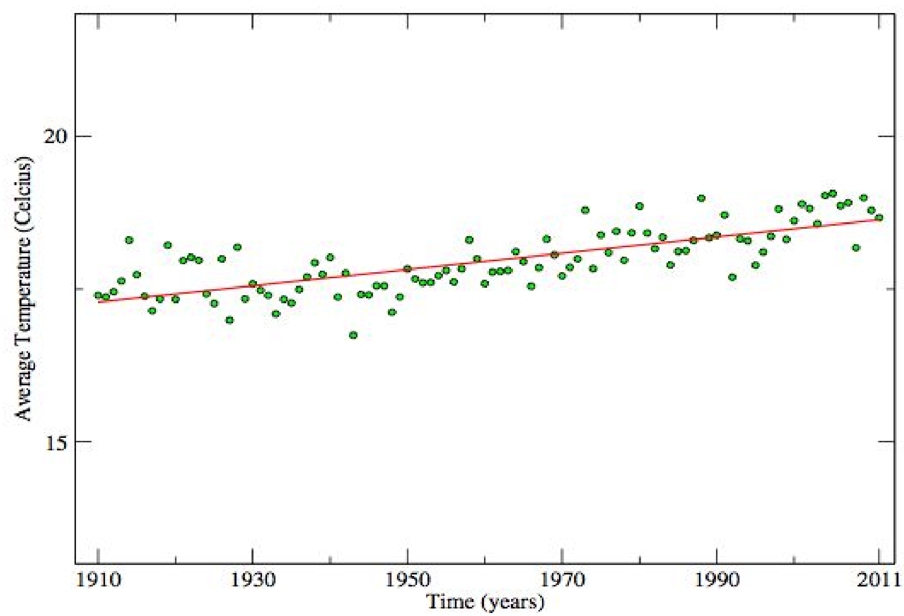
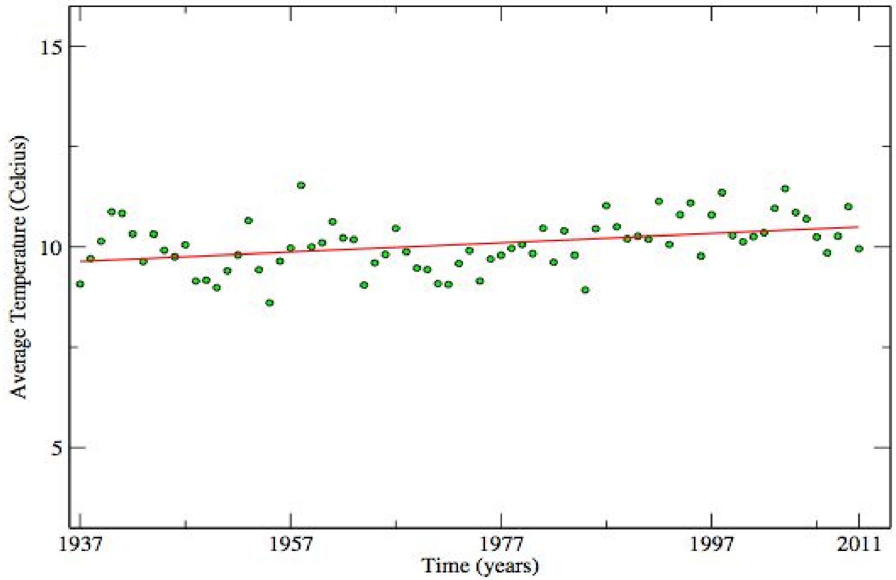



FIG. 35: Average of daily measured mean temperature of every single year in Sydney 1910-2011




For Vancouver, the average global warming rate 0.11 ◦C per decade.

FIG. 36: Average of daily measured mean temperature of every single year in Vancouver 1937-2011

For Ottawa the average global warming rate 0.18 ◦C per decade.

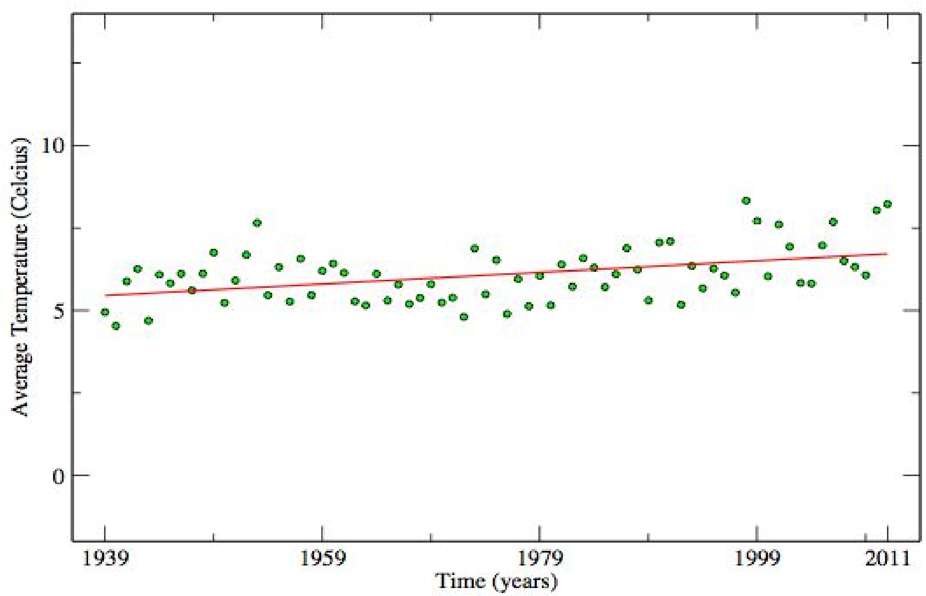



FIG. 37: Average of daily measured mean temperature of every single year in Ottawa 1939-2011

# CONCLUSIONS

◆ Our superstatistical distributions are consistent with the Köppen-Geiger climate classification system which is one of the most widely used climate classification systems.

◆ The details of the distribution of course heavily depend on the climate zone of the location.

◆ These environmentally important distributions are different from standard examples of distribution functions discussed so far in the literature, such as the $\chi 2$, inverse $\chi 2$ or lognormal distribution.

Part-II

- Daily rainfall amount distributions at various locations
- Waiting time distributions between rainy days
- Extreme event statistics for exponential and q-exponential distributions
- Conclusions

We present histograms of rainfall statistics, extracted from experimentally measured time series of rainfall at various locations on the Earth.

We performed a systematic investigation of time series of rainfall data for eight different example locations on Earth.

| | |
|---|---|
| Central England (1931-2014) | Vancouver (1937-2014) |
| Darwin (1975-2014) | Sydney (1910-2014) |
| Dubai (1968-2014) | Ottawa (1939-2014) |
| Eureka (1951-2014) | Hongkong (1997-2014) |

When making a histogram of  the amount of daily rainfall observed, a suprising feature arises. All distributions are power law rather than exponential.

They are well fitted by so-called q-exponentials;

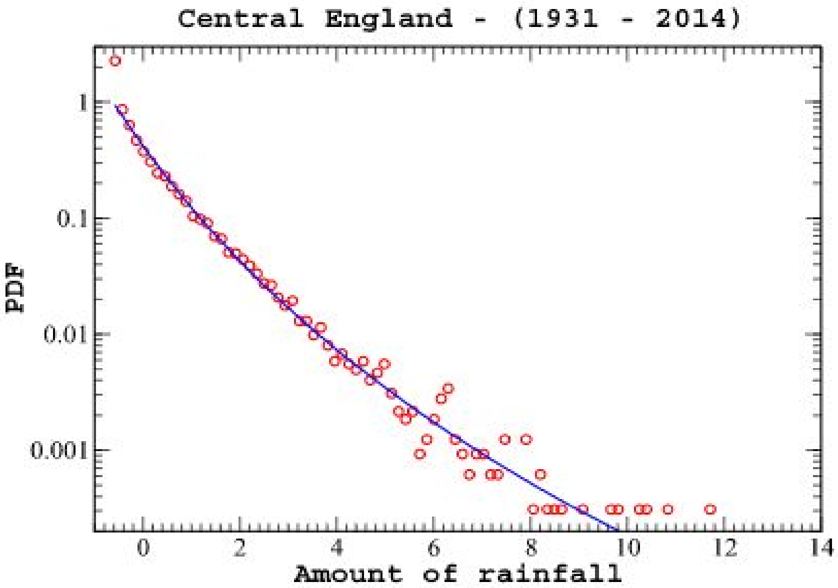
$$e_q(x) = (1 + (1 - q)x)^{1/(1-q)}$$

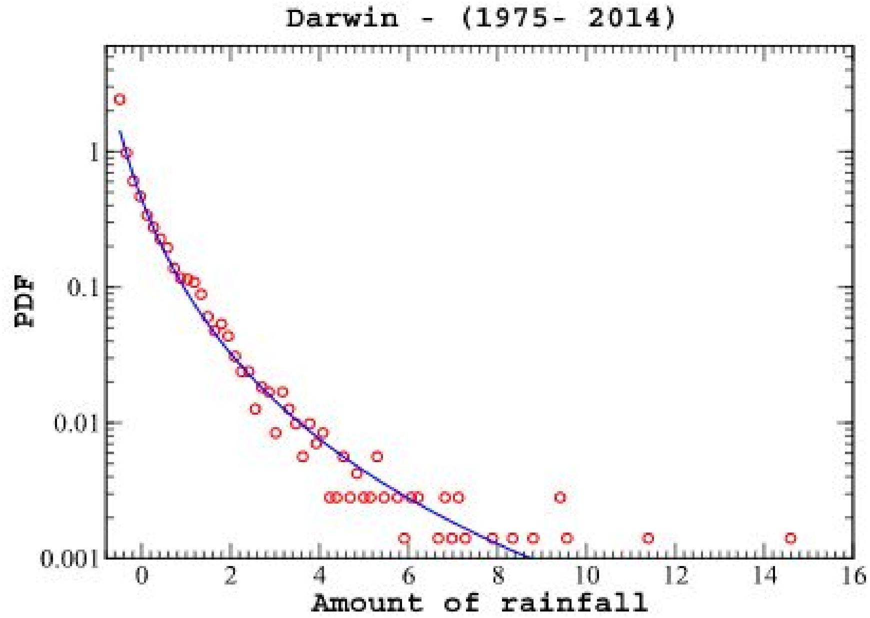
Let’s remember:

Consider the following well-know formula:

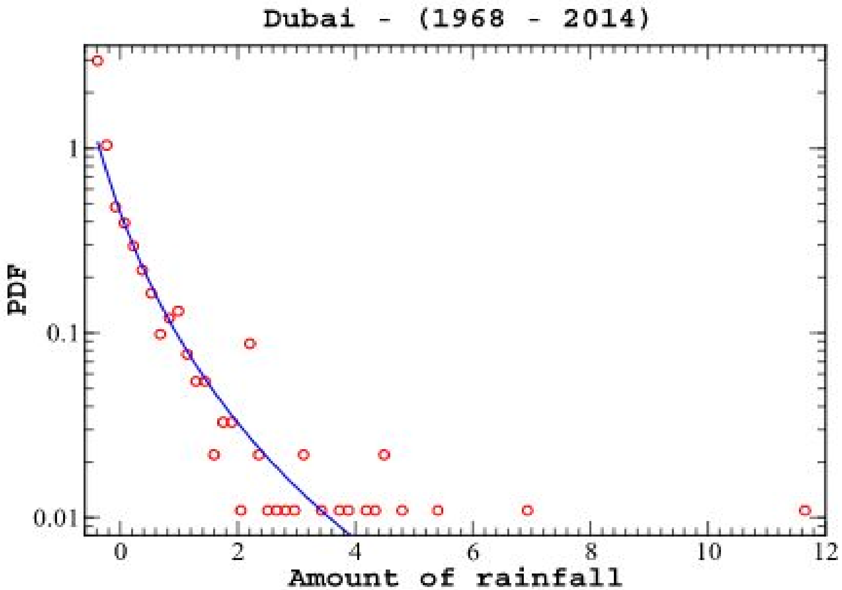
$$\int_0^\infty \mathrm{d}\beta f(\beta) \mathrm{e}^{-\beta E} = \frac{1}{(1 + (q-1)\beta_0 E)^{1/(q-1)}},$$

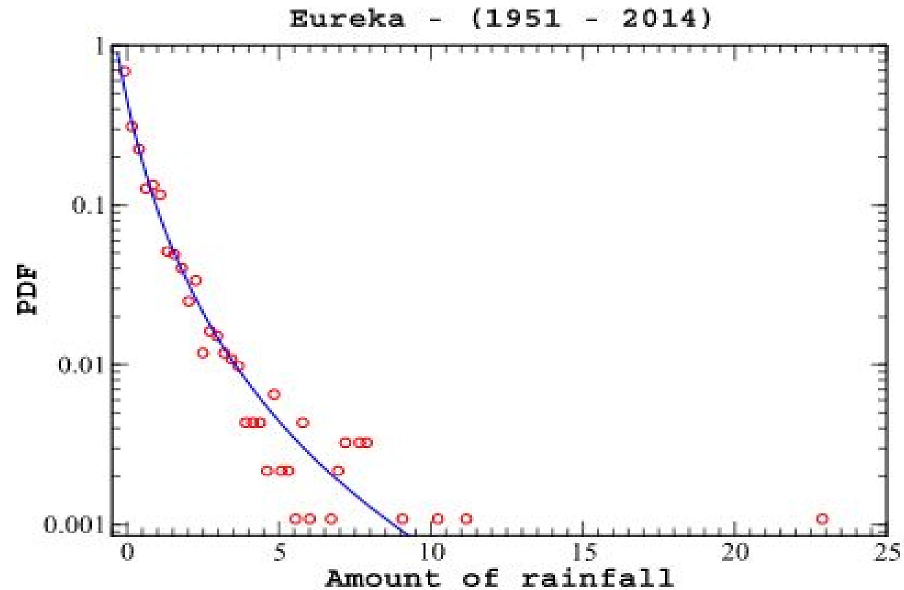
where $$f(\beta) = \frac{1}{\Gamma\left(\frac{1}{q-1}\right)} \left\{\frac{1}{(q-1)\beta_0}\right\}^{1/(q-1)} \beta^{(1/(q-1)-1)} \exp\left\{-\frac{\beta}{(q-1)\beta_0}\right\}$$

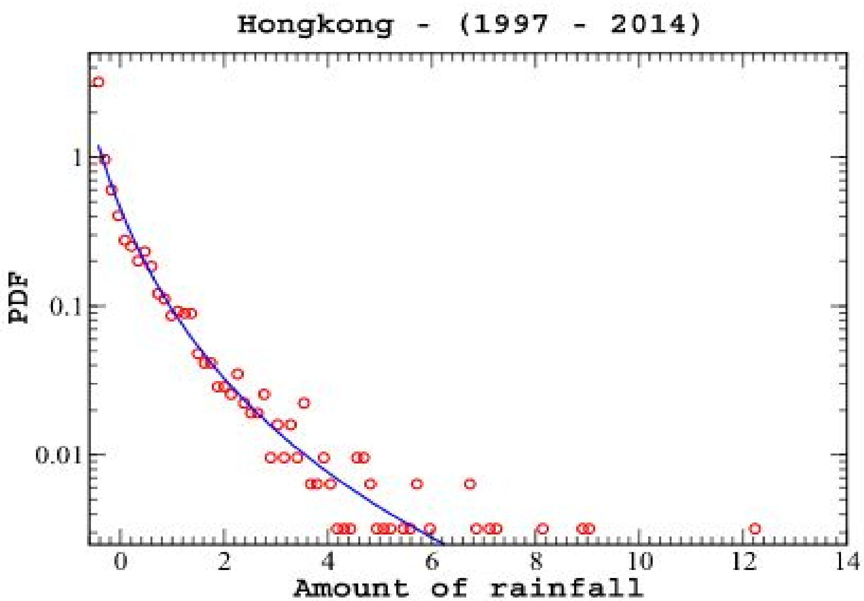


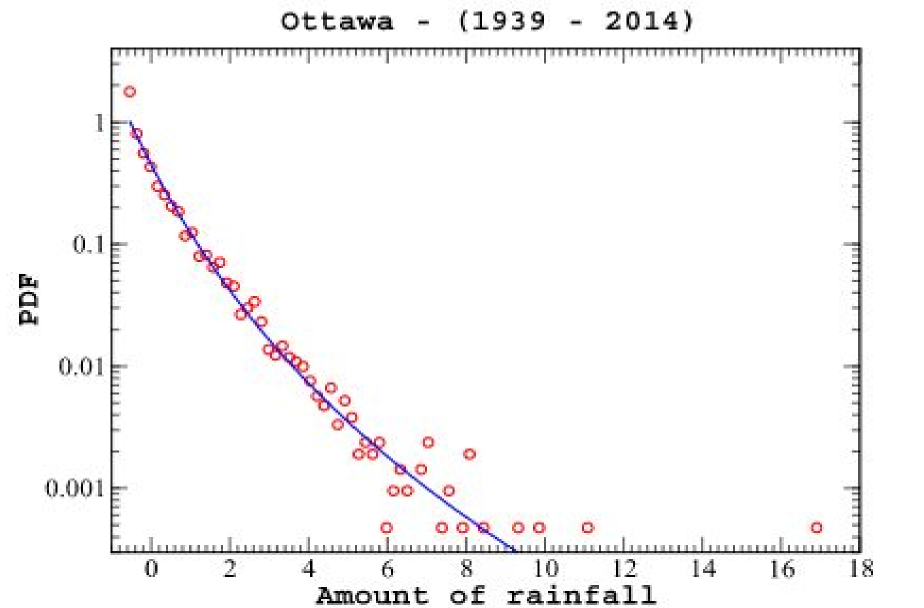




The actual value of q for the observed rainfall statistics reflects characteristics effective properties in climate and temporal precipitation pattern.

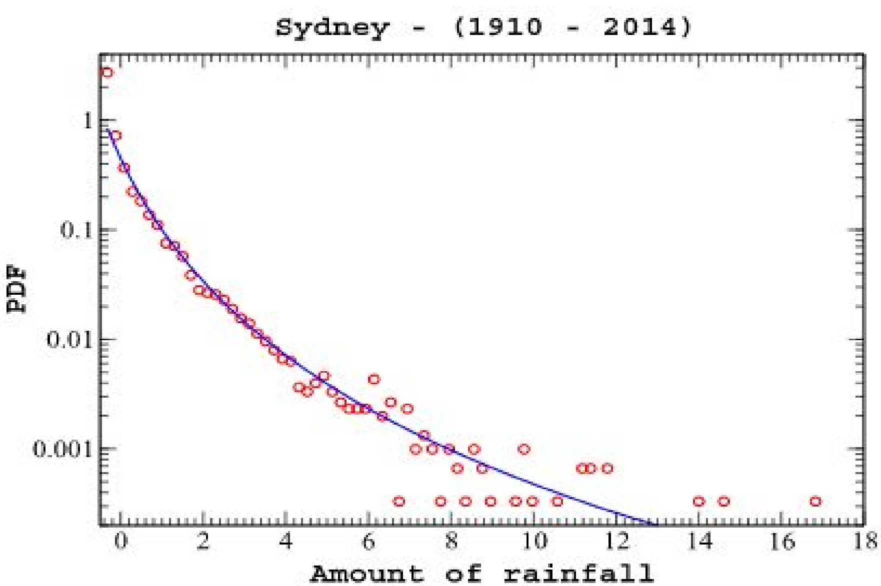


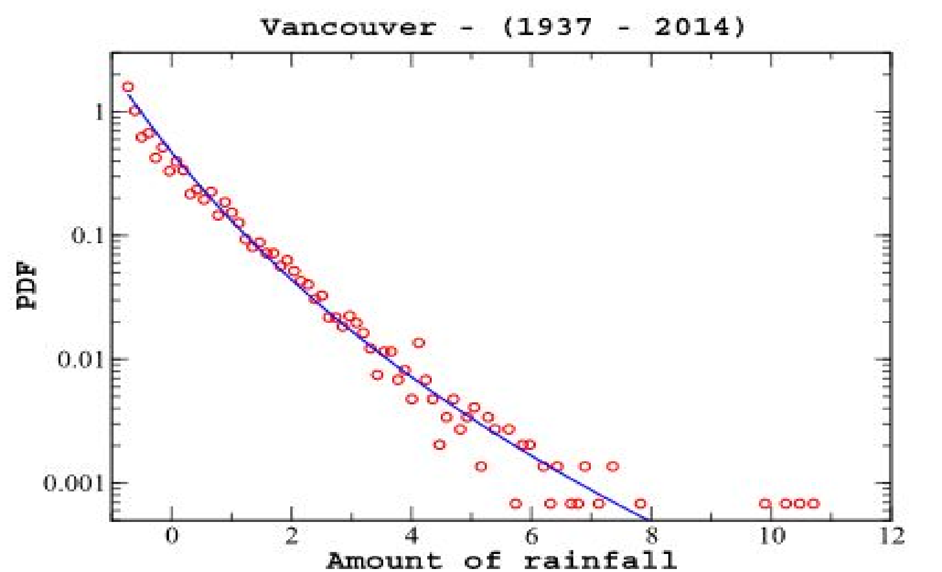





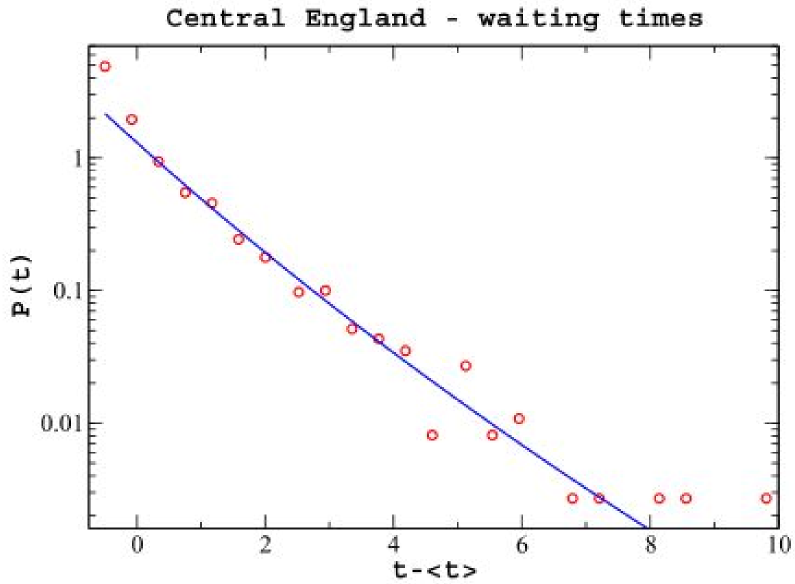

Whereas the data of most locations are well fitted by q≈ 1.3,
Central England and Vancouver somewhat lower values of q closer to 1.13.

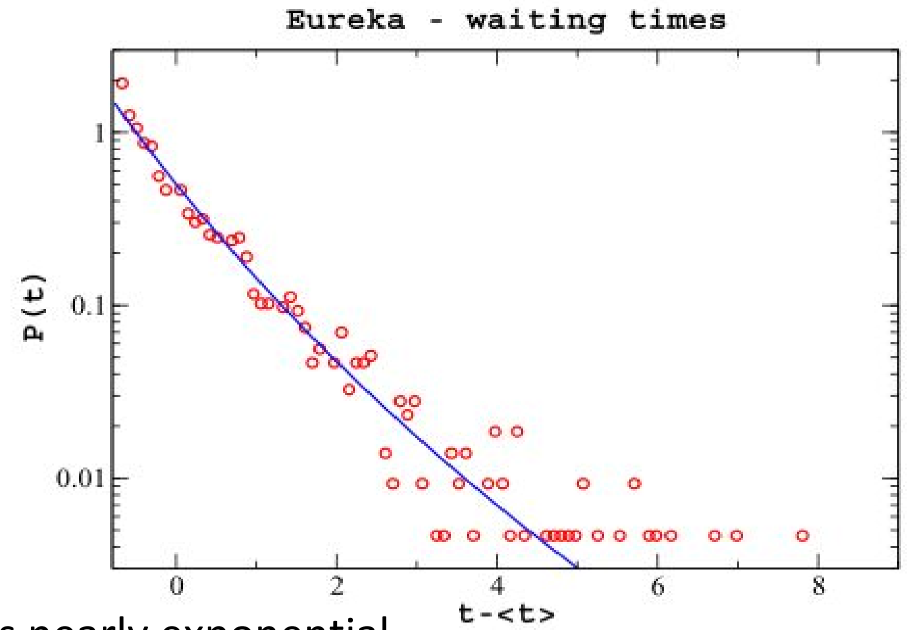


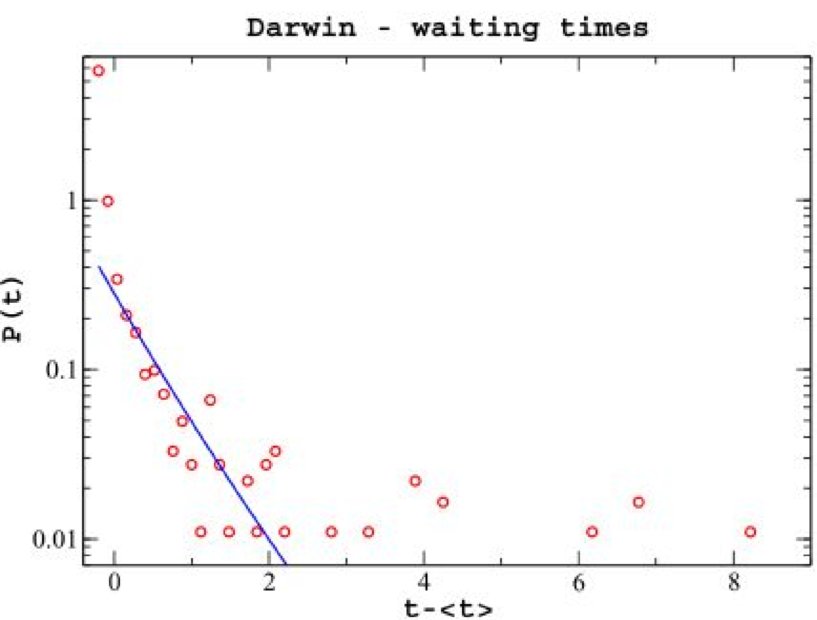
- Waiting time distributions between rainy days

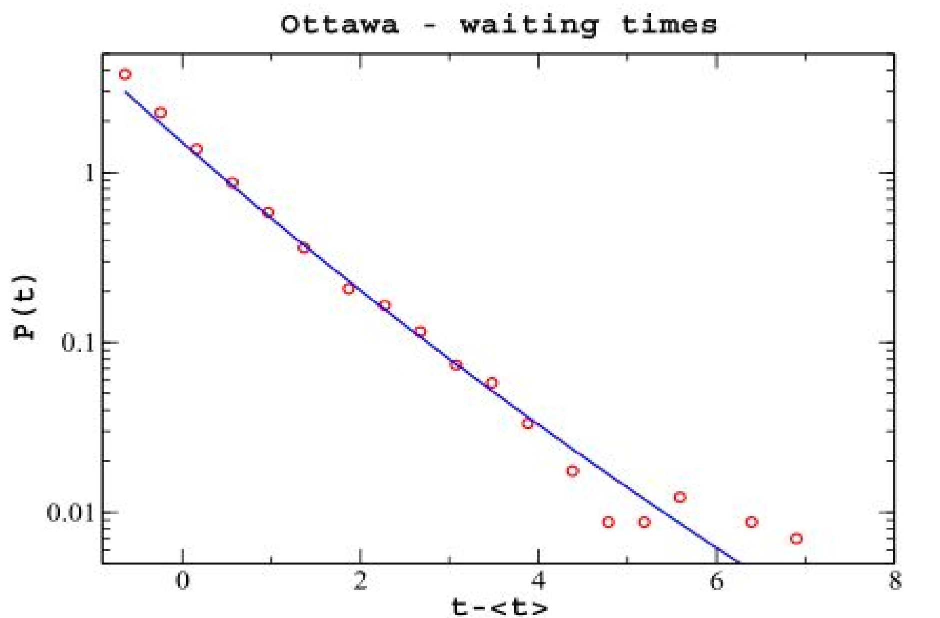
Another interesting observable that we extracted from the data is the waiting time distribution between rainy episodes. The waiting time is then the number of days one has to wait until it rains again. This is a random variable with a given distribution which we can extract from the data.

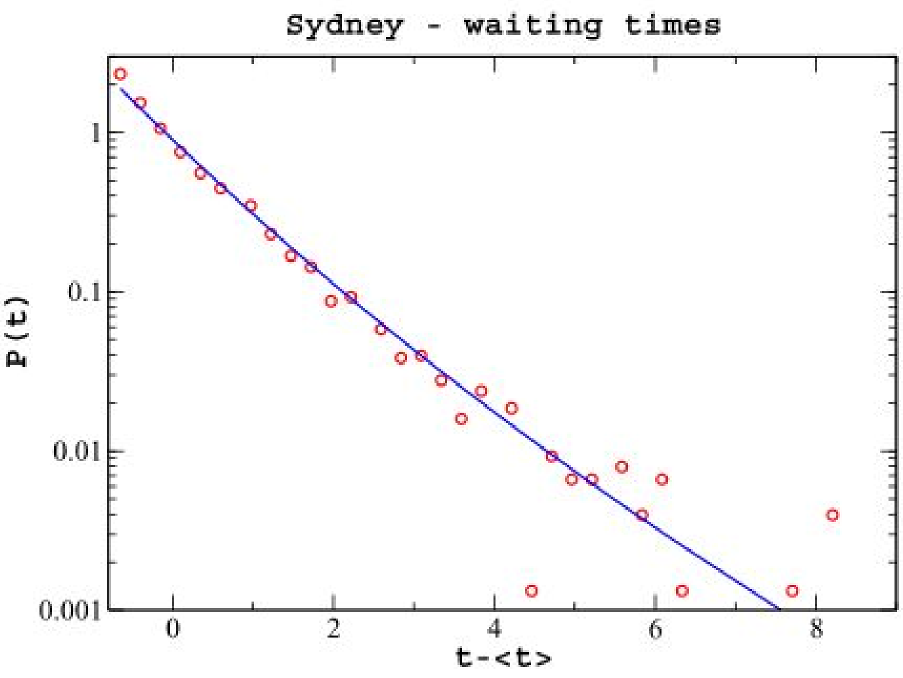


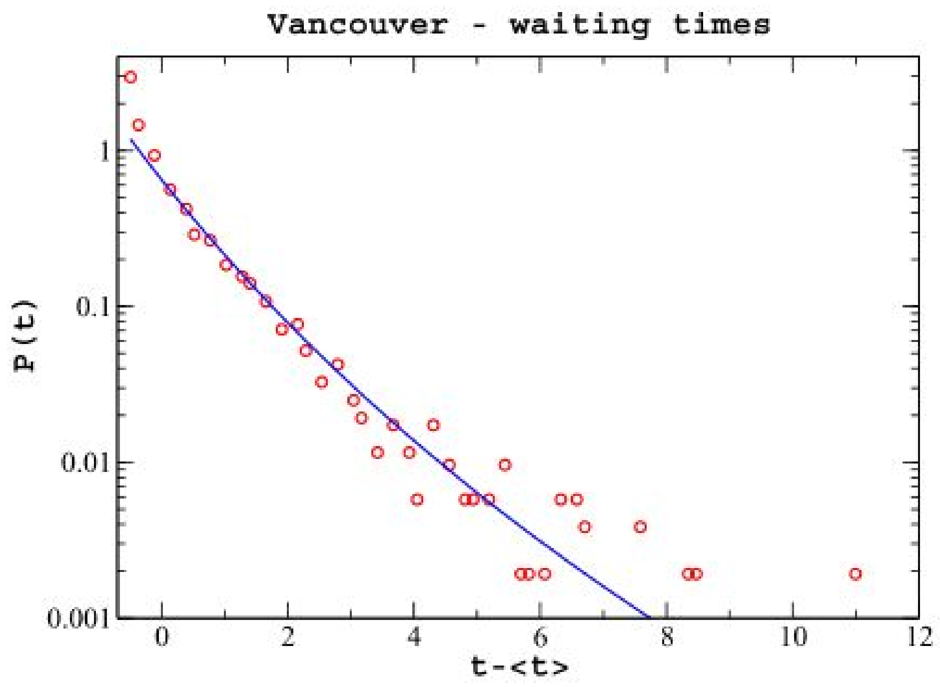




What one observes here is that the distribution is nearly exponential.
That means the Poisson process of nearly independent point events of rainy days is a reasonably good model.






At closer inspection, however, one sees that again a slightly deformed q-exponential, this time with q≈1.05, is a better fit of the waiting time distribution.

# Conclusions

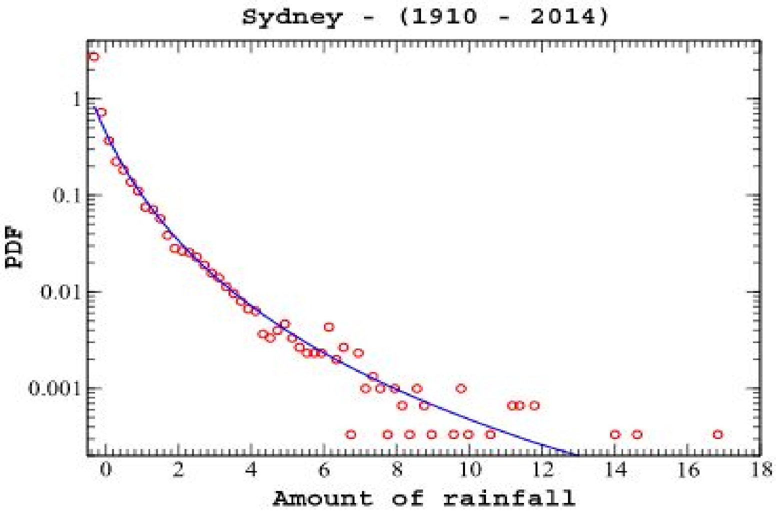
The probability densities of daily rainfall amounts at a variety of locations on Earth are not Gaussian or exponentially distributed, but follow an asymptotic power law, the q-exponential distribution.

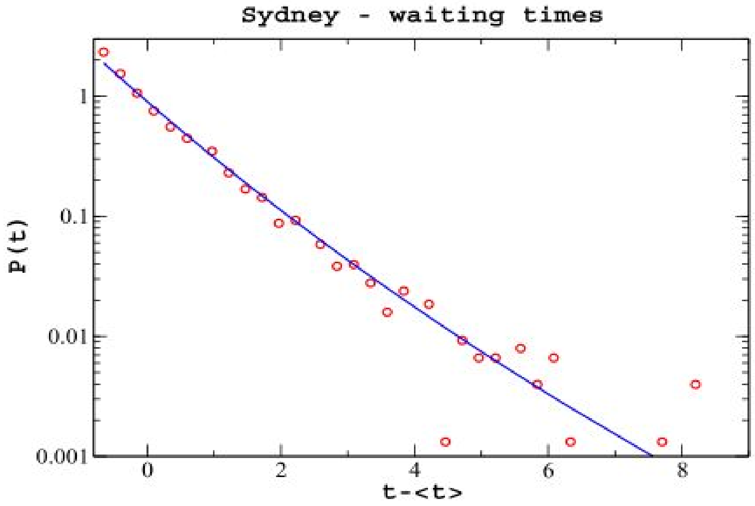
The corresponding entropic exponent q is close to q≈ 1.3.
The waiting time distribution between rainy episodes is observed to be close to an exponential is a better fit, this time with q close to 1.05.

For extreme event statistics;

power laws (q-exponential) exponential laws (ordinary exponential)

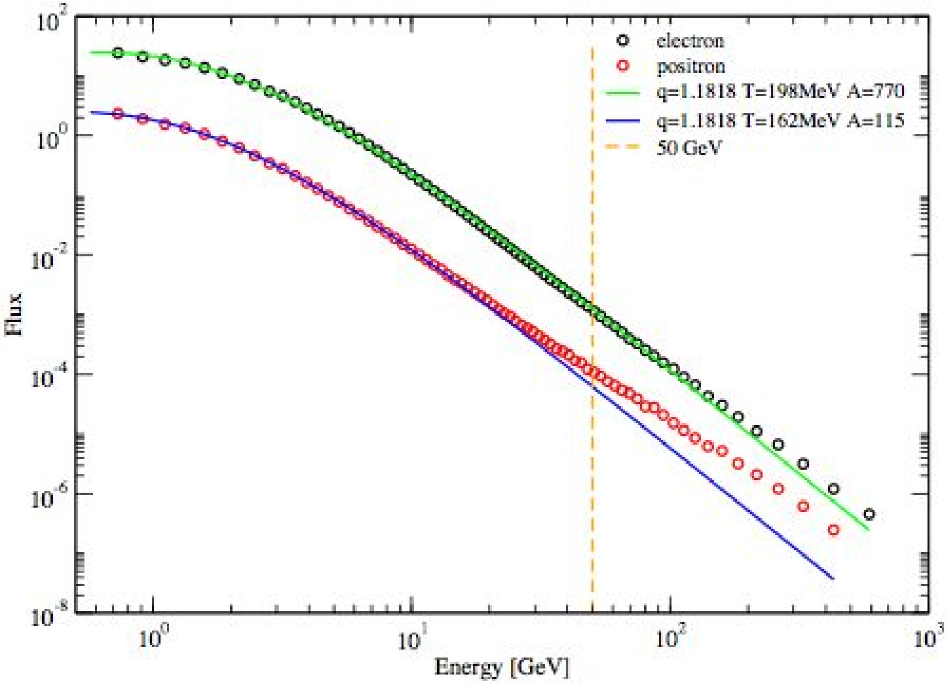
q≈1.3 → Gumbel distributions q≈1.05 → Frechet distributions



**Generalized statistical mechanics of cosmic rays: Application to positron-electron spectral indices (May, 2017)** **Arxiv : 1705.01334**

We apply generalized statistical mechanics developed for complex systems to theoretically predict energy spectra of particle and anti-particle degrees of freedom in cosmic ray fluxes based on a q-generalized Hagedorn theory for transverse momentum spectra and hard QCD scattering processes.

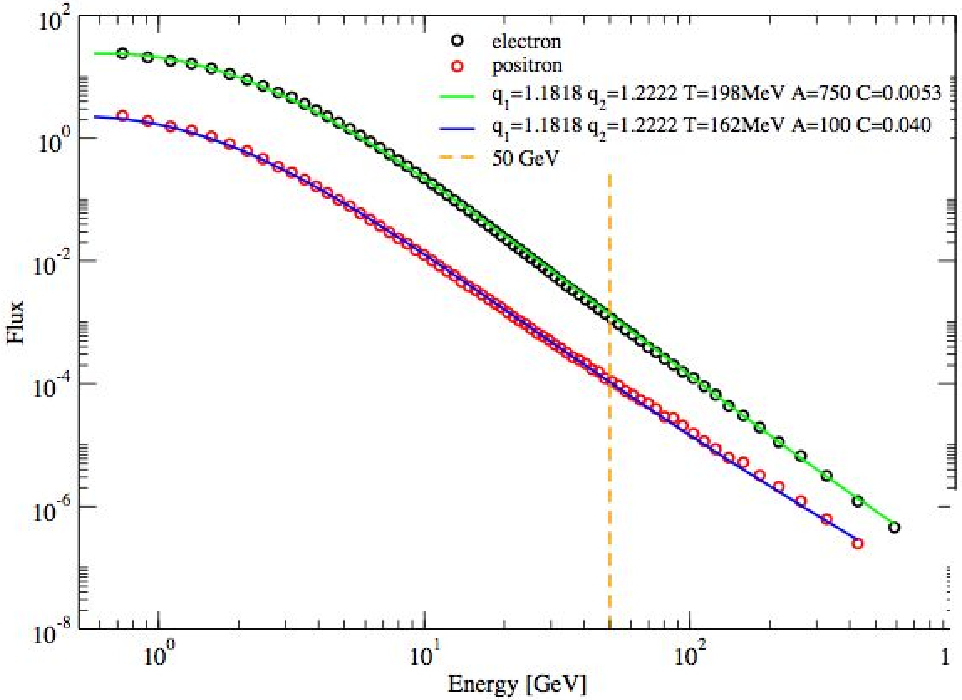



We start from q-generalized canonical distributions of the form

$$p(E) \sim \frac{E^2}{(1+(q-1)\beta_0 E)^{\frac{1}{q-1}}}.$$

Here q is the entropic index and $\beta_0$ is an inverse temperature parameter. E is the energy of the particle.

Flux Φ(E) of e+ and e− primary cosmic ray particles of energy E as measured by AMS-2 and theoretical prediction of the q-generalized Hagedorn theory (solid lines)




The measured AMS-02 data are very well fitted by linear combination of escort and non-escort distributions (solid lines)

We found that this crossover is very well described by a linear combination of generalized canonical distributions where the entropic index takes on two values, namely the QCD value 13/11 = 1.1818 and the escort value 11/9 = 1.2222, evaluated at temperature T for positrons and Tˆ for electrons.

This figure shows that in the entire energy range the measured cosmic ray flux is very well fitted by the linear combination

$$P_{\pm}(E) = A_{\pm}\left(\frac{E^2}{(1+(q-1)\beta_0 E)^{\frac{1}{q-1}}} + C_{\pm}\frac{E^2}{(1+(\hat{q}-1)\beta_0 E)^{\frac{1}{\hat{q}-1}}}\right)$$

Thank you…