Scaling, multi-scaling, and non-stationarity

Steve UHLIG

Scaling

- Scaling: no timescales plays a significant role compared to other timescales
- If the increments of a process X(t) are replaced by the wavelet coefficients *d*(j,k), scaling can be studied through the statistic $E\left|d(j,k)\right|^q$
- q is the moment of the statistic:
  - small q values capture smoothness
  - large q values capture irregularity
- The power-law can be linear or non-linear:
  - self-similarity : $E\left|d(j,k)\right|^q \propto \exp(qH\ln(2^j))$
  - multi-scaling : $E\left|d(j,k)\right|^q \propto \exp(H(q)\ln(2^j))$

Wavelets reminder

- A wavelet $\psi_{j,k}$ is a bandpass oscillating function where j is the timescale and k is time
- $\psi_{j,k}(t)$ form an orthonormal basis of L2
- The internal product $\left\langle X(t), \psi_{j,k}(t) \right\rangle$ matches irregularities at scale j and time k
- Useful properties:
  - Built-in scaling: $\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k)$
  - N vanishing moments: $\int \psi_0(t)dt = 0, k = 0, \ldots, N-1$
  - Almost decorrelation: $d_{j,k}$ are Gaussian distributed

Scaling vs. physical models

- Scaling is a statistical property of the signal
- Scaling models do not need to have any relationship with the physics (generative process) of the signal
- Scaling is a formalism aimed at describing the statistical properties of a signal

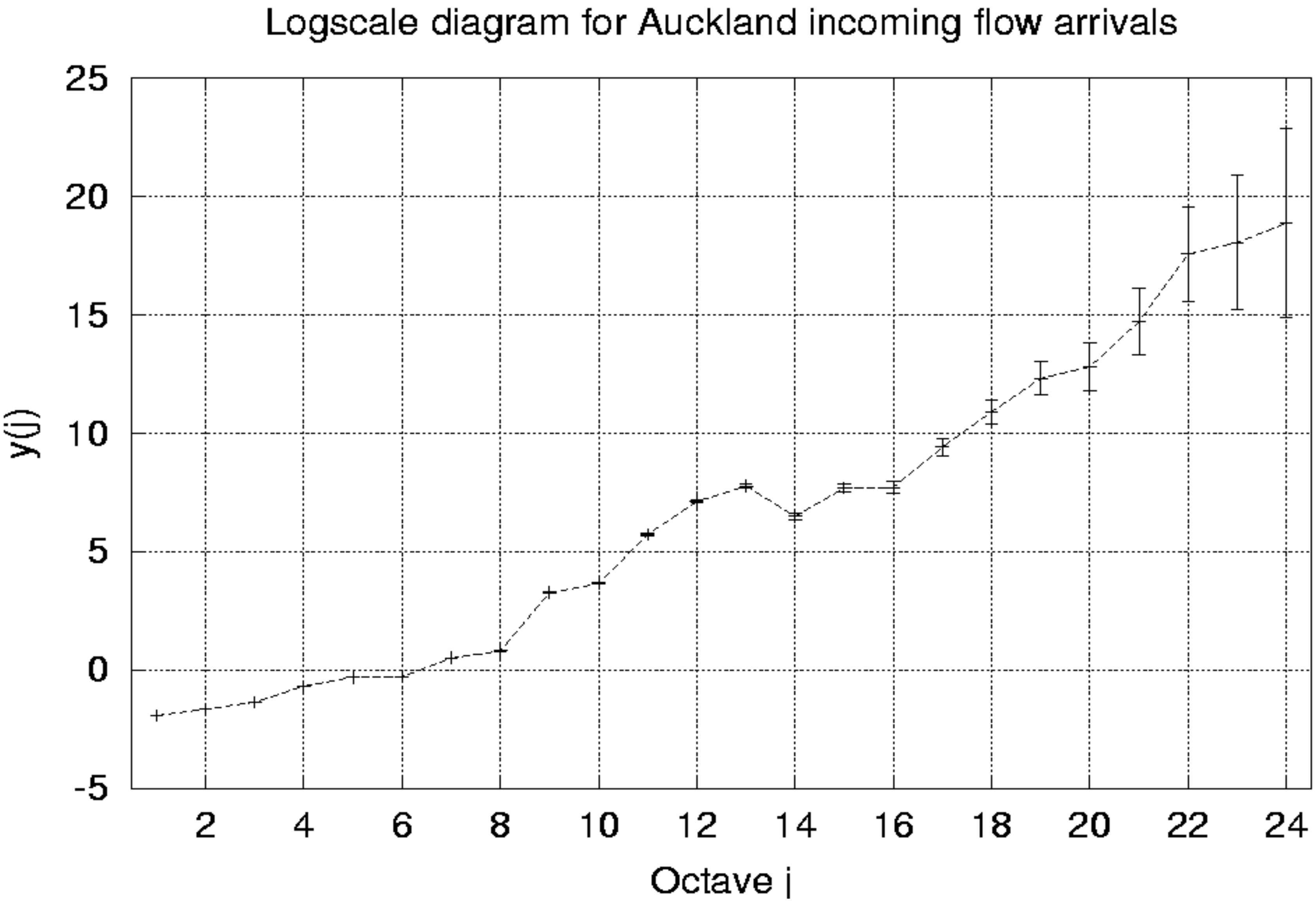
2nd order scaling

- 2nd order scaling can be identified with the *logscale diagram* (LD) [VA99]:

  - plot $\log_2(\mu_j)$ as a function of the timescale *j where* $\mu_j = \sum_k |d_{j,k}|^2$

- Scaling detection requires alignment of $\log_2(\mu_j)$, including the confidence intervals

Example

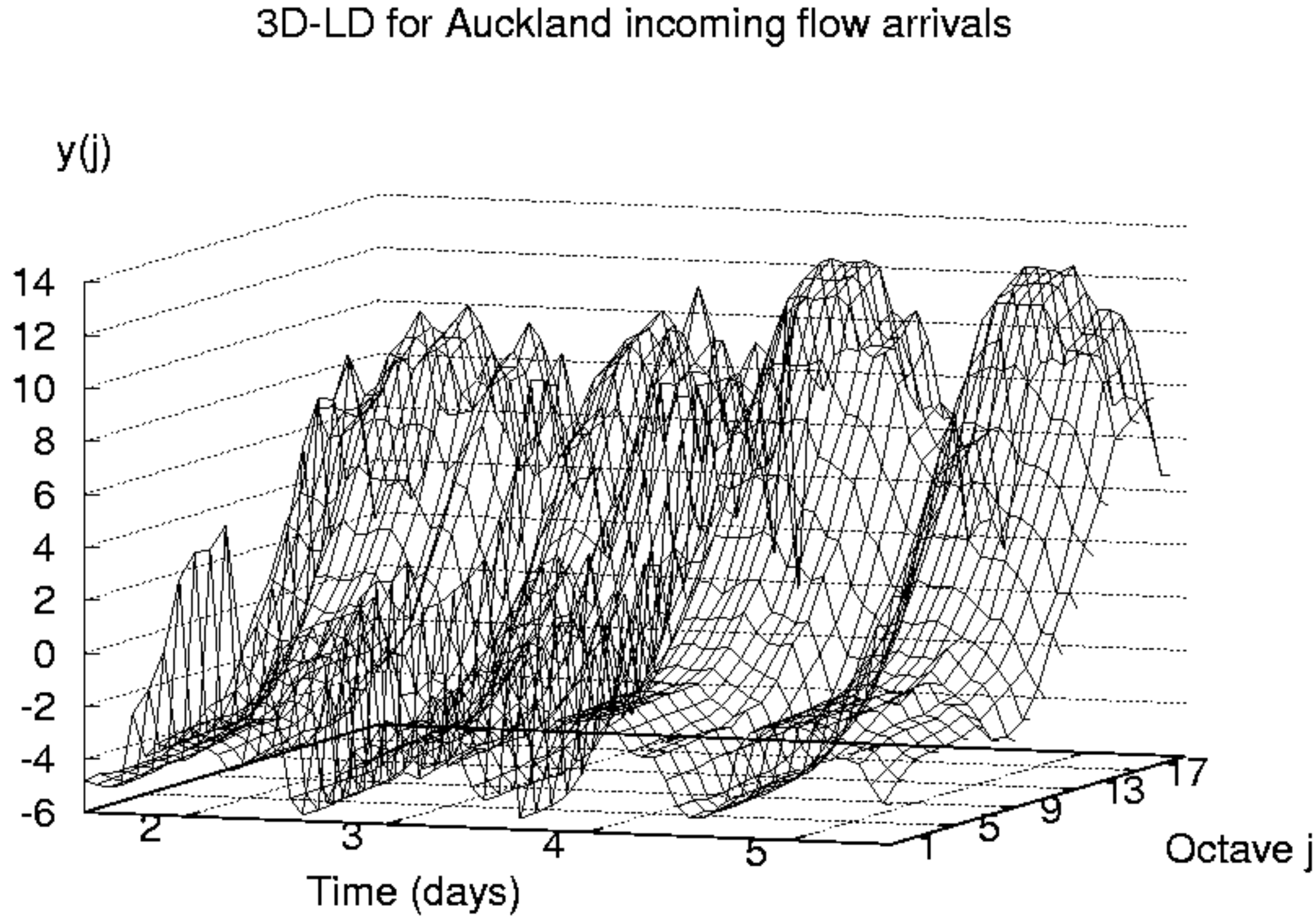


Logscale diagram for Auckland incoming flow arrivals
25
20
15
10
5
0
-5
y(j)
2 4 6 8 10 12 14 16 18 20 22 24
Octave j

2^{nd} order scaling and non-stationarity

- What if strict scaling is not identified?
- No scaling or time-dependent scaling?
- 3D-LD:
  - break the process into constant-size time intervals
  - compute the LD over each interval
  - plot in 3D the evolution over time of the LD

Non-stationary scaling



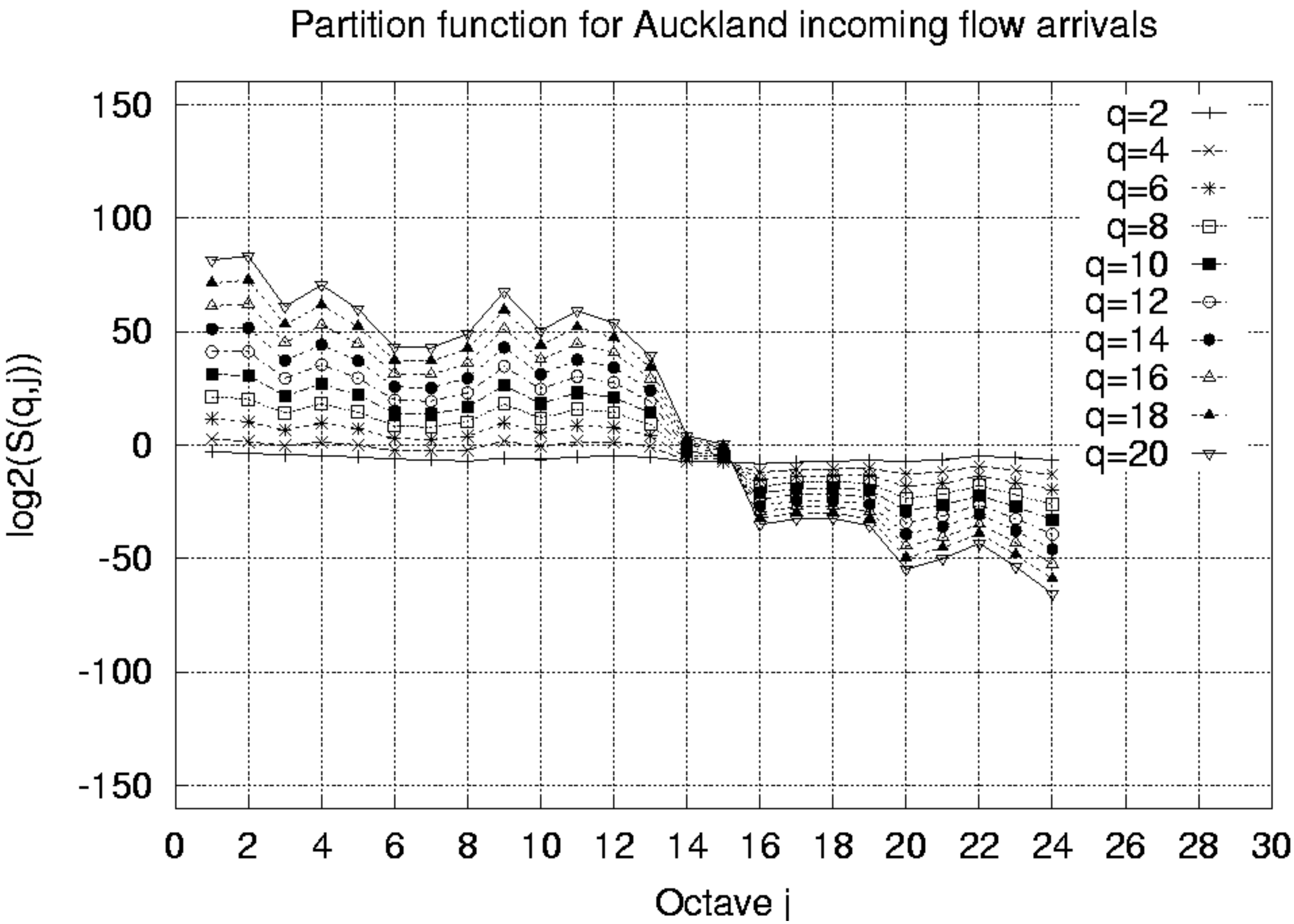
Scaling components

- 2 different scaling regimes have been identified in Internet traffic:
  - Long-range dependence at time scales larger than the average flow size [WPRT02]
  - Multi-scaling at small time scales [FGW98]
- At large time scales, the cause is heavy-tailed flow sizes coupled with ON/OFF durations [TWS97]
- At small time-scales, TCP is conjectured to be responsible [FGW98]

High-order scaling

- Distinguishing between different types of scaling is possible by checking the linearity of the moments *q* as a function of time scales *j*
- Partition function S(q,j) = $\sum_k \left| 2^{-j/2} d(j,k) \right|^q$
- M(j) = $\left\langle \frac{\log(S(q+1,j)) - \log(S(q,j))}{\log(S(2,j)) - \log(S(1,j))} \right\rangle_q$
  - No scaling: M(j) < 1
  - self-similarity: M(j) ~1
  - multi-scaling: M(j) > 1

Example

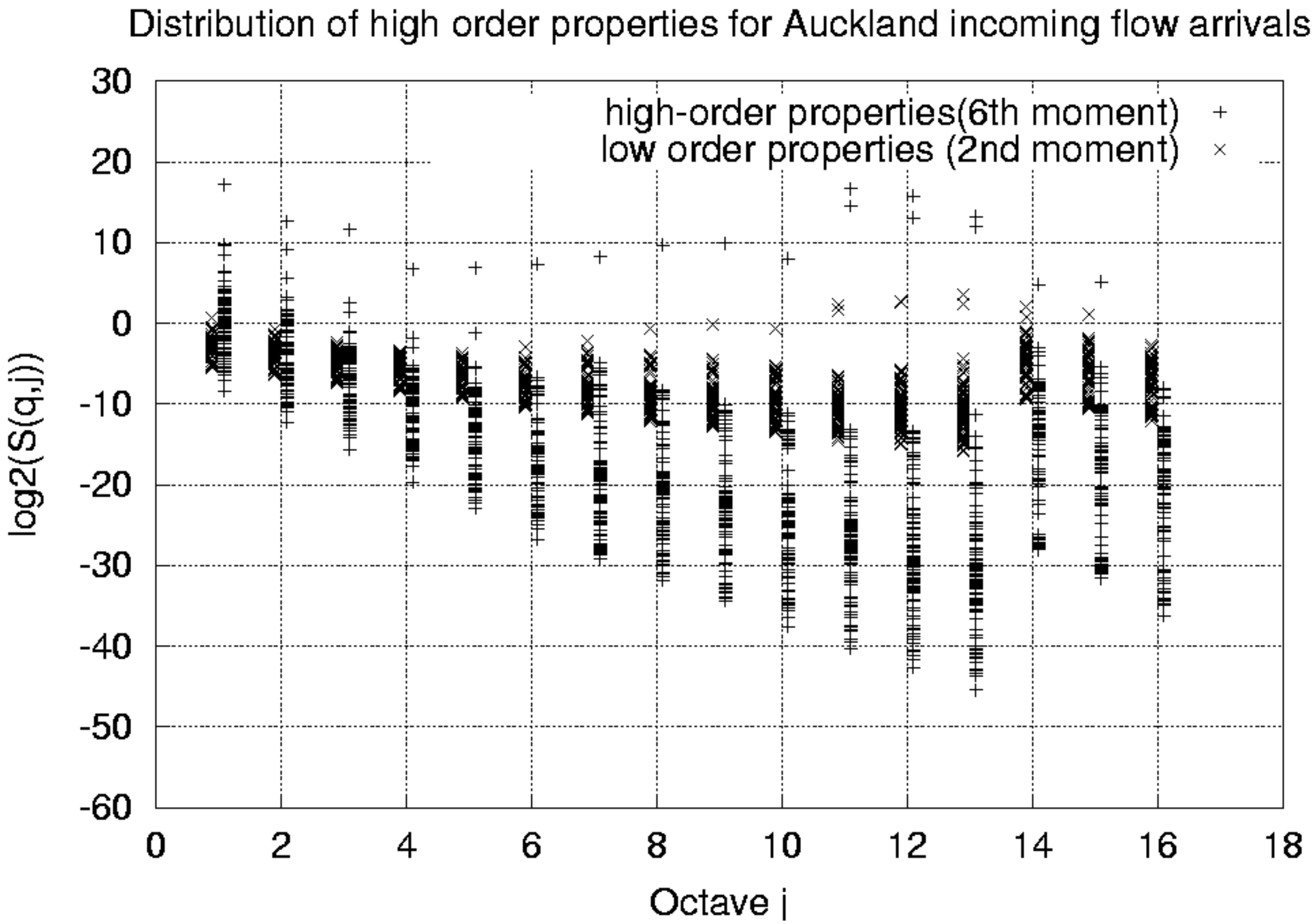


Partition function for Auckland incoming flow arrivals
150
100
50
0
-50
-100
-150
log2(S(q,j))
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30
Octave j
q=2
q=4
q=6
q=8
q=10
q=12
q=14
q=16
q=18
q=20

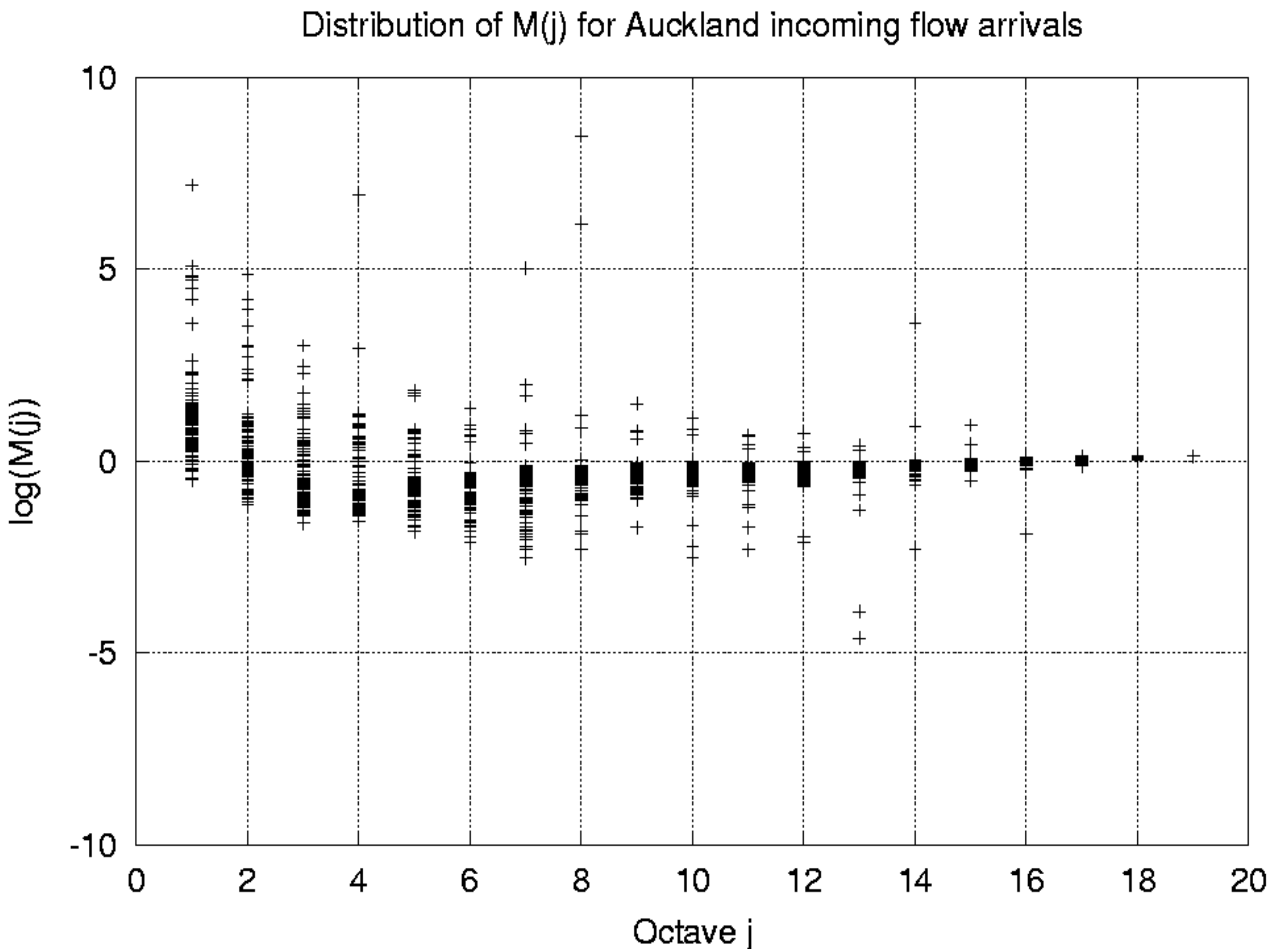
High-order scaling and non-stationarity

- High-order scaling can also suffer from non-stationarity
- Same principle as for 3D-LD:
  - break the process into constant-size time intervals
  - compute the MD over each interval
  - plot the MD over each time interval

Bias from non-stationarity



M(j) over time



Distribution of M(j) for Auckland incoming flow arrivals
10
5
0
-5
-10
log(M(j))
0
2
4
6
8
10
12
14
16
18
20
Octave j

Conclusions

- Non-stationarity is present in real data
- Always check your data!
- Scaling easily biased by non-stationarity, despite wavelet properties
- Tools exist to check for stationarity

Selected references

- [FGW98] A. Feldmann, W. Willinger and A. Gilbert. *Data networks as cascades: Investigating the Multifractal Nature of Internet WAN Traffic.* Proc. of ACM SIGCOMM, 1998.
- [TWS97] M. Taqqu, W. Willinger and R. Sherman. *Proof of a Fundamental Result in Self-Similar Traffic Modeling.* Comp. Comm. Rev., 1997.
- [VA99] D. Veitch and P. Abry. *A wavelet based joint estimator for the parameters of LRD.* IEEE Transactions on Information Theory, 45(3), April 1999.
- [AFTV00] Abry, Flandrin, Taqqu, Veitch. *Wavelets for the analysis, estimation and synthesis of scaling data.* Book chapter of "Self Similar Network Traffic Analysis and Performance Evaluation", Wiley, 2000.
- [WPRT02] W. Willinger, V.Paxson, R. Riedi and M. Taqqu. *Long-range dependence and data network traffic.* Book chapter of "Theory and applications of long-range dependence", Birkhäuser, Boston, 2002.