Nonlinear analysis of very long time series: opportunities and challenges

NIBIO
NORWEGIAN INSTITUTE OF BIOECONOMY RESEARCH

Holger Lange, Marlene Schramm, Sebastian Sippel and Junbin Zhao

Department for Terrestrial Ecology,
Norwegian Institute of Bioeconomy, Ås, Norway; holger.lange@nibio.no

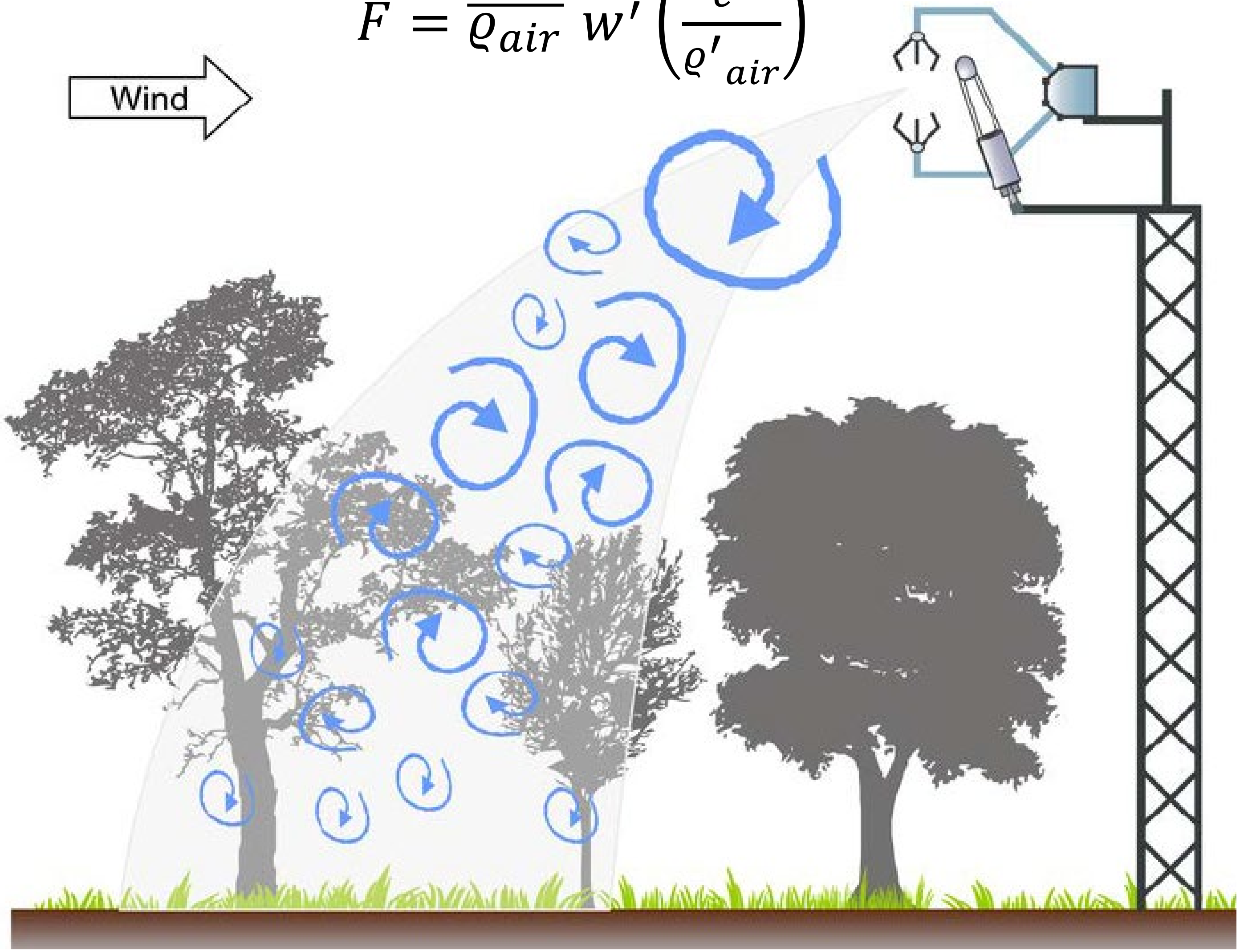
- Why long time series are a blessing
- Eddy Covariance and the ICOS network
- Our pilot measurements at Hoxmark
- Analysis: Tarnopolski diagrams, q-complexity
- Conclusions

Data-oriented approach to difficult problems in environmental science

- Characterize the dynamical system through its observables: «Knowledge through observations» (ICOS' motto)
- Signal and Noise characterization / identification / separation
- What is the optimal temporal resolution («maximizing signal, minimizing noise»)?
- Analysis methods from nonlinear dynamics
- Caveat: many methods require very long time series…
- …we have them!

The Eddy Covariance Method:
vertical fluxes can be presented as a covariance of the vertical wind velocity and the gas concentration

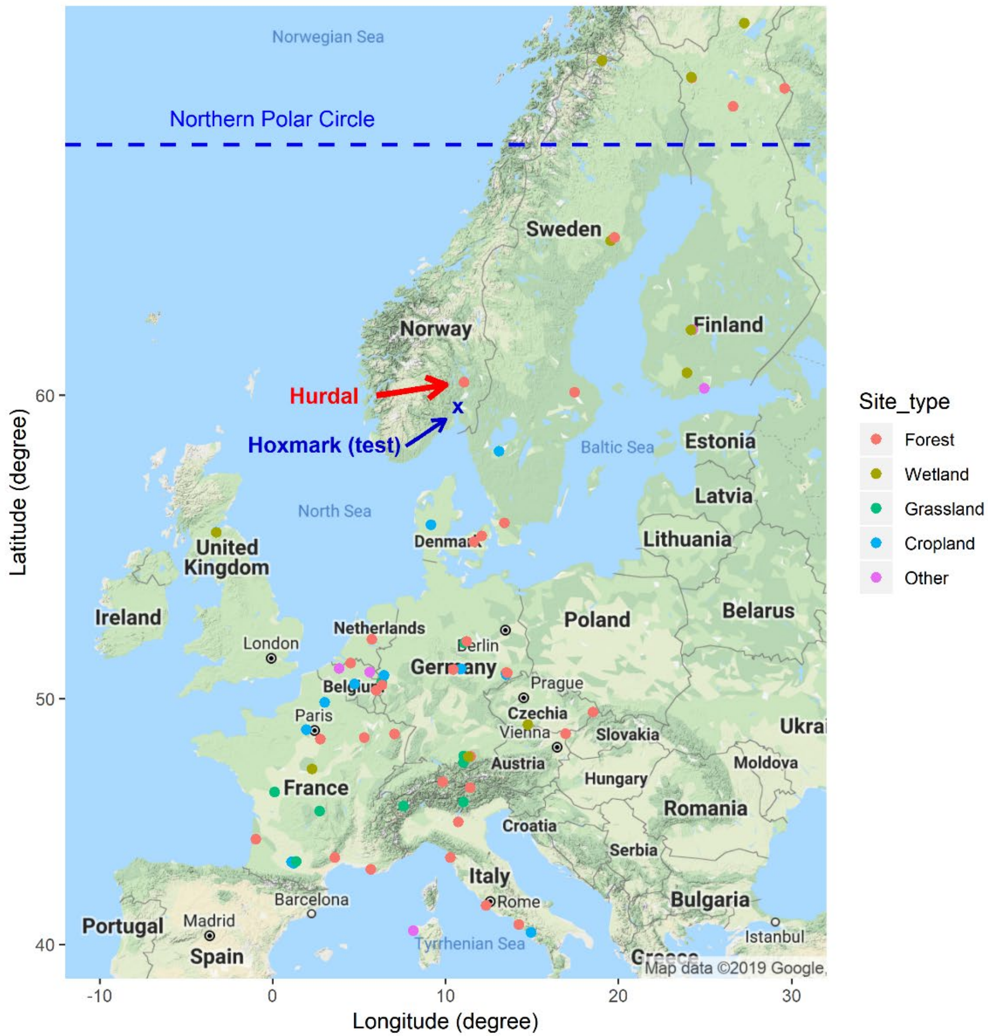
$$F = \overline{\varrho_{air}}\ \overline{w'\left(\frac{c'}{\varrho'_{air}}\right)}$$



Important Eddy Covariance variables

- Carbon Dioxide concentration
- Windspeed and –direction (3D)
- Water vapor
- Latent Heat
- Sensible Heat
- Net Radiation
- **Net Ecosystem Exchange** (NEE)
- …

Map of ICOS Stations

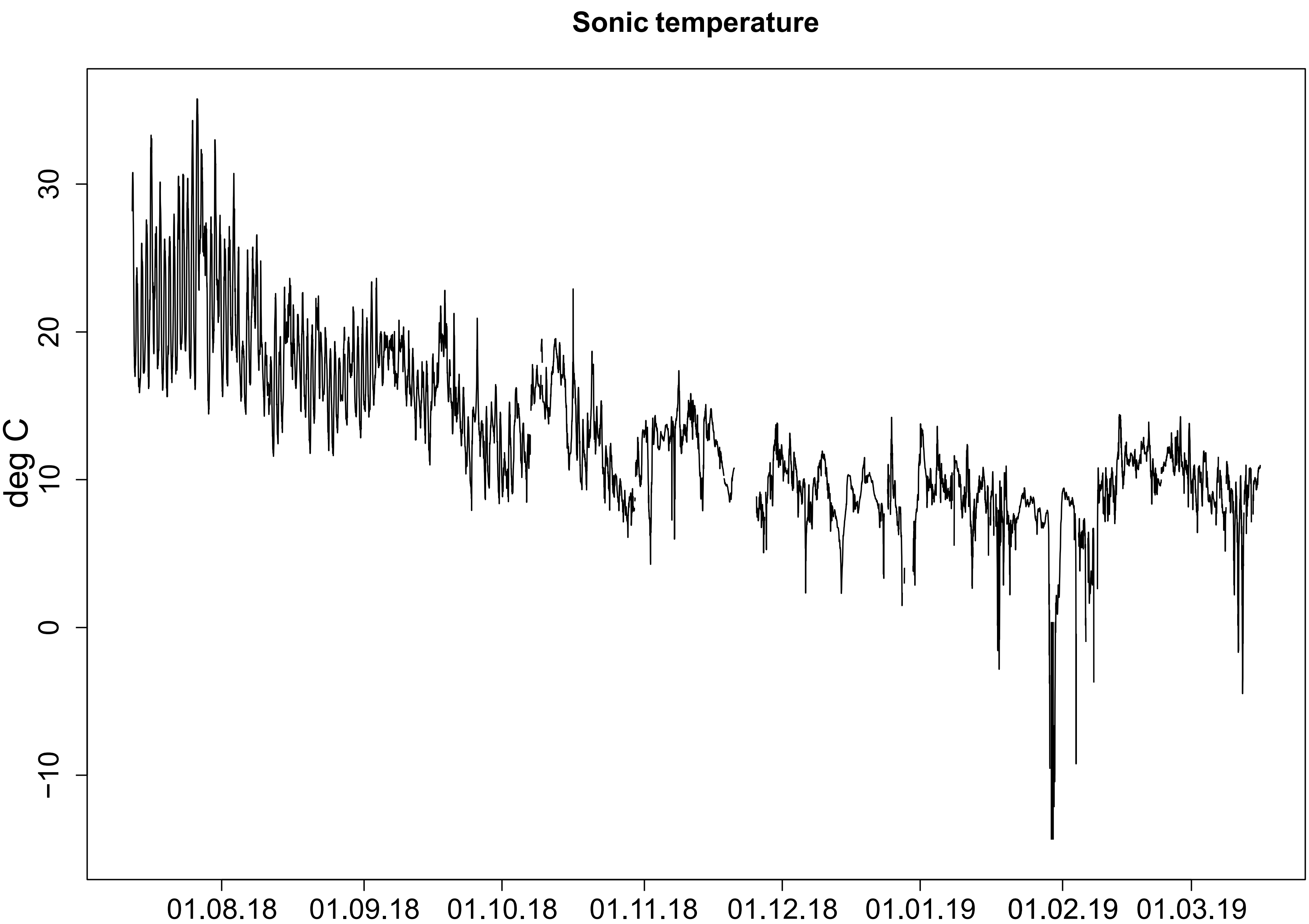


Our «pilot» EC system at Hoxmark

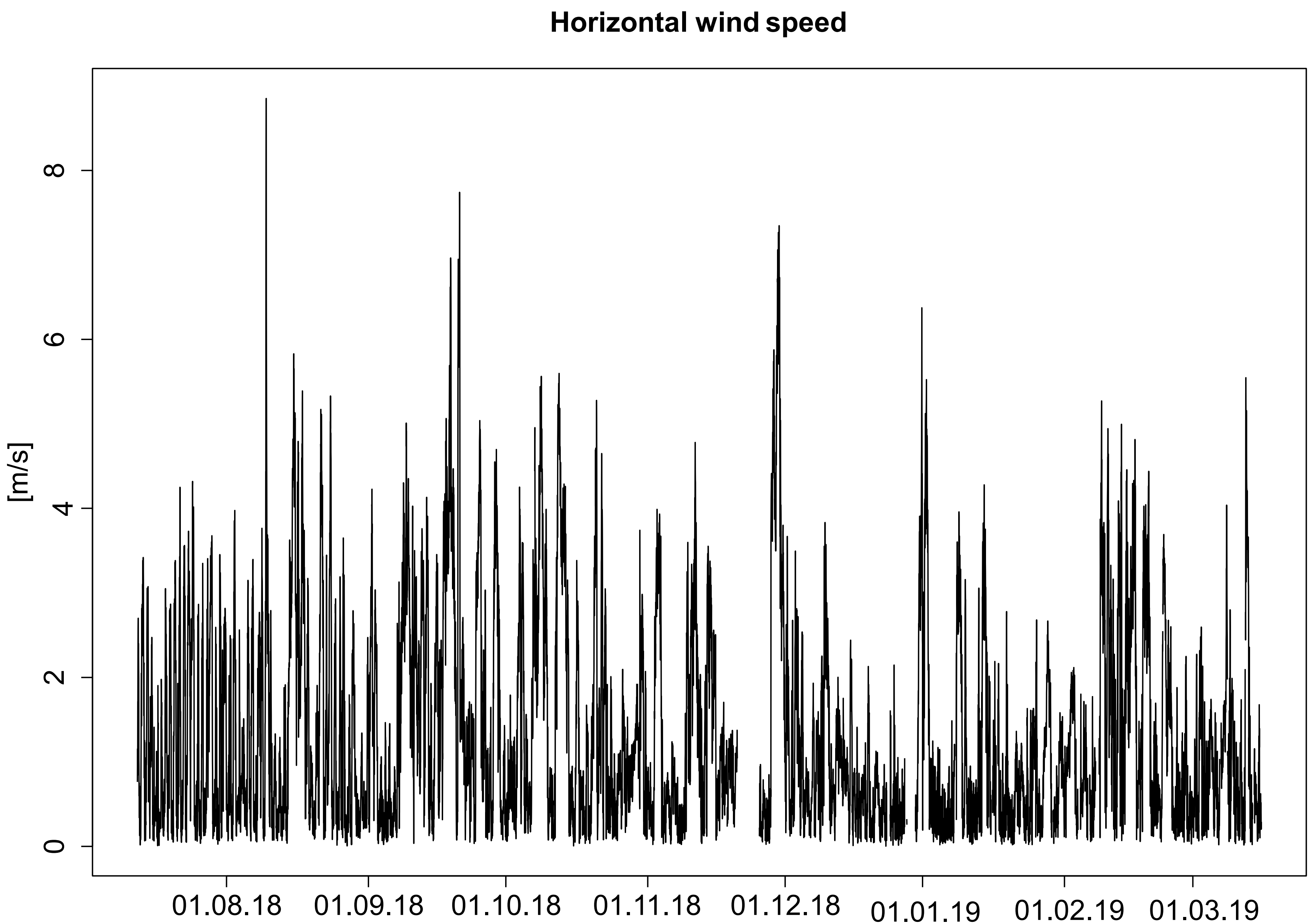
NIBIO
NORSK INSTITUTT FOR
BIOØKONOMI

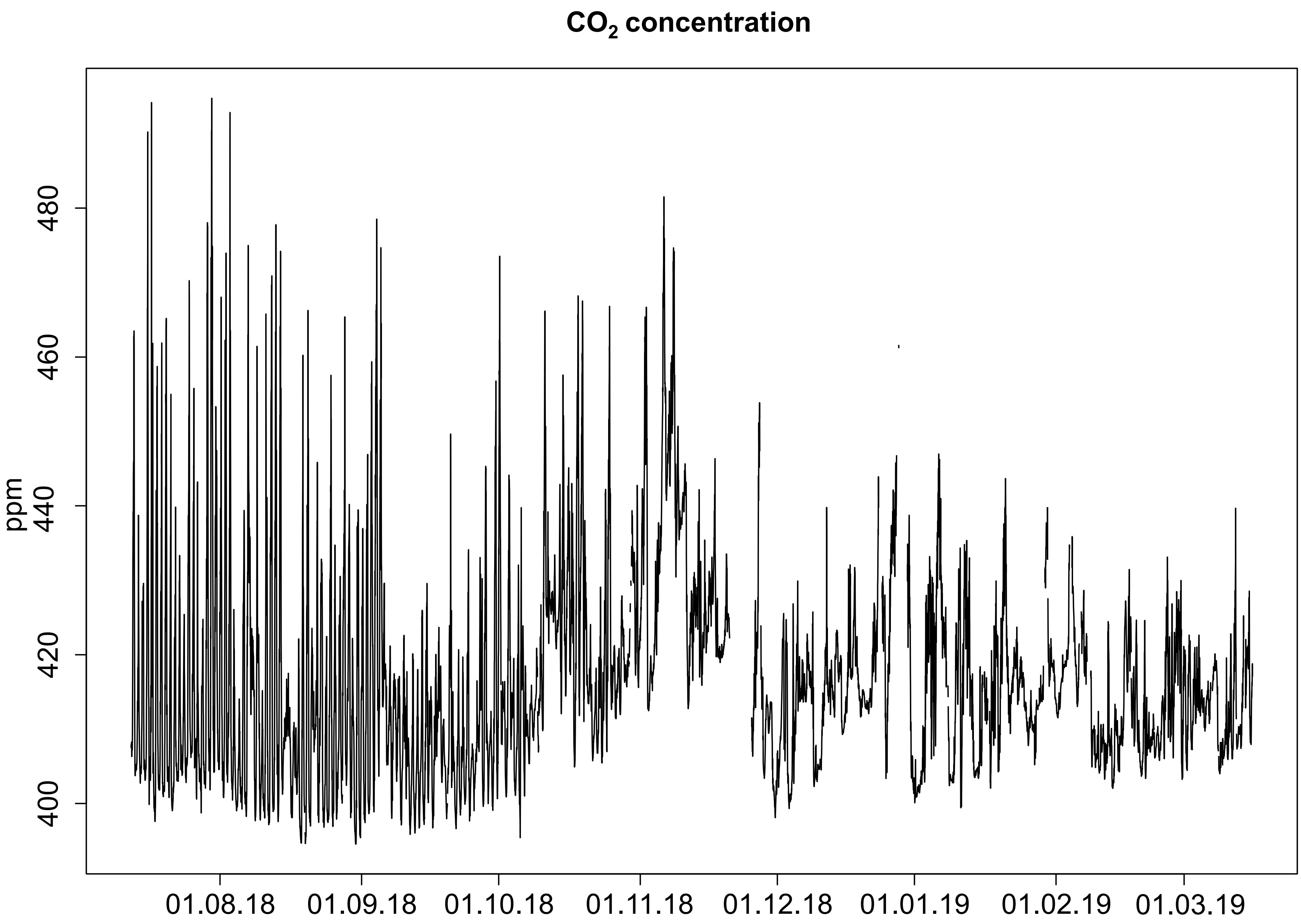
Our time series from Hoxmark

- Sonic: Gill HS-50
- IrGA: Licor RS-7200
- 21 variables @ 20 Hz measurement frequency
- First observation day 13.7.2018
- Today (8.4.2019): Time series with $> 4 \cdot 10^8$ values
- Additional 1 728 000 values per variable per day



Sonic temperature
deg C
30
20
10
0
−10
01.08.18
01.09.18
01.10.18
01.11.18
01.12.18
01.01.19
01.02.19
01.03.19


Horizontal wind speed
[m/s]
0
2
4
6
8
01.08.18
01.09.18
01.10.18
01.11.18
01.12.18
01.01.19
01.02.19
01.03.19



CO_2 concentration
ppm
480
460
440
420
400
01.08.18
01.09.18
01.10.18
01.11.18
01.12.18
01.01.19
01.02.19
01.03.19

Finding the optimal temporal resolution: Decimation and Aggregation

Original resolution: 0.05 s

- **Decimation**: taking each nth value
- **Aggregation**: non-overlapping mean of each n values

$$(1 \leq n \leq 10^7)$$

Effective resolution: 0.05 s … 5 days

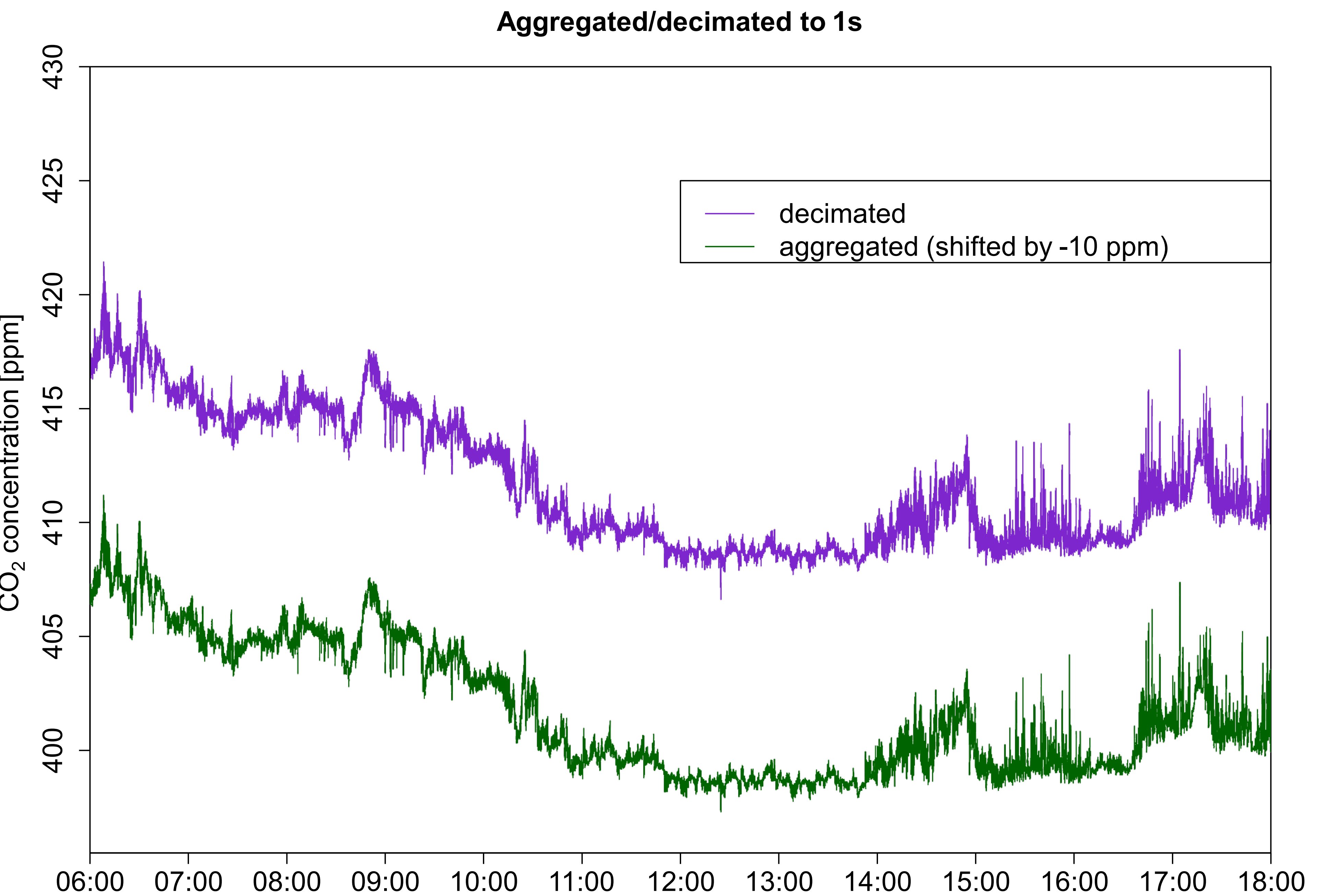
Characterizing properties of EC variables

Examples:

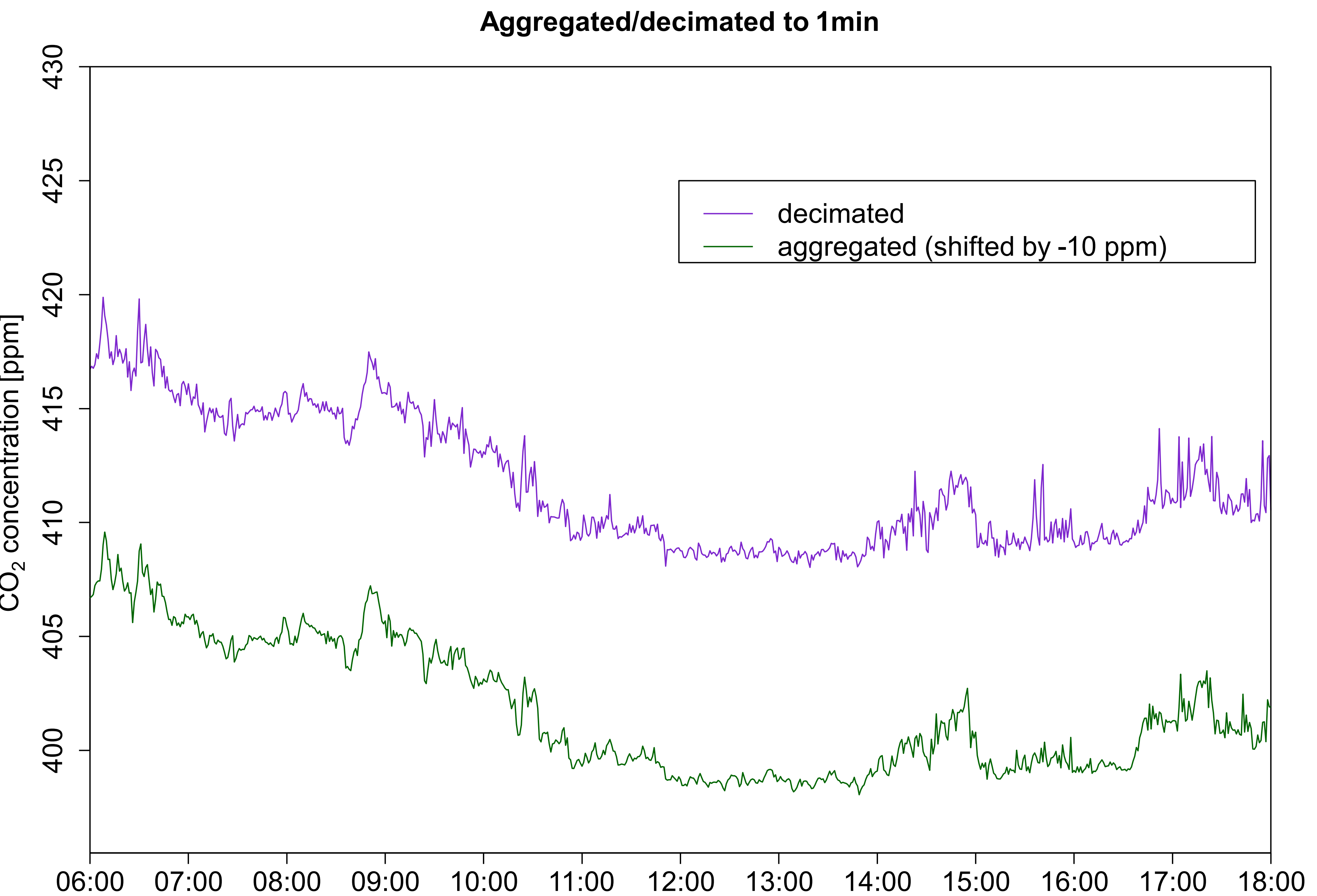
- **Tarnopolski diagrams**
- **q-Entropy and q-complexity**
- Renyi entropy and complexity
- MPR Complexity-Entropy plane
- Fisher-Entropy plane
- Hurst exponents
- …

Reference Noise Processes

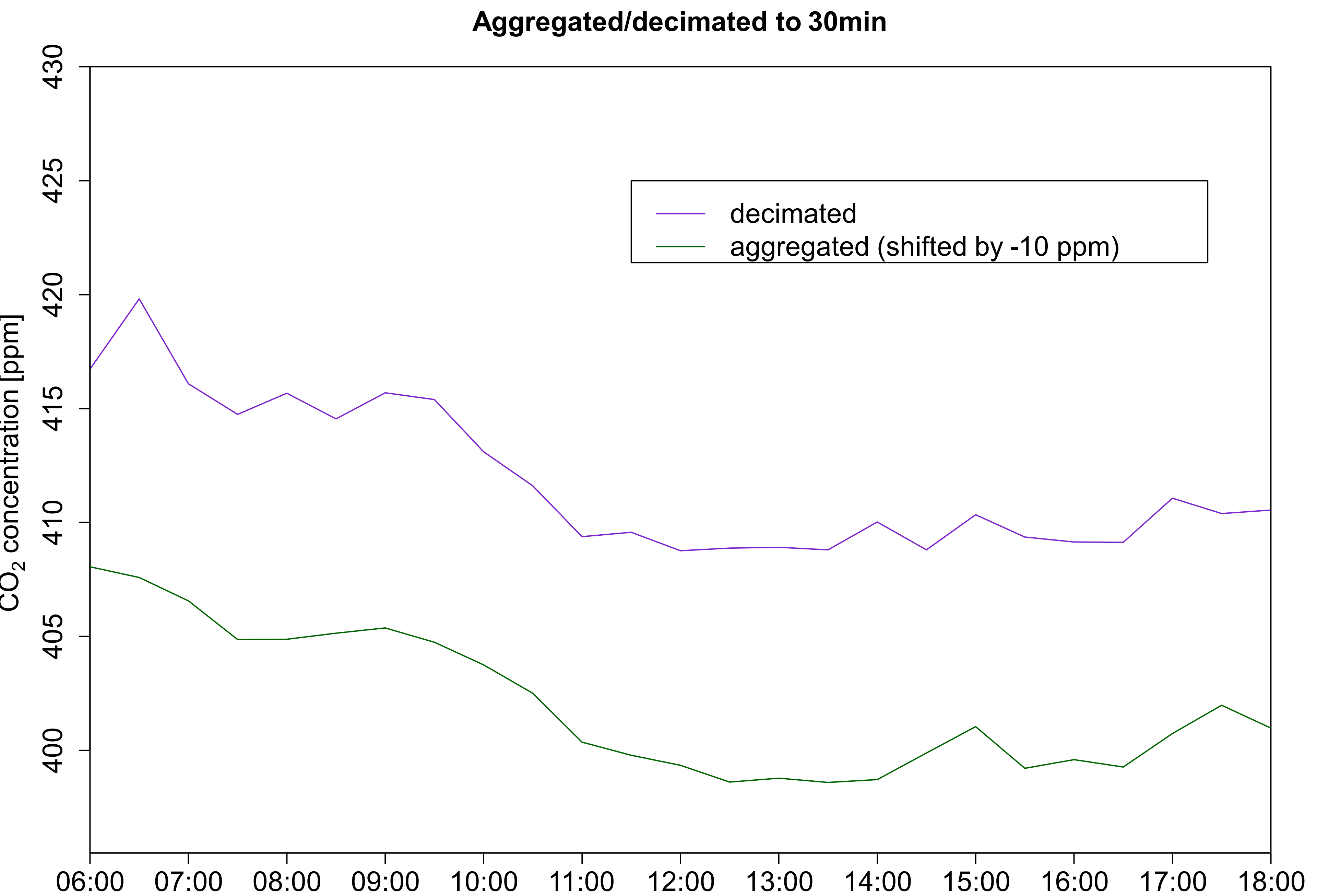
- k noise: long-range correlations, $P(f) \propto f^{-k}$
- fractional Brownian motion (fBm): Hurst parameter $H, 0 \leq H \leq 1$ $P(f) \propto f^{-(2H+1)}$
- fractional Gaussian noise (fGn)



Aggregated/decimated to 1s
decimated
aggregated (shifted by -10 ppm)
CO_2 concentration [ppm]
430
425
420
415
410
405
400
06:00
07:00
08:00
09:00
10:00
11:00
12:00
13:00
14:00
15:00
16:00
17:00
18:00



Aggregated/decimated to 1min
decimated
aggregated (shifted by -10 ppm)
CO_2 concentration [ppm]
430
425
420
415
410
405
400
06:00
07:00
08:00
09:00
10:00
11:00
12:00
13:00
14:00
15:00
16:00
17:00
18:00


Aggregated/decimated to 30min
decimated
aggregated (shifted by -10 ppm)
CO2 concentration [ppm]
430
425
420
415
410
405
400
06:00
07:00
08:00
09:00
10:00
11:00
12:00
13:00
14:00
15:00
16:00
17:00
18:00

Digging into the noise: Tarnopolski analysis

For a given time series, calculate two quantities:

- The Abbe value $\mathbb{A}$: ratio of the mean square successive difference to the variance

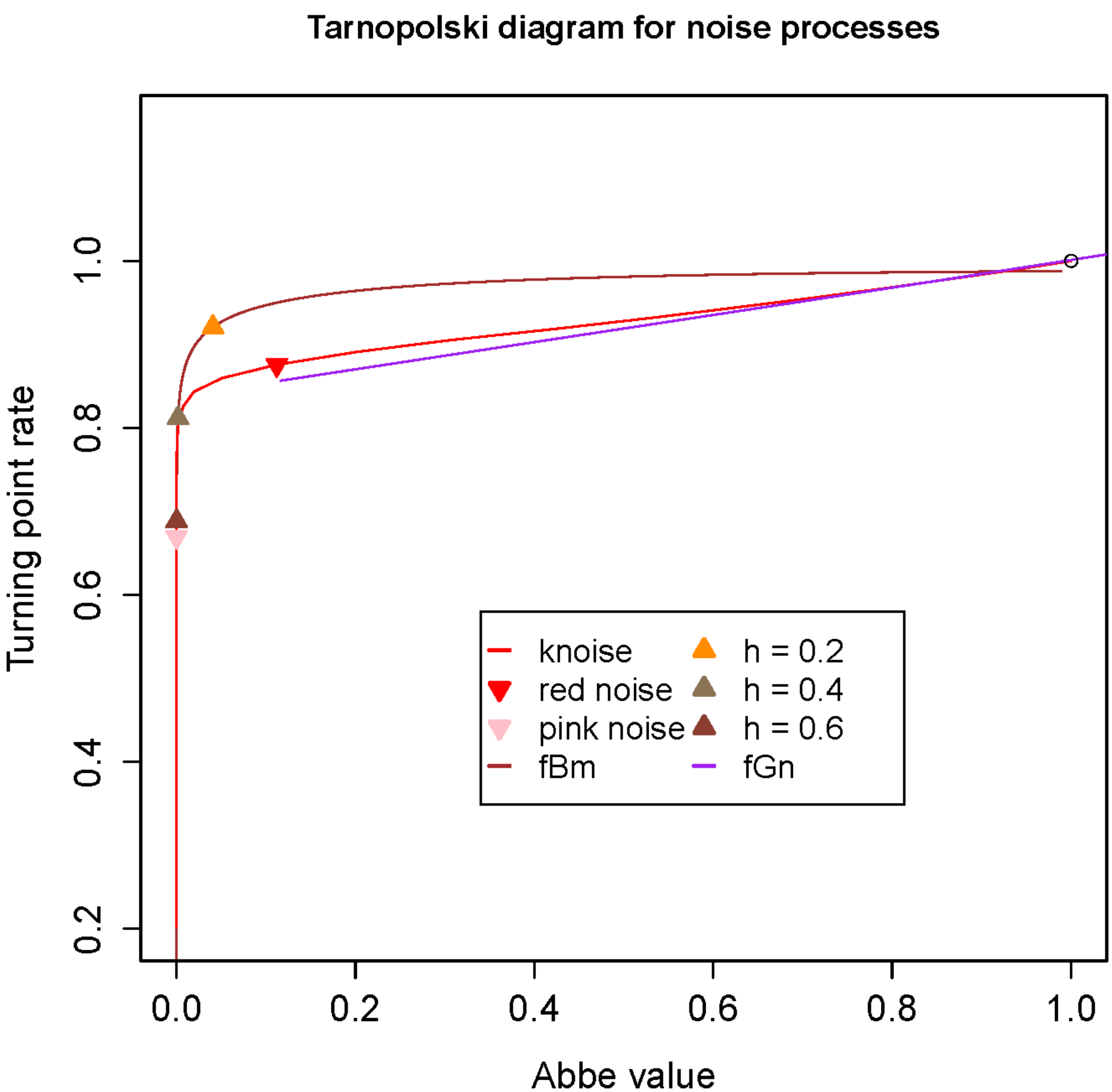
$$\mathbb{A} = \frac{\delta^2}{2s^2} = \frac{\frac{1}{n-1}\sum_{i=1}^{n-1}(x_{i+1}-x_i)^2}{\frac{2}{n}\sum_{i=1}^{n}(x_i-\hat{x})^2}$$

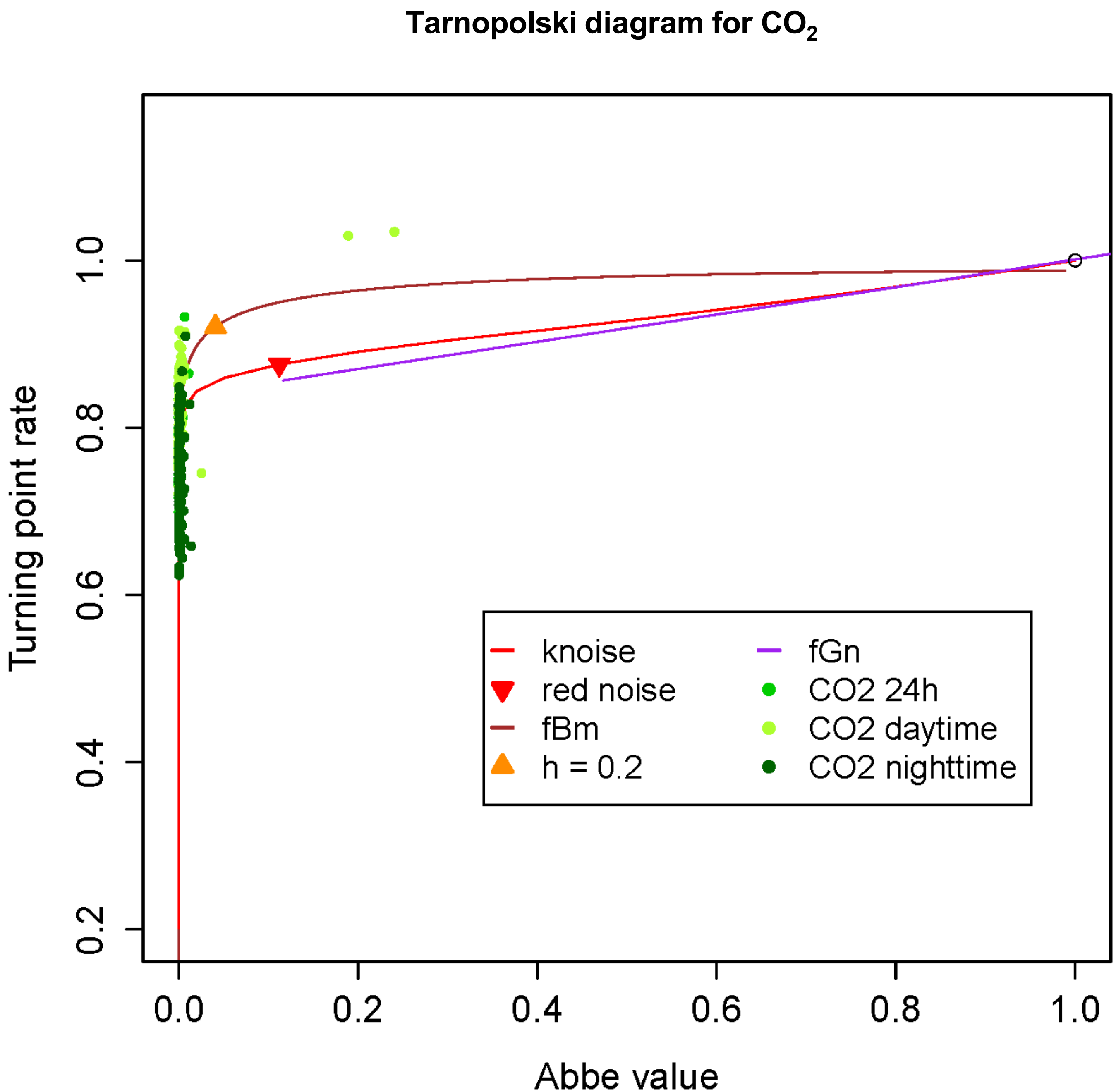
- The number of turning points $\mathbb{T}$: count the local «extrema» N_E

$$\mathbb{T} = \frac{3N_E}{2(n-2)}$$

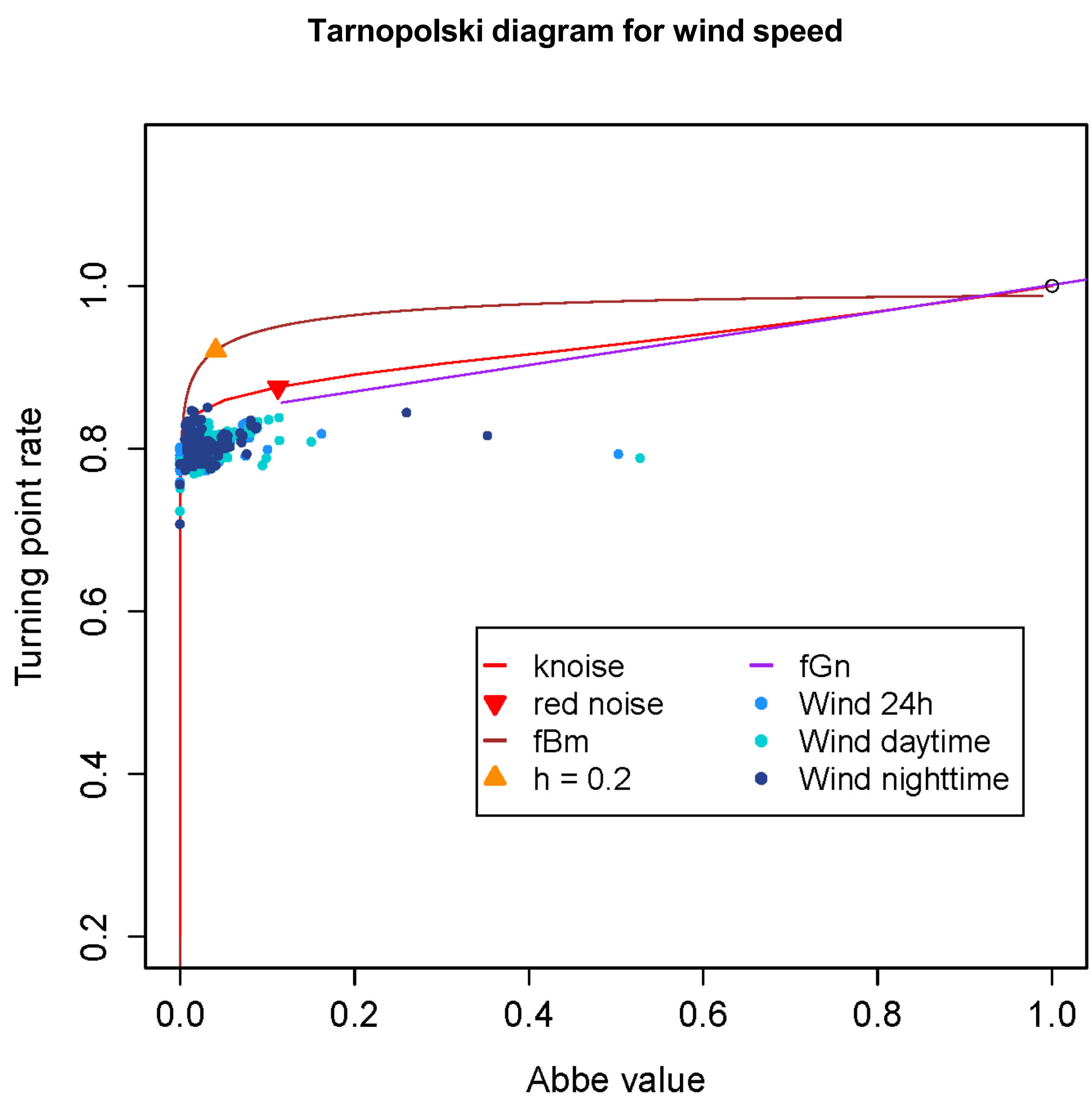
- Normalized such that for white noise, ($\mathbb{A}$, $\mathbb{T}$) = (1,1)
- Plot in a 2D diagram
- This easily distinguishes different types of noise

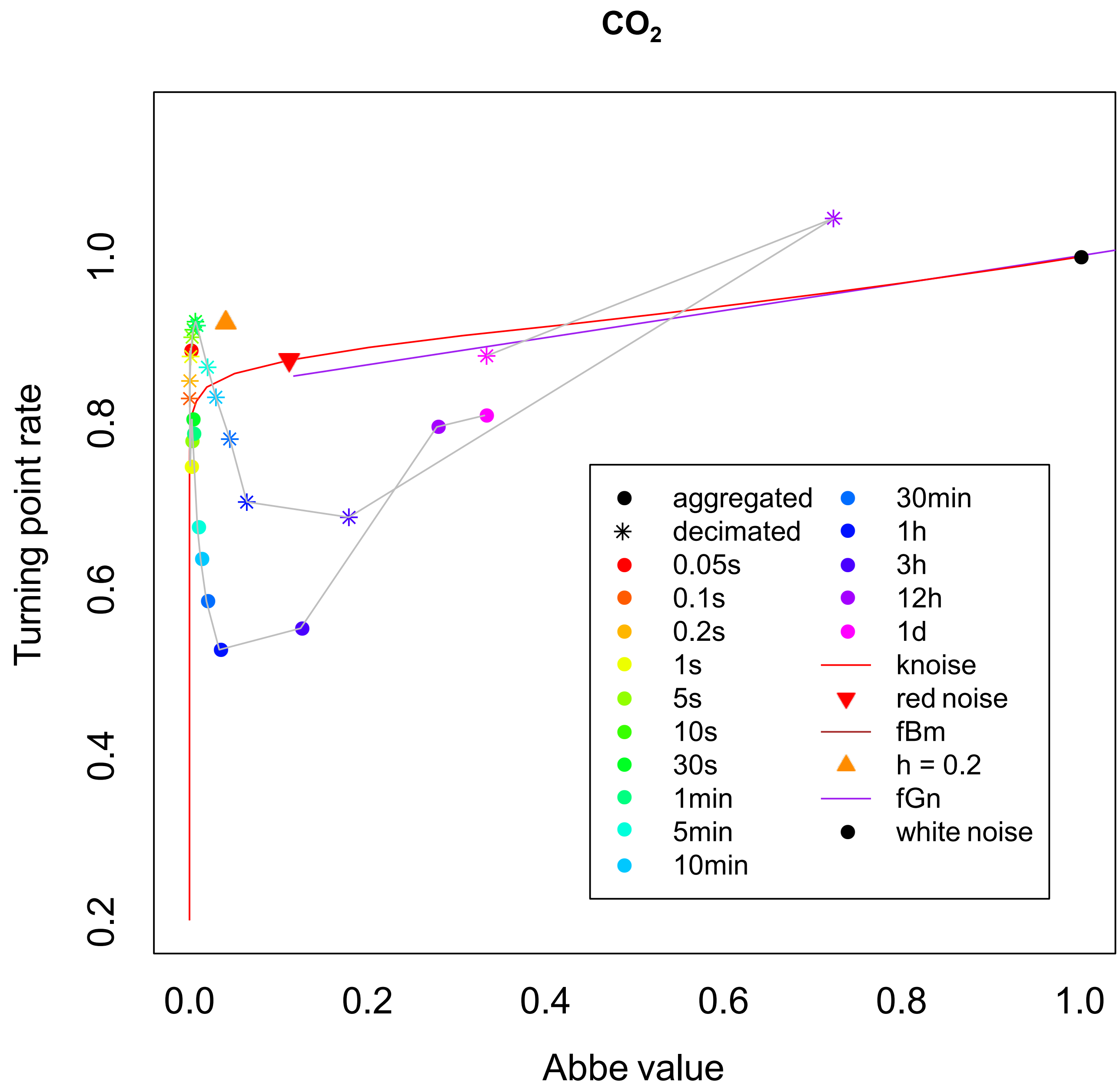
M. Tarnopolski, Physica A 416 (2016), 662-673


Tarnopolski diagram for noise processes
Turning point rate
Abbe value
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
knoise
red noise
pink noise
fBm
h = 0.2
h = 0.4
h = 0.6
fGn

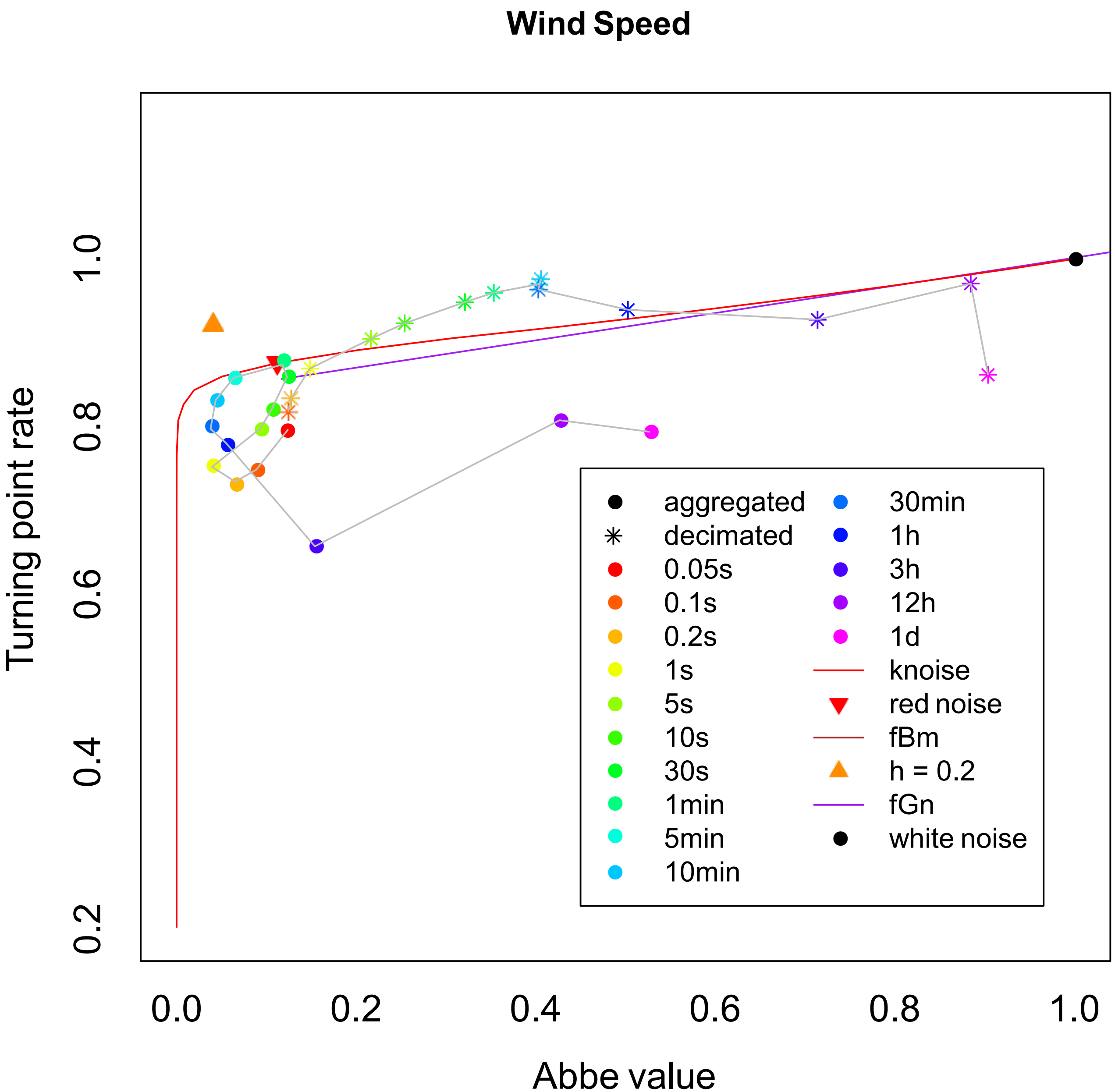


Tarnopolski diagram for CO_2
Turning point rate
Abbe value
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
knoise
red noise
fBm
h = 0.2
fGn
CO2 24h
CO2 daytime
CO2 nighttime

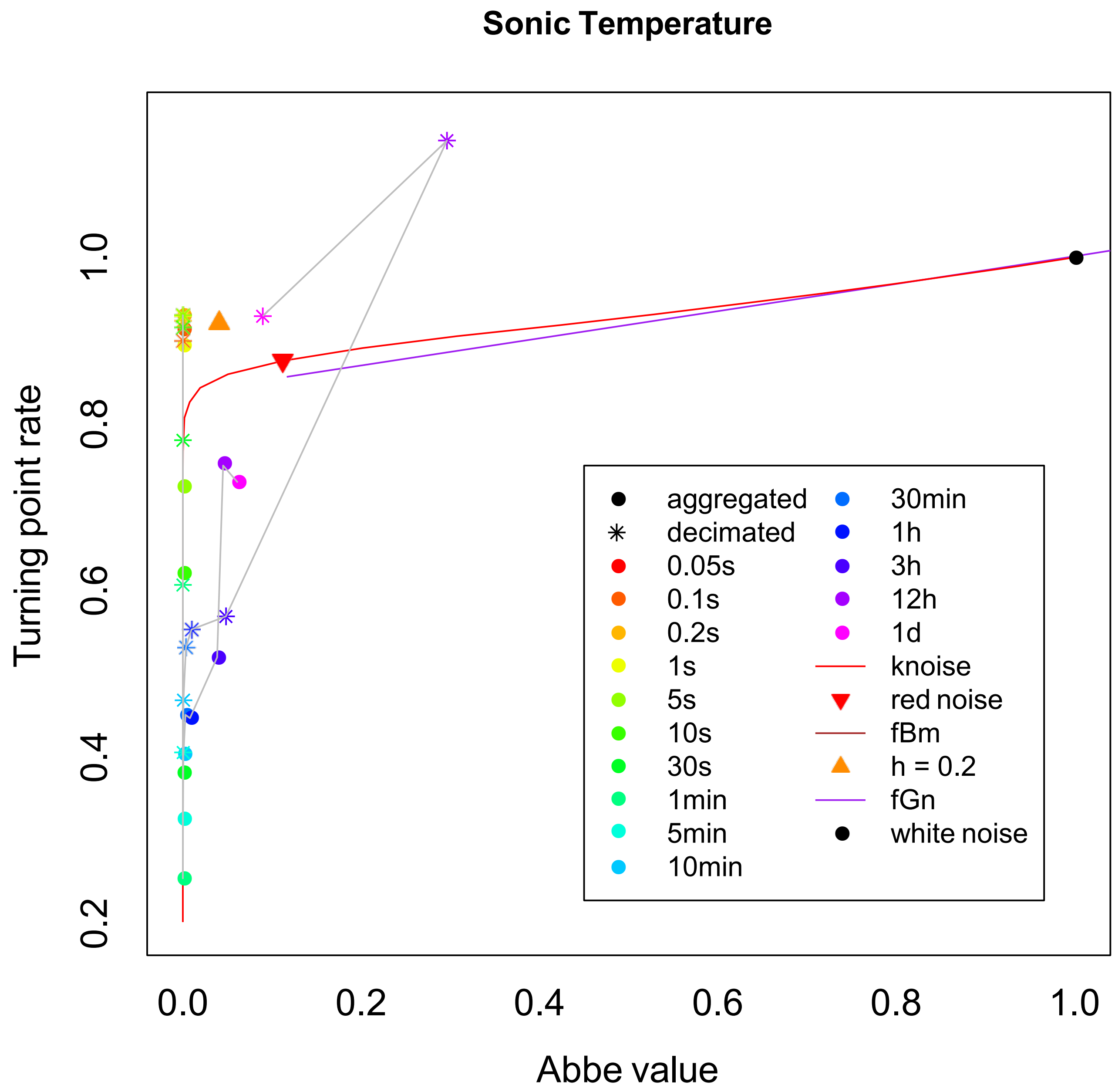

Tarnopolski diagram for wind speed
Turning point rate
Abbe value
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
knoise
red noise
fBm
h = 0.2
fGn
Wind 24h
Wind daytime
Wind nighttime



CO2
Turning point rate
Abbe value
0.2
0.4
0.6
0.8
1.0
0.0
aggregated
decimated
0.05s
0.1s
0.2s
1s
5s
10s
30s
1min
5min
10min
30min
1h
3h
12h
1d
knoise
red noise
fBm
h = 0.2
fGn
white noise



Wind Speed
Turning point rate
Abbe value
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
aggregated
decimated
0.05s
0.1s
0.2s
1s
5s
10s
30s
1min
5min
10min
30min
1h
3h
12h
1d
knoise
red noise
fBm
h = 0.2
fGn
white noise



Sonic Temperature
Turning point rate
Abbe value
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
aggregated
decimated
0.05s
0.1s
0.2s
1s
5s
10s
30s
1min
5min
10min
30min
1h
3h
12h
1d
knoise
red noise
fBm
h = 0.2
fGn
white noise

The Bandt-Pompe order statistics

Given a time series x_i of length *N*

Choose an embedding dimension (aka word length) d

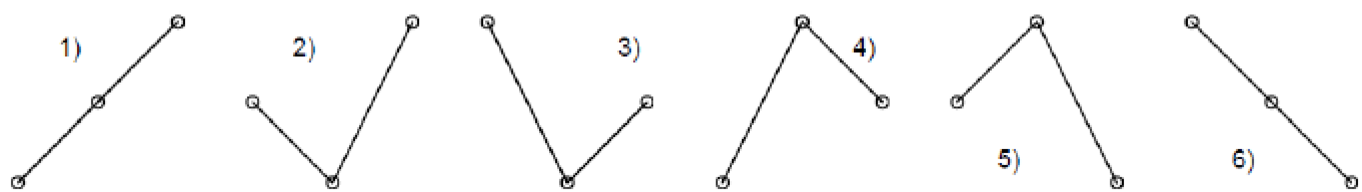
required: $N > 5d!$

For each set of d consecutive values, determine their *relative ranks*, e.g.

$\{0.75,0.32,0.66,0.84\} \to \{3,1,2,4\}$

Each group of ranks forms an order pattern; there are d! different patterns

d= 3:



The frequencies of patterns forms the *ordinal pattern distribution* $\{p_i\}_{j=1,\ldots,d!}$

q-entropy and q-complexity

- q-entropy (Tsallis & Co.)
- $log_q x = \int_1^x t^{-q} dt$
- $H_q(P) = \frac{1}{d!}\sum_{j=1}^{d!} p_j log_q \frac{1}{p_j}$
- q-complexity (*Ribeiro et al. 2017*)
- $C_q(P,R)=\frac{D_q(P,R)H_q(P)}{D_q^*}$
- $D_q(P,R) = \frac{1}{2}\left(K_q\left(P|\frac{P+R}{2}\right) + K_q\left(R|\frac{P+R}{2}\right)\right)$

K_q is the q-version of the Kullbäck-Leibler distance:

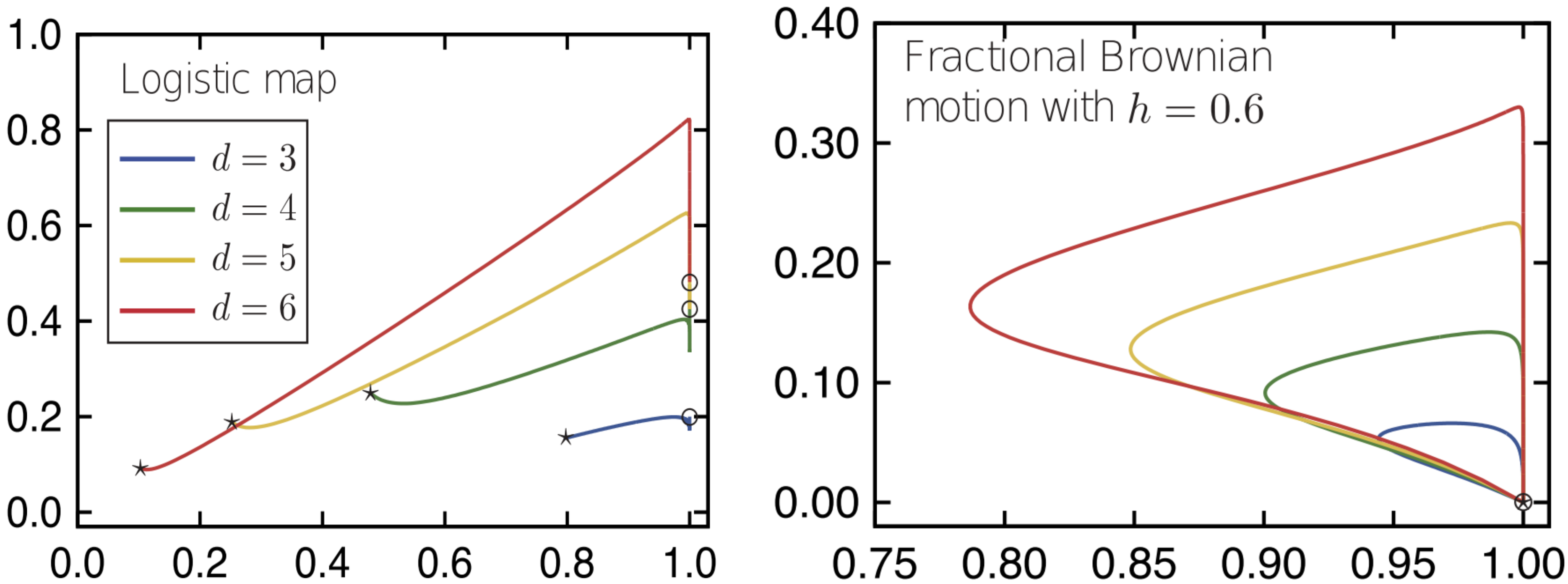
$$K_q(P|R) = -\sum p_i log_q \frac{r_i}{p_i}$$

The typical choice for R is the uniform distribution

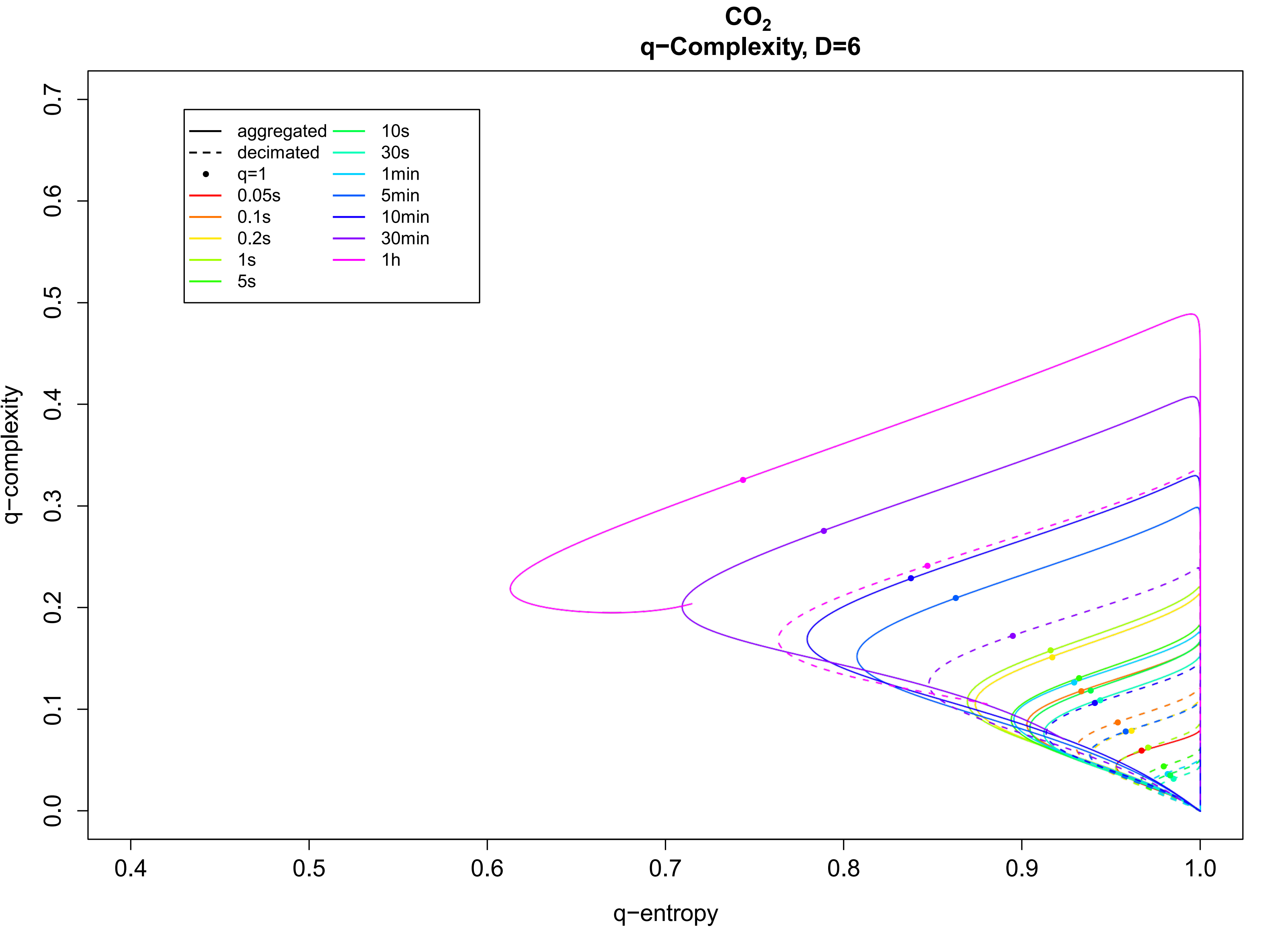
$$U = \{1/d!\}_{j=1,\ldots,d!}$$

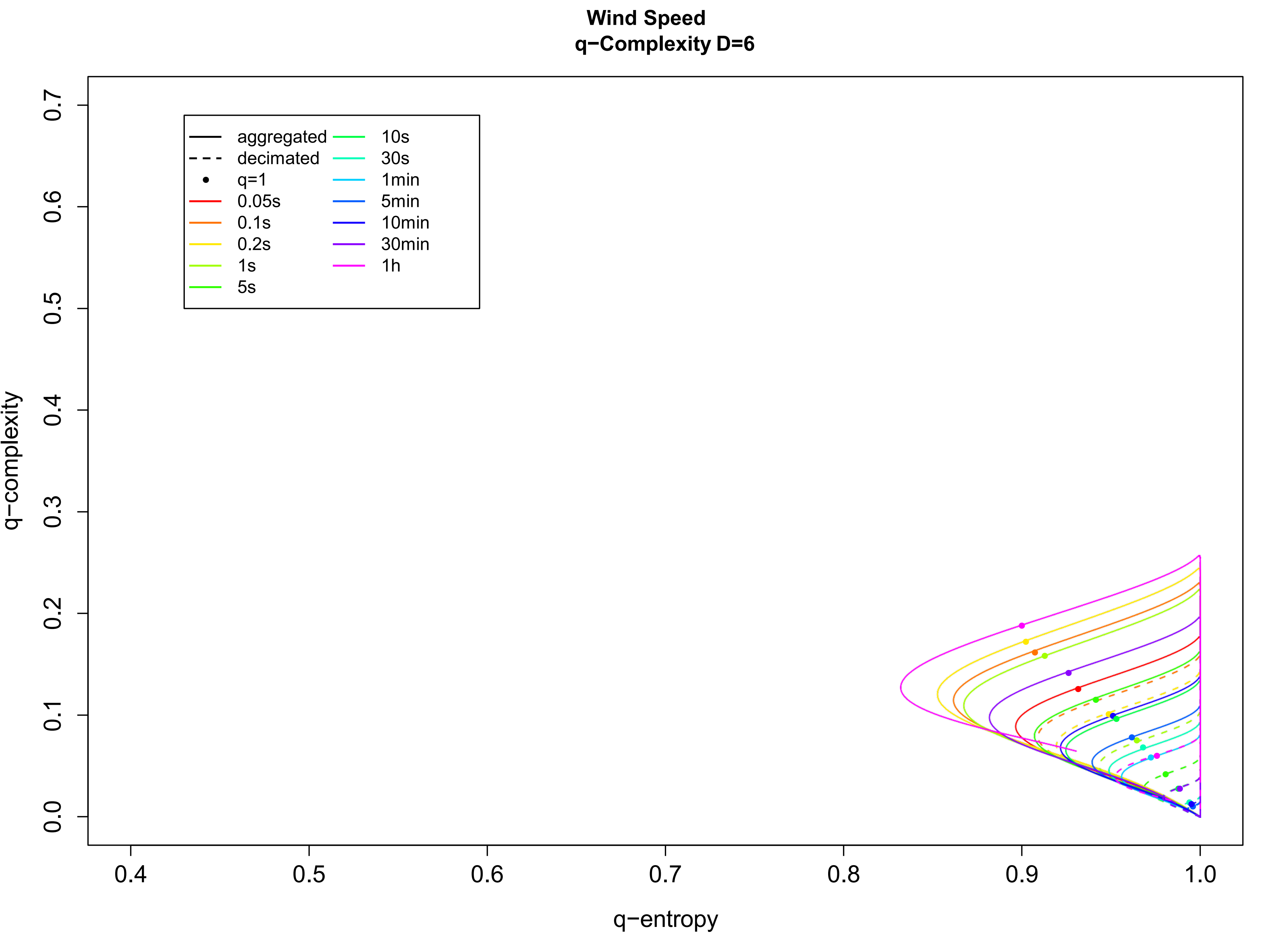
$$D_q^* = \max_P D_q(P,U)$$

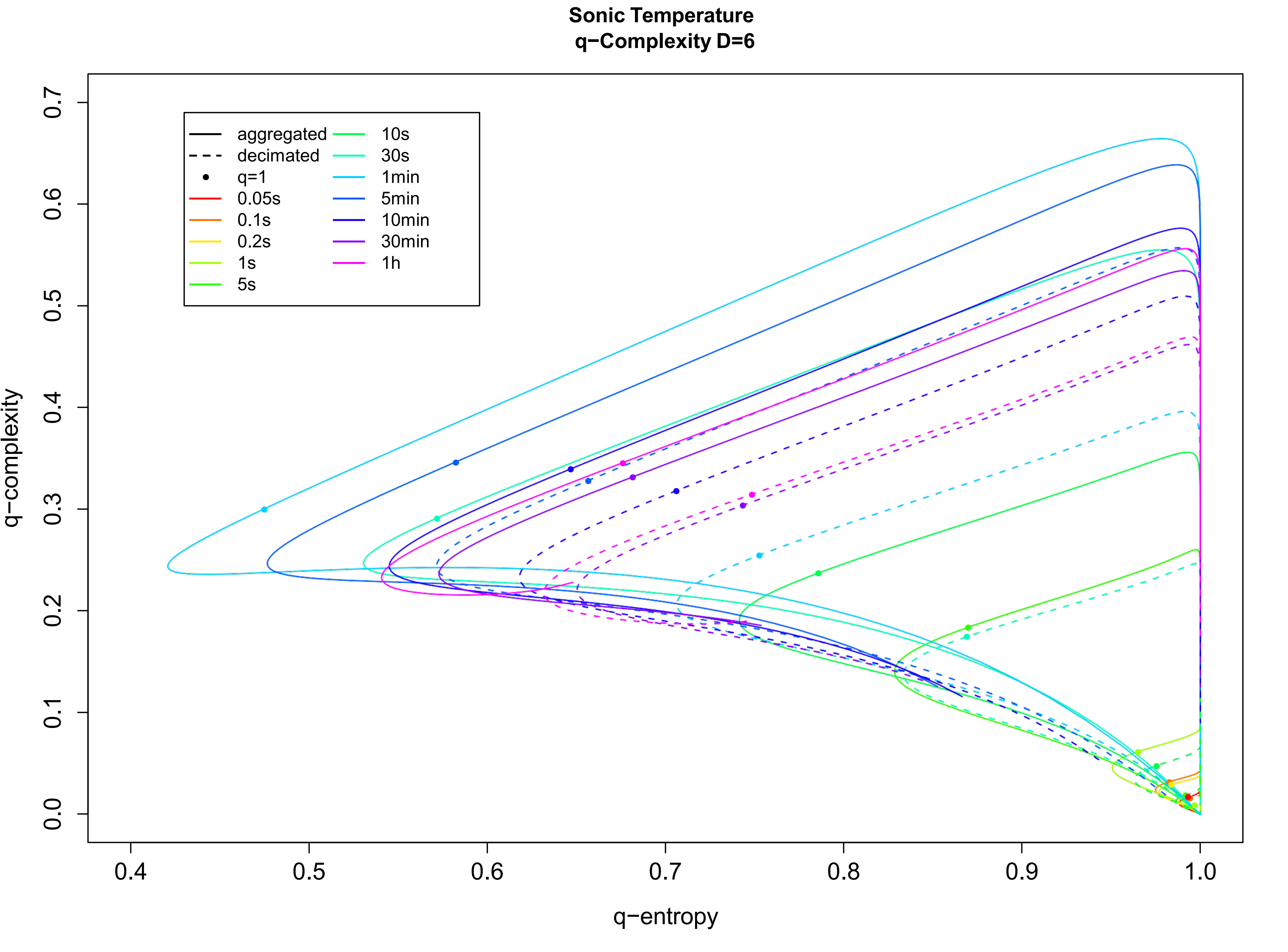
qH-qC plane: deterministic processes form open curves, Stochastic processes form loops

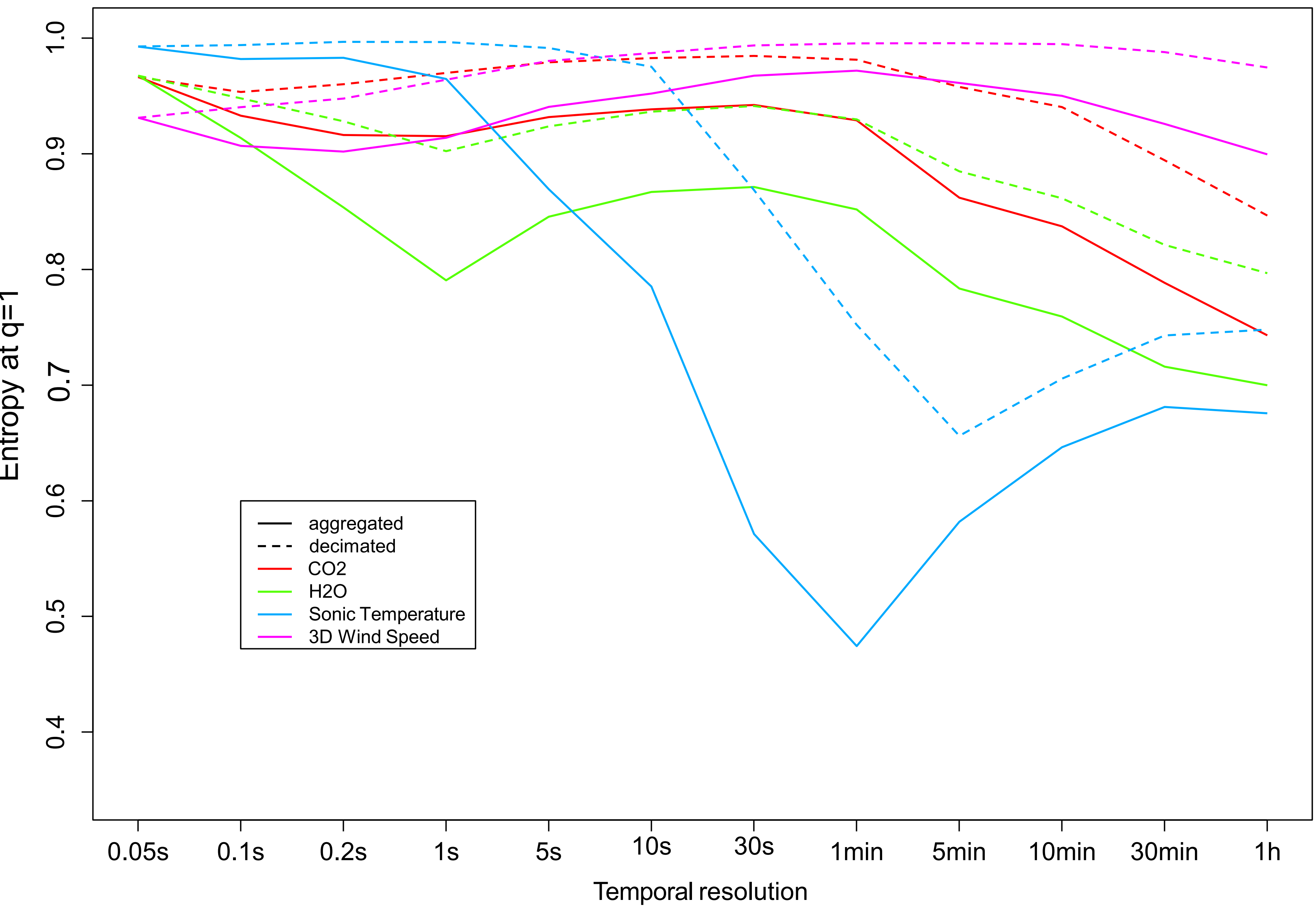


Ribeiro et al. (2017)


CO_2
q−Complexity, D=6
aggregated
decimated
q=1
0.05s
0.1s
0.2s
1s
5s
10s
30s
1min
5min
10min
30min
1h
q−complexity
q−entropy
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.4
0.5
0.6
0.7
0.8
0.9
1.0


Wind Speed
q−Complexity D=6
aggregated
decimated
q=1
0.05s
0.1s
0.2s
1s
5s
10s
30s
1min
5min
10min
30min
1h
q−complexity
q−entropy
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.4
0.5
0.6
0.7
0.8
0.9
1.0


Sonic Temperature
q−Complexity D=6
aggregated
decimated
q=1
0.05s
0.1s
0.2s
1s
5s
10s
30s
1min
5min
10min
30min
1h
q−complexity
q−entropy
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.4
0.5
0.6
0.7
0.8
0.9
1.0



Entropy at q=1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.05s
0.1s
0.2s
1s
5s
10s
30s
1min
5min
10min
30min
1h
Temporal resolution
aggregated
decimated
CO2
H2O
Sonic Temperature
3D Wind Speed

Conclusions and Outlook

NIBIO
NORWEGIAN INSTITUTE OF BIOECONOMY RESEARCH

- Eddy Covariance is generating very long time series of many variables
- Long-range correlations and persistence are the rule even at high temporal resolution
- EC time series are NOT simply fBm or k noise
- Temporal scaling over several orders of magnitude can be investigated
- Optimal resolution can be determined using Tarnopolski and q complexity analysis:
- $CO_2 \approx 1$ s , Wind ≈ 0.2 s , Temperature ≈ 1 min
- Environmental science is a playground for theoretical physicists

Thank you for your attention!

NIBIO
NORWEGIAN INSTITUTE OF BIOECONOMY RESEARCH

View from top of (ICOS) tower at Hurdal