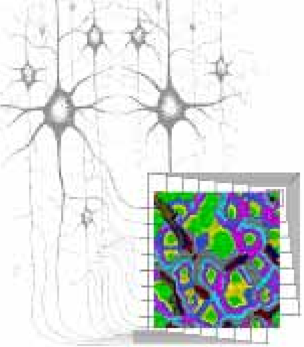


Climb or Glide?

Data Mining & Knowledge Discovery in Flight Records

Question:
What is the probability to find a thermal of certain strength on a typical cross country day?

Prof. Dr. Alfred Ultsch

Arbeitsgruppe Datenbionik

Universität Marburg

The Data

1635 Flight Records (IGC files)

Source: Online Contest (OLC) of BY, HE and TH and Coburg Competition

Recording period :

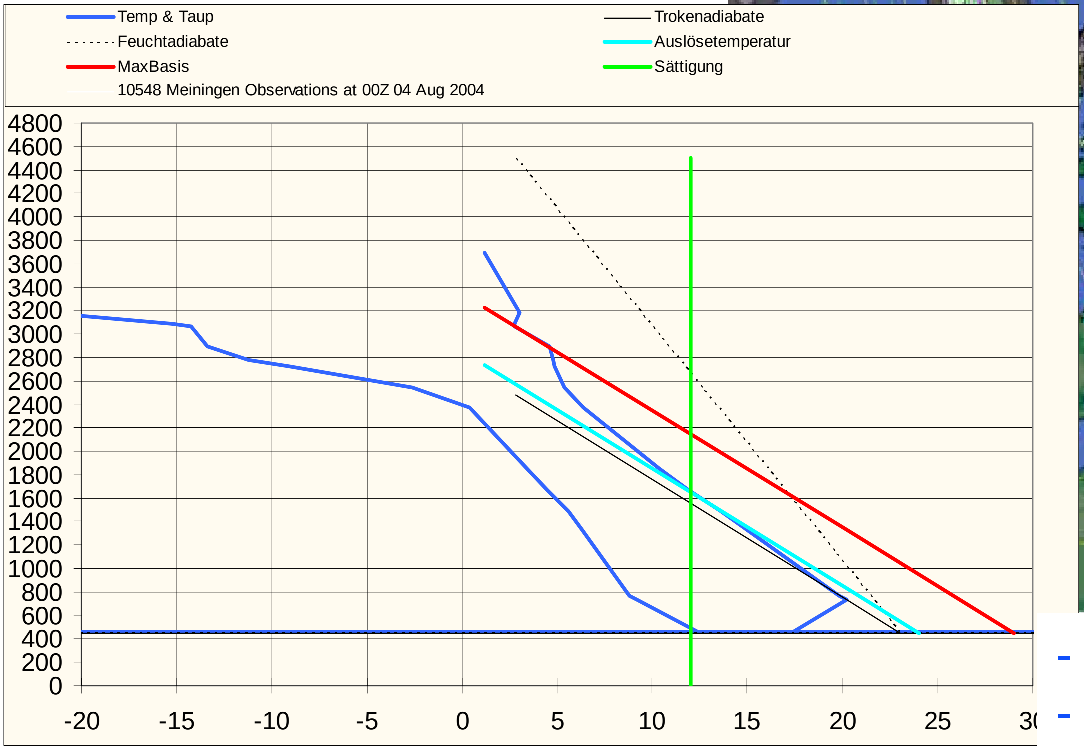
July 30th and August 8 2004

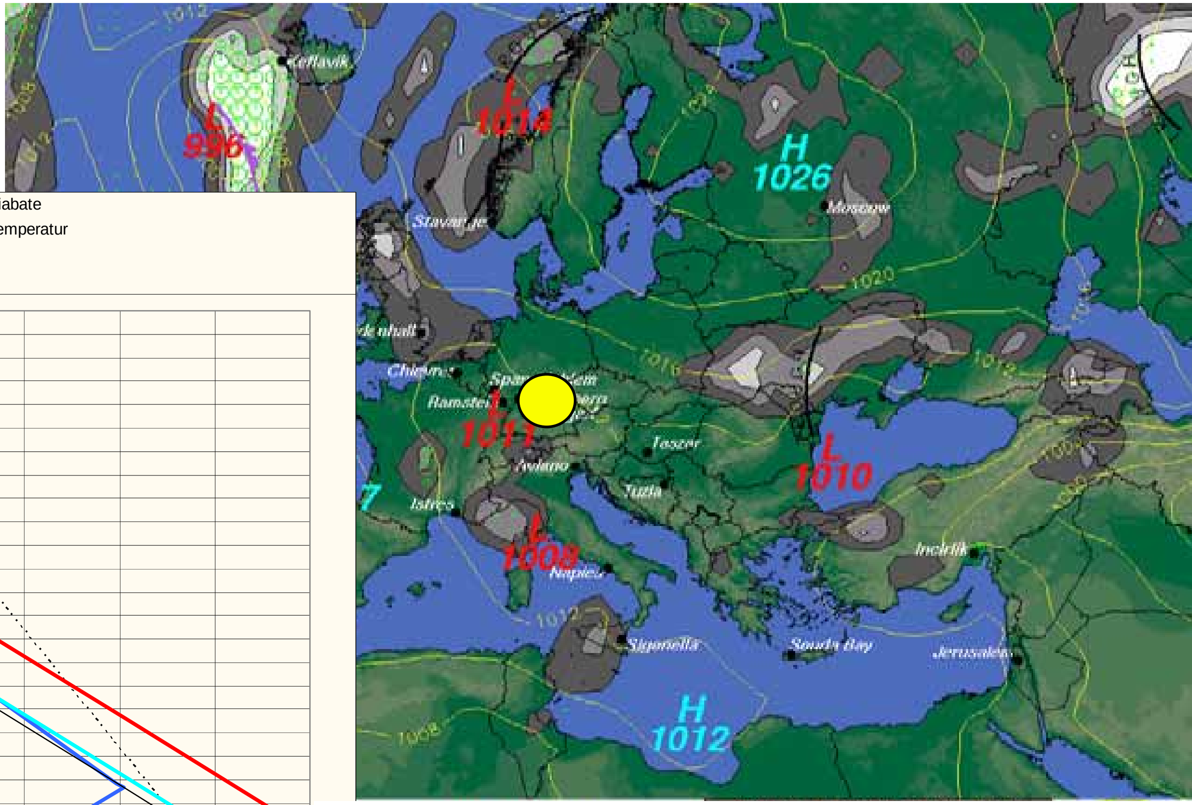
= during Coburg Gliding Competition (Bayerische Segelflugmeisterschaften der Club- und Doppelsitzerklasse)

Area: restricted to 49.5-51.2°N / 10-12.8°E

Prof. Dr. Alfred Ultsch
Datenbionik
ultsch@informatik.uni-marburg.de

Weather during period



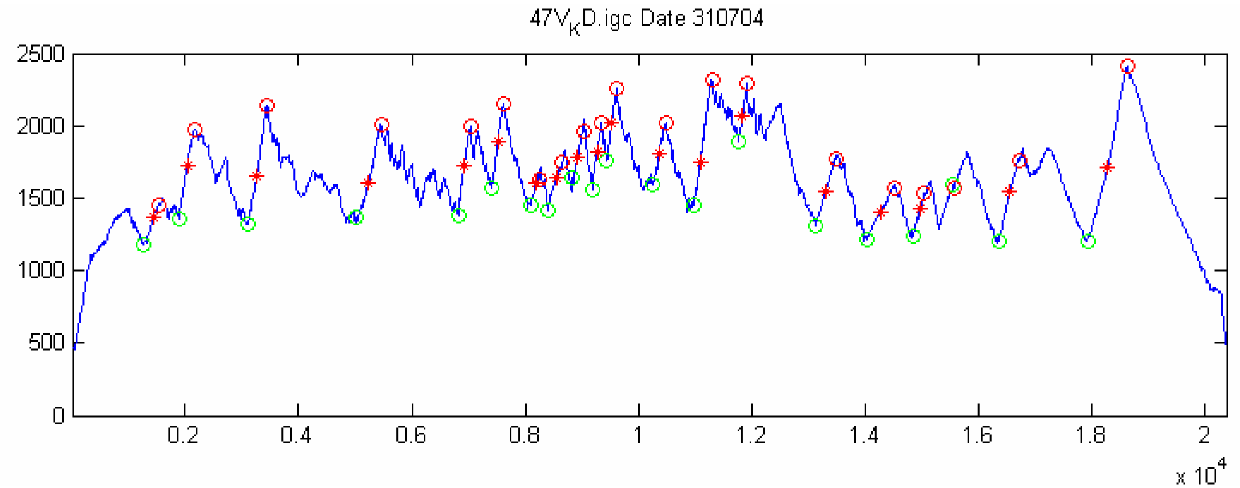


- mainly High pressure regime,
- domintaed by East situation
- thermal strengths prognosed at Coburg (A.Ultsch): typ. 2m/sec and more

- competition tasks of 300-500km fulfilled by most pilots on all days

Prof. Dr. Alfred Ultsch
Datenbionik
ultsch@informatik.uni-marburg.de

Identification of thermals

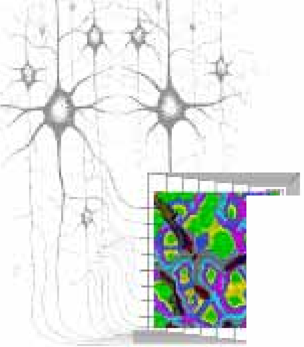


Problems:

- **noisy height due to turbulence**

- **„dolphin“- flight stile**

- **Minimum altitute gain as paramter (250m)**



Prof. Dr. Alfred Ultsch
Datenbionik
ultsch@informatik.uni-marburg.de

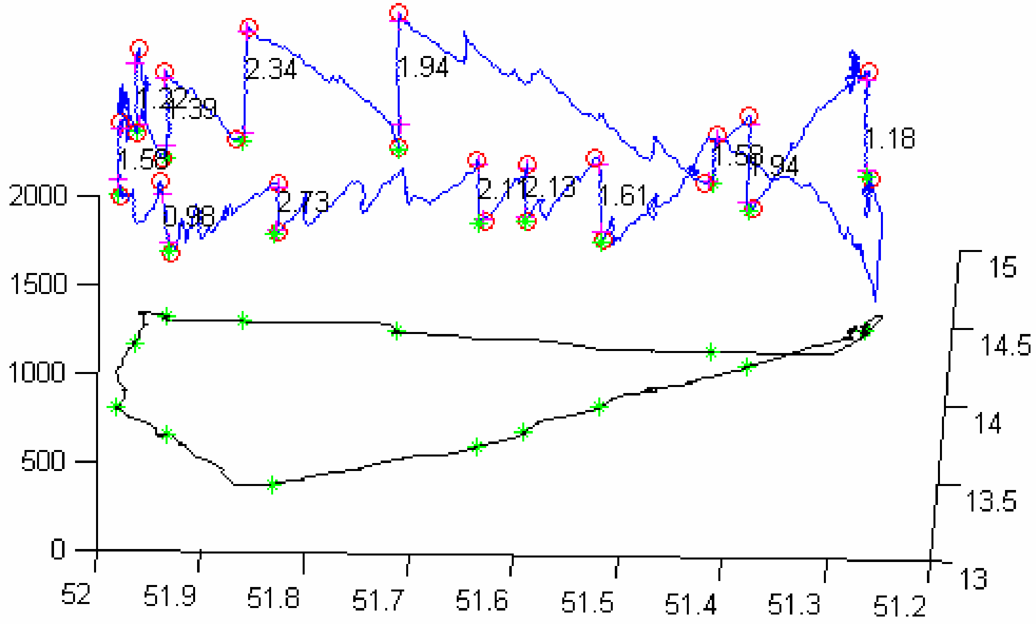
Thermal Strength

Footpoints of 9677 thermals of the total 21695 thermals in area around Coburg

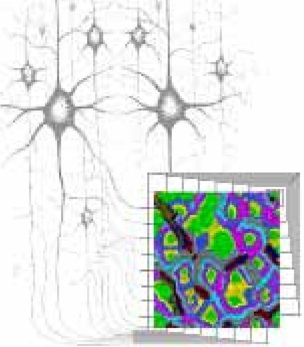
Filtered for uniqueness (no two pilots in the same thermal)

ThermalCoreStrength = ThermalCoreHeightGain/ThermalCoreDuration;

Core = central 80% of thermal time



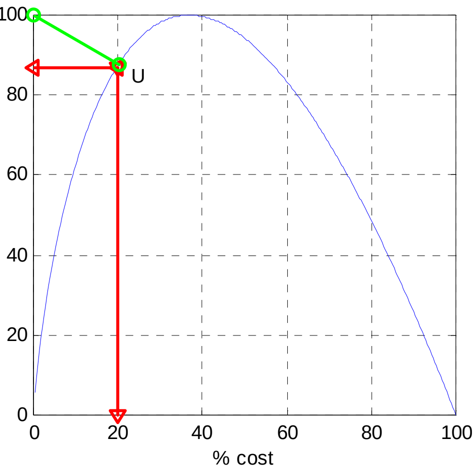
Prof. Dr. Alfred Ultsch
Datenbionik
ultsch@informatik.uni-marburg.de



Estimation of data density

Pareto Density Estimation (PDE) a kernel based density estimation with fixed kernel (Parzen window).

Kernel such that entopic yiel of subset is optimized



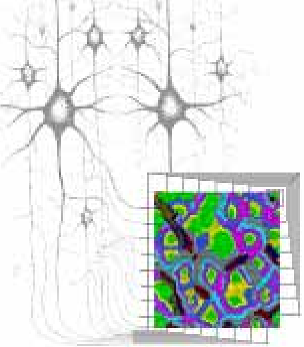
Properties of Pareto Density Estimation PDE:

- **Optimal density estimation of Gauss mixture models (GMM)**
- **very good to analyze overlapping**
- **shows modal points**
- **in particular good for the detection of clusters**

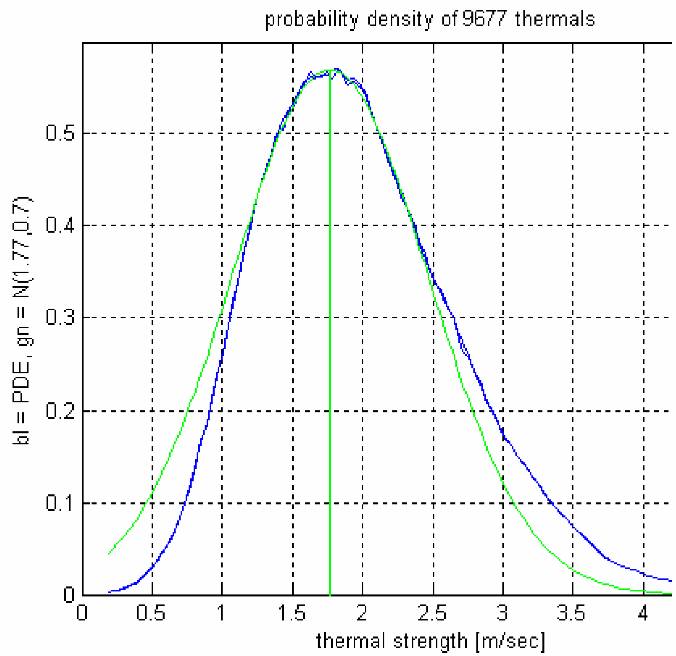
Applications:

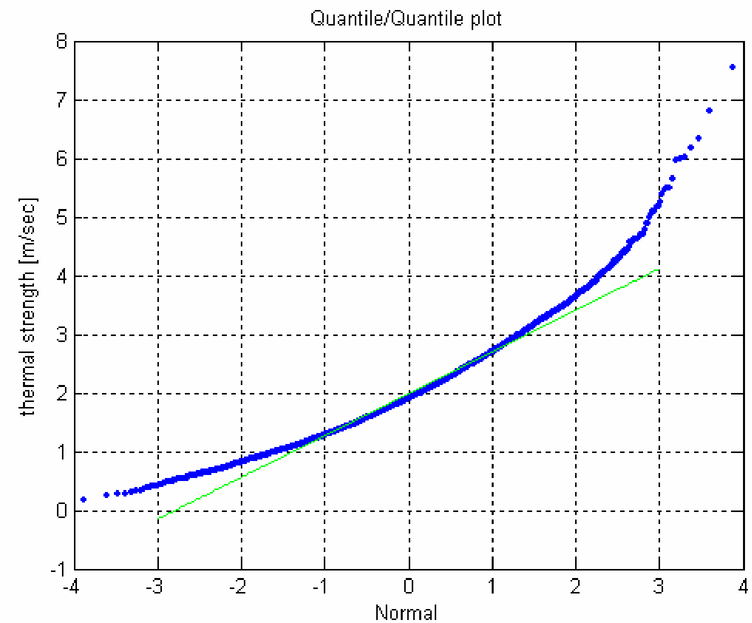
- **– one dimensional data: PDEplot –approximation of probability density**
- **– two dim data PDEscatter**
- **– high dimensional data: P-Matrix**

Prof. Dr. Alfred Ultsch
Datenbionik
ultsch@informatik.uni-marburg.de



Distribution of thermal strength





Comparison to Normal distribution:

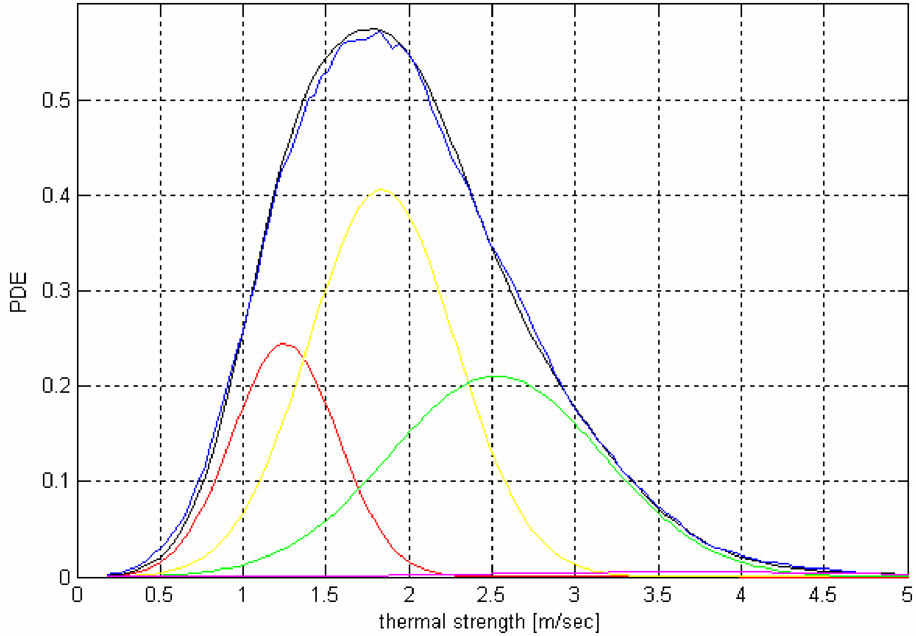
⇒Definitive not a Normal (Gaussian) distribution

⇒To many big ones , to few small ones

⇒Smooth and systematic deviation from Normal distribution

Prof. Dr. Alfred Ultsch
Datenbionik
ultsch@informatik.uni-marburg.de

Gauss Mixture Model

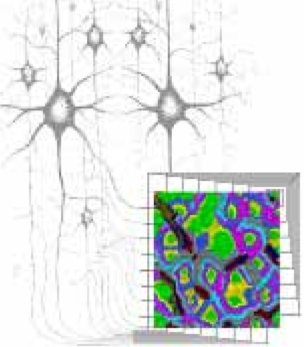


From 1…10 Mixtures wird tried´:

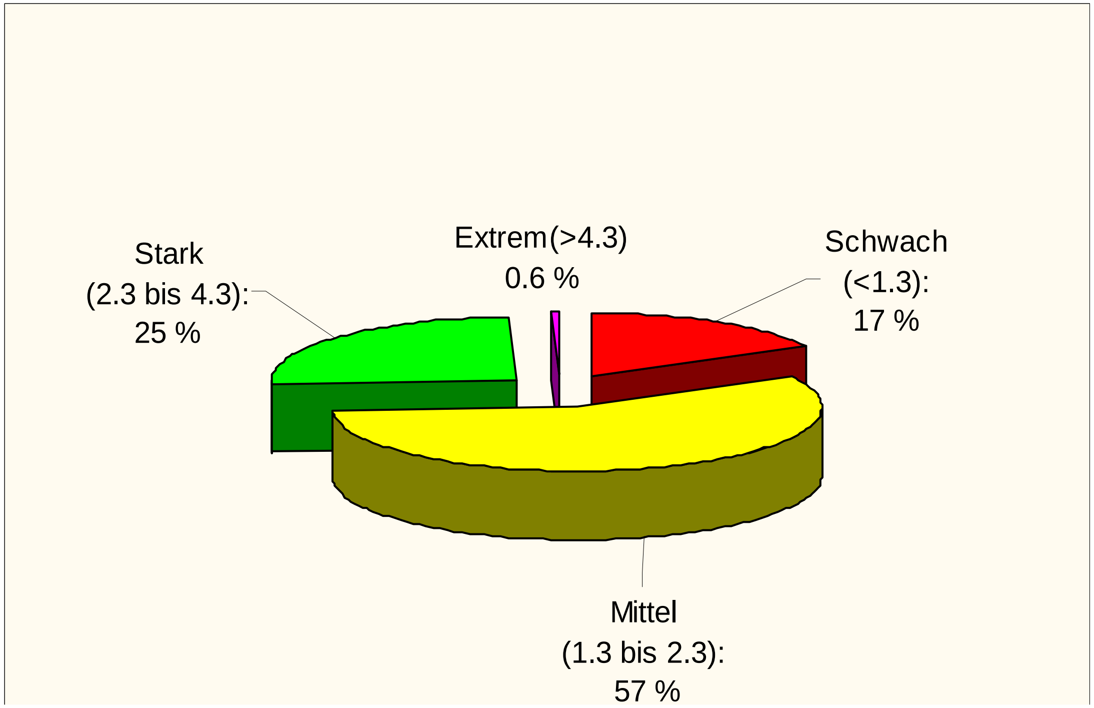
Best model has 4 Gaussians.

Mean	STD	Weight
1.25	0.32	0.19
1.84	0.44	0.45
2.52	0.64	0.34
3.69	1.18	0.02

Prof. Dr. Alfred Ultsch
Datenbionik
ultsch@informatik.uni-marburg.de



4 Classes of Thermals?

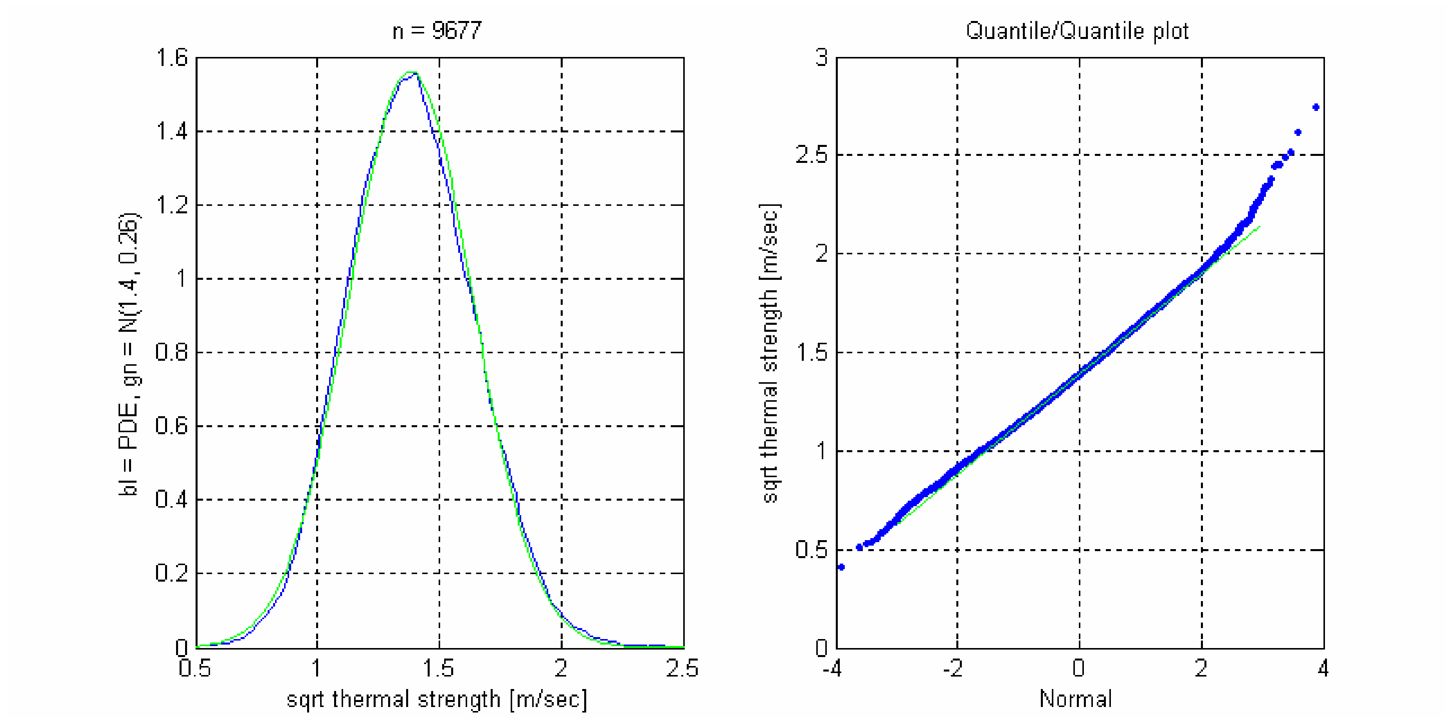


Maximum likelihood decision defines 4 classes of thermals:

Weak	**<1.3 [m/sec]**
Average	**1.3 ... 2.3[m/sec]**
Strong	**> 13 [m/sec]**
Extreme	**> 4.3 [m/sec]**

Prof. Dr. Alfred Ultsch
Datenbionik
ultsch@informatik.uni-marburg.de

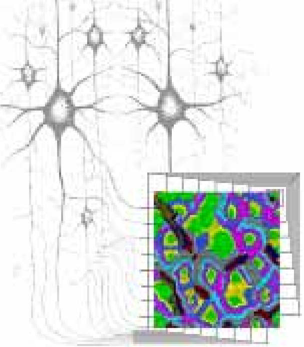
Square root of thermal strength



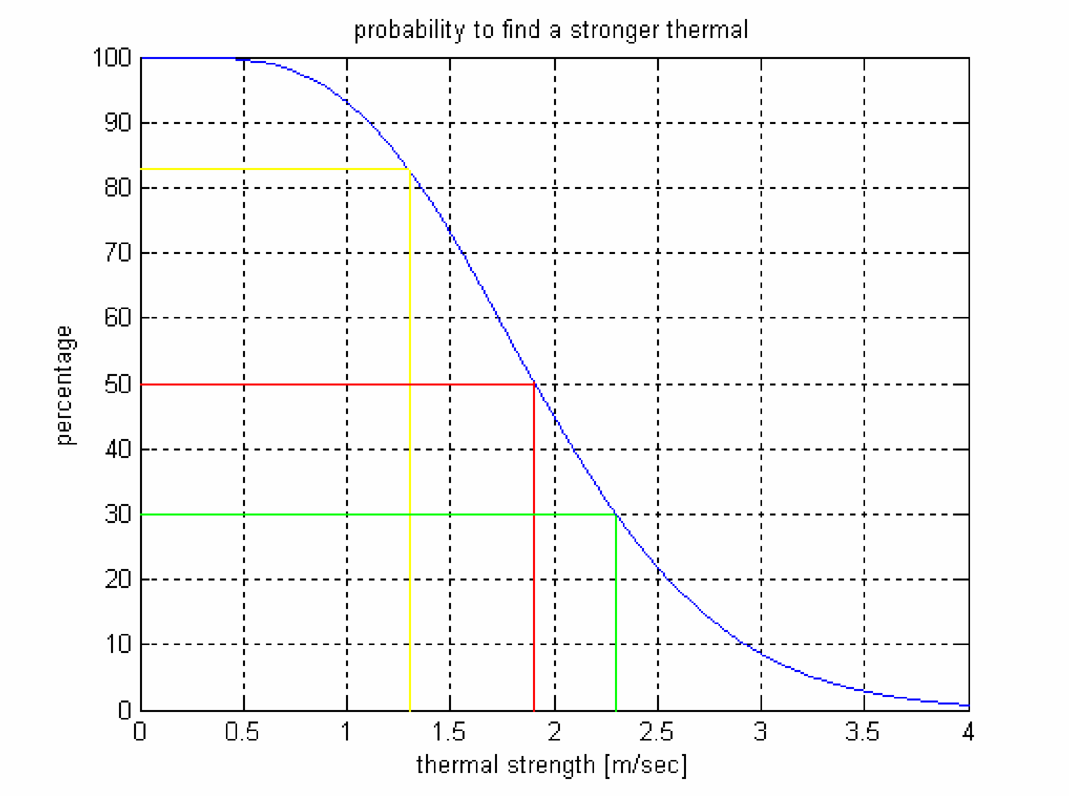
Surprising fit to Normal distribution !

=> Simple model of probability to find a thermal

Prof. Dr. Alfred Ultsch
Datenbionik
ultsch@informatik.uni-marburg.de

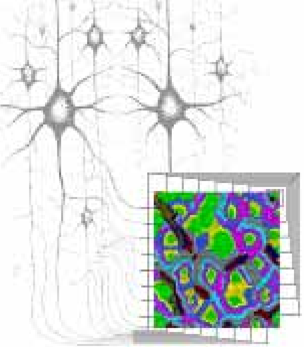


Probability Distribution of Thermals



Model: square root Normal distribution

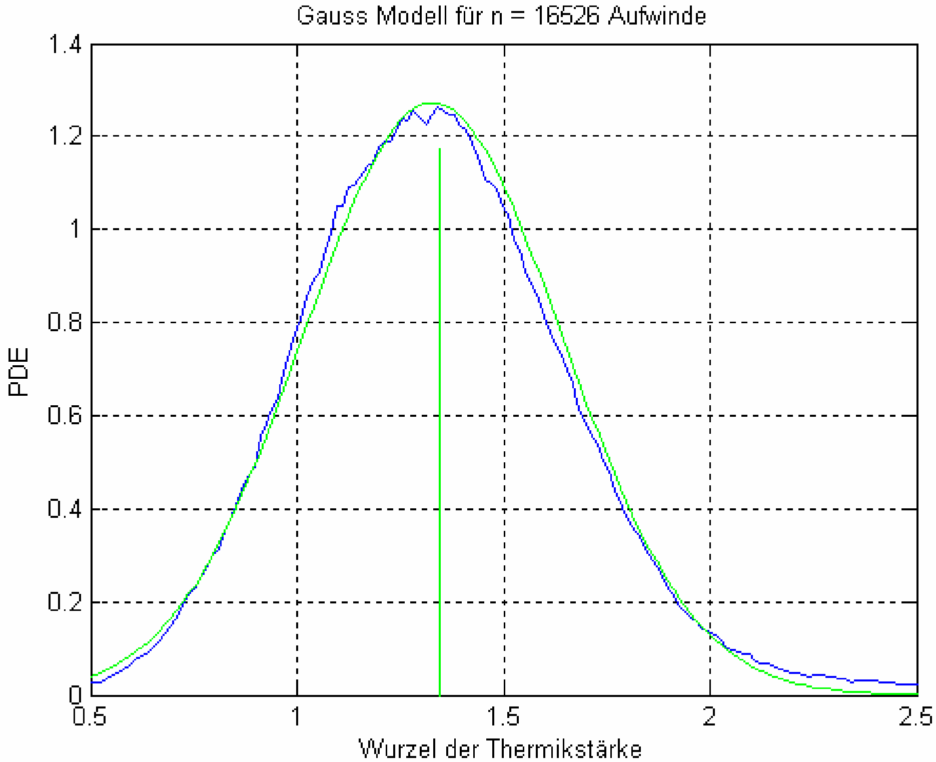
Prof. Dr. Alfred Ultsch
Datenbionik
ultsch@informatik.uni-marburg.de



Discussion: What is the reason behind sqare normal distribution?

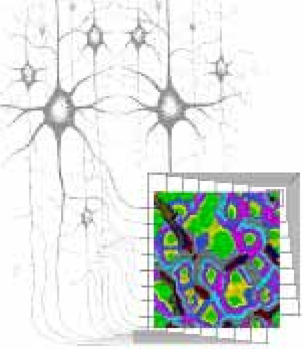
1) **data processing:** other authors (e.g. Jon Meis Lüsse 1996) used 80 m as min height gain to identify thermals in IGC files

- Distribution of sqrt(CoreStrength) using this limit: n= 16,526



⇒ Fit to sqrt-normal even better!

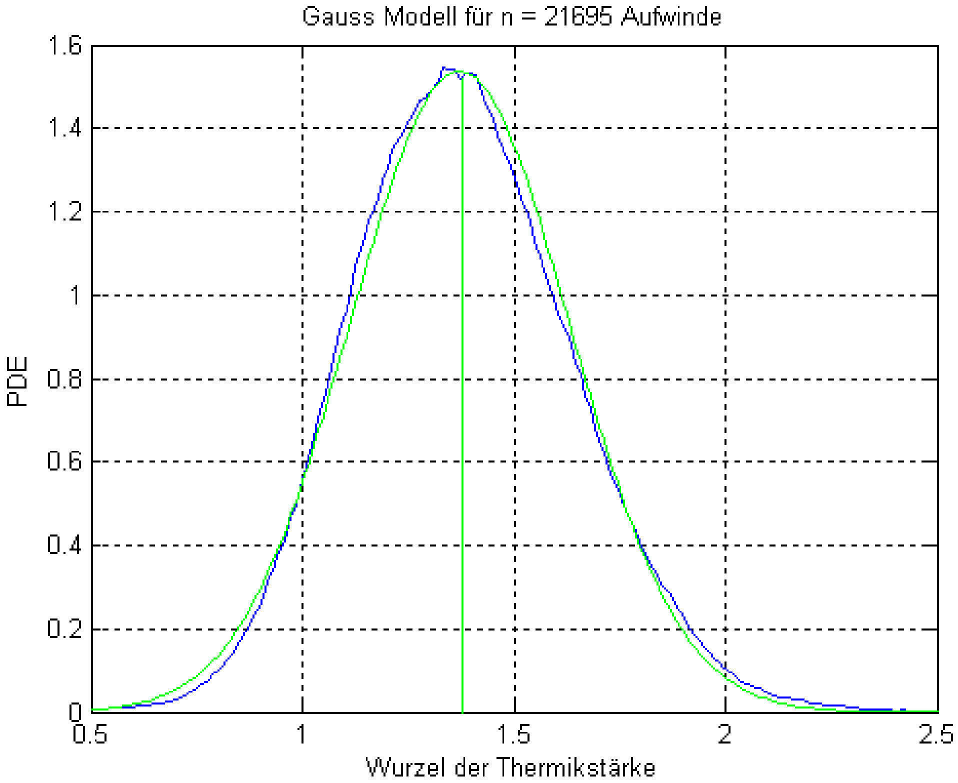
Prof. Dr. Alfred Ultsch
Datenbionik
ultsch@informatik.uni-marburg.de



Discussion: What is the reason behind sqare normal distribution?

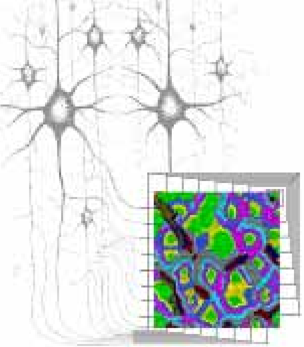
2) **Bias by glider pilots: they use only the strong thermals**

- **Distribution of all 21,695 thermals sqrt(CoreStrength)**



=> **Model fits the data!**

Prof. Dr. Alfred Ultsch
Datenbionik
ultsch@informatik.uni-marburg.de



Discussion: What is the reason behind sqare normal distribution?

3) Meteorological reasons ?

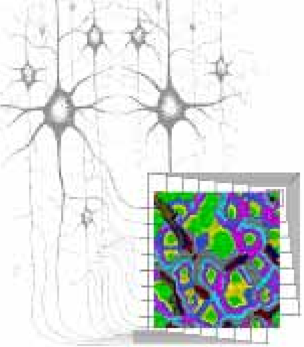
⇒ Left to discuss here

Proposal:

- **Nature generates thermals using a solar heating plate .**
- **The diameter D of this heating plate is drawn from a Normal distribution Gauss(m, s)**
- **The strength S of the thermal in m/sec is direct proportional to area of the heating plate:**

$$S = c\,D^2$$ **with D from Gauss(m, s)**

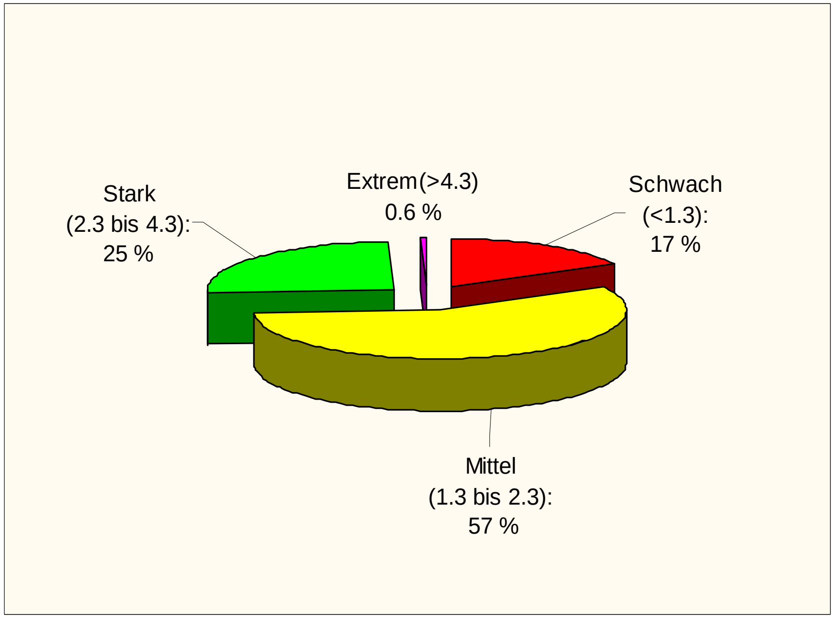
Prof. Dr. Alfred Ultsch
Datenbionik
ultsch@informatik.uni-marburg.de



Application

-Compare distribution of thermals of a flight to expected distributions

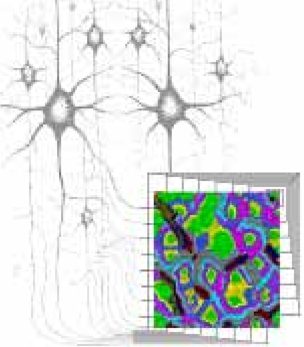
e.g ¼ should be strong thermals



- give precise semantics to „weak“, „average“ „strong“ in gliding forcasts

- model can be used for prediction

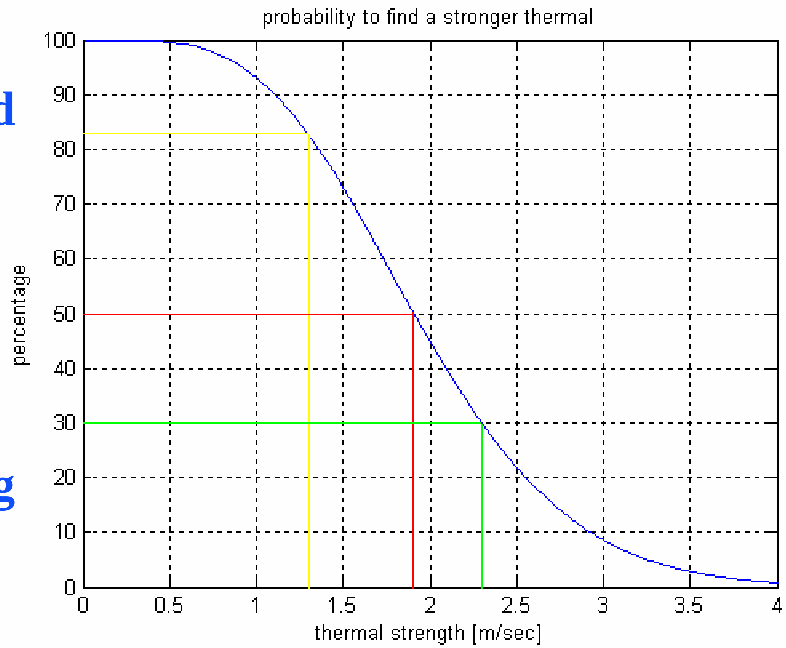
Prof. Dr. Alfred Ultsch
Datenbionik
ultsch@informatik.uni-marburg.de



Rules of Thumb

derived rules of thumb for cross country soaring:

- Don‘t use thermals below 1 m/sec. The probability to find a bette one is over 90% !
- Go for thermals with at least 1.4 m/sec
- If the integrator shows less than 1.9 m/sec consider flying on. The probability to find a better thermal is > 50%
- Definitivly take thermals with 2.3 or better !
- Use thermals < 1.3 m/sec only in difficult situations (wether/ outlanding)

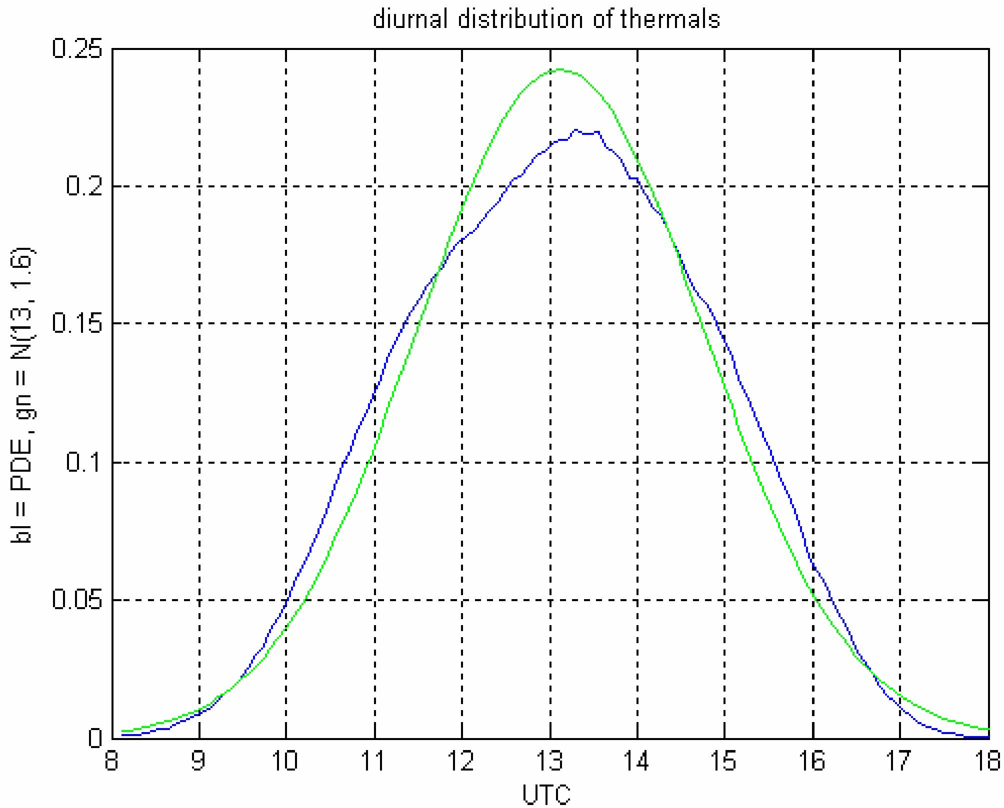


Prof. Dr. Alfred Ultsch
Datenbionik
ultsch@informatik.uni-marburg.de

Future Work

1) Diurnal Distribution of Thermals

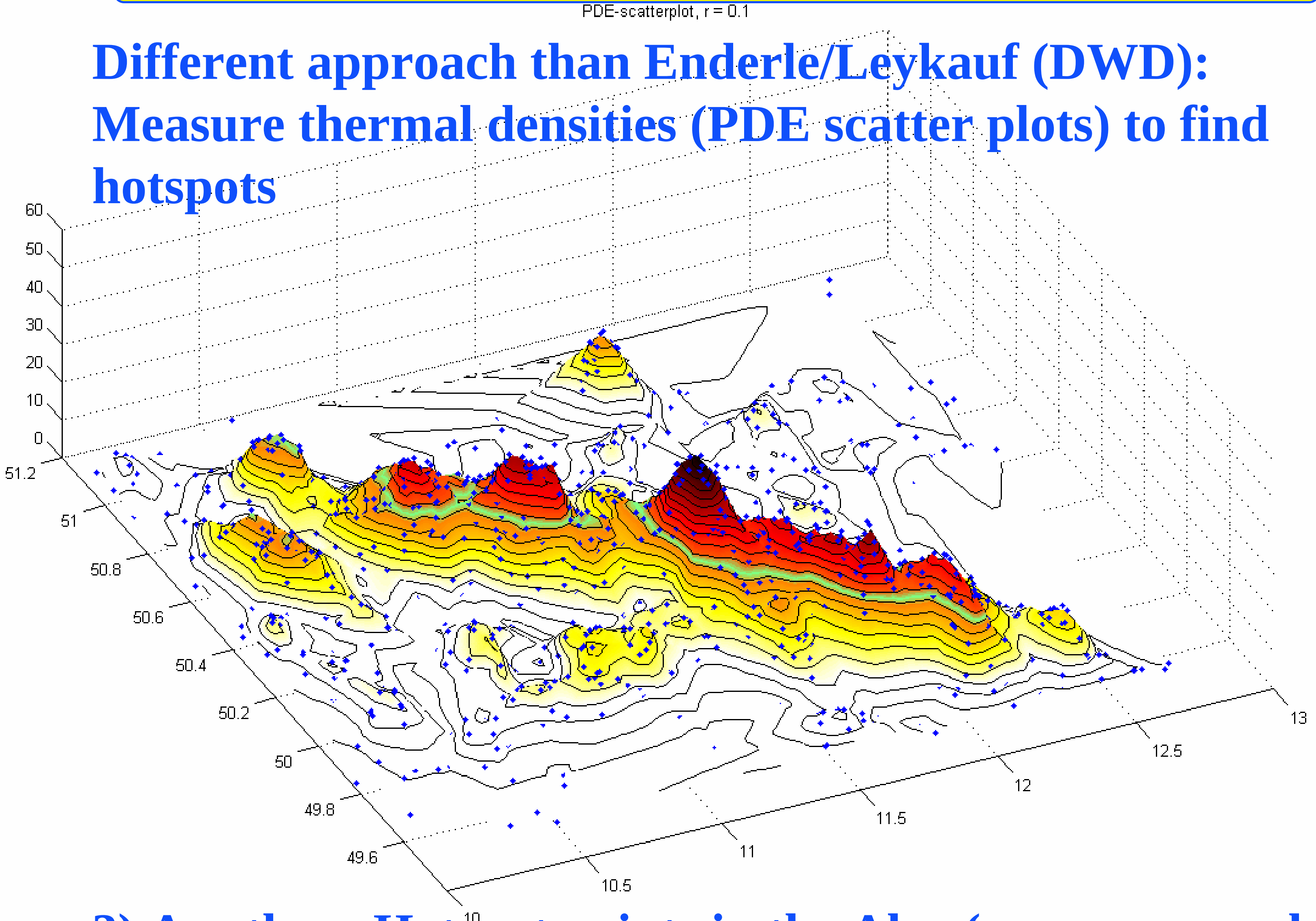
A first look at circadian distribution of occurences:



Prof. Dr. Alfred Ultsch
Datenbionik
ultsch@informatik.uni-marburg.de

Future Work 2: Are there Hotspots?

Different approach than Enderle/Leykauf (DWD): Measure thermal densities (PDE scatter plots) to find hotspots



3) Are there Hotspot points in the Alps (sponsors sought)?

Summary

The empirical distribution of a large collection of thermals in typical cross county areas /weather is consistent with either

1) **four types of thermals**
   (small, average large, extra) or

2) **a Sqrt-Normal Distribution**

Open question: Is there a meterological law behind this findings?

(e.g.: $S = c\, D^2$ 👌 with D from Gauss(m, s))

Prof. Dr. Alfred Ultsch
Datenbionik
ultsch@informatik.uni-marburg.de