Conservative Cascades: an Invariant of Internet Traffic

Steve **UHLIG**

E-mail: suh@info.ucl.ac.be
URL: http://www.info.ucl.ac.be/~suh/

Computer Science and Engineering Dept.
Université catholique de Louvain, Belgium

Schedule

1) Scaling theory ⇔ Scaling theory

2) Cascades and TCP traffic ⇔ Scaling practice

3) TCP traffic analysis ⇔ Traffic analysis

4) Conclusions

Scaling

~ absence of characteristic (time)scale

~ dependence among (time)scales

Many different types of scaling flavours :

- self-similarity
- long-range dependence
- multiscaling
- multifractality
- infinitely divisible cascades

Terminology

Let X(t) be a regularly sampled process :

- **d(j,k)** : process increment at timescale **j** and time **k**
- **q** : moment of the process
- **j** : timescale, **j** = 1,...,n.
- **E|X(t)|**^{q} : expectation of moment **q** of process **X** at time **t** (over all possible realizations of the process).

Assumption :

- **d(j,k) is** a homogeneous (stationary) zero-mean process.

Self-similarity

$$E|d(j,.)|^q = C_q (2^j)^{qH} \propto \exp(q\,H\,\ln(2^j))$$

- linear scaling among timescales and linear scaling among moments both driven by a single parameter H

⇒ called mono-scaling since H is constant

Multiscaling

$$E|d(j,.)|^q = C_q (2^j)^{H(q)} \propto \exp(H(q)\ \ln(2^j))$$

- linear scaling among timescales j, but driven by a function H(.) of the moment q

⇒ mono-scaling becomes multi-scaling since scaling depends on the moment q

Infinitely divisible cascade

$$E|d(j,.)|^q = C_q (2^j)^{H(q)} \propto \exp(H(q)\, n(2^j))$$

where

- H(q) represents 1 step of the cascade
- $n(2^j)$ represents cascade depth at timescale j

⇒ scaling does not depend linearly on the moment q nor the timescale j anymore

The missing link

$$E|d(j,.)|^q \propto \exp(f(q)\, g(2^j))$$

- **separability** of the moments (**q**) and the timescales (**j**)
- higher-order moments emphasize on larger irregularities
- larger timescales focus on smoothed versions of the process (zooming out)

⇒ **framework to study irregularities and timescales independently**

Probabilistic interlude

- <u>Cascade :</u> multiplicative process that breaks a process into smaller and smaller fragments according to some (deterministic or random) rule.

- <u>Link with infinitely divisible distributions (Feller vol. 2) :</u>

F is infinitely divisible if for every n there exists a distribution F_n such that $F = F_n^{n*}$ where "*" denotes the convolution operator.

or equivalently

F is infinitely divisible iff for each n it can be represented as the distribution of the sum $S_n = X_{1,n} + ... + X_{n,n}$ of n independent random variables with a common distribution F_n.

Conservative cascades

2 objectives :

- preservation of total mass of the process
- randomness in the splitting of the mass of the process

Example :

Step 1 :

Step 2 :

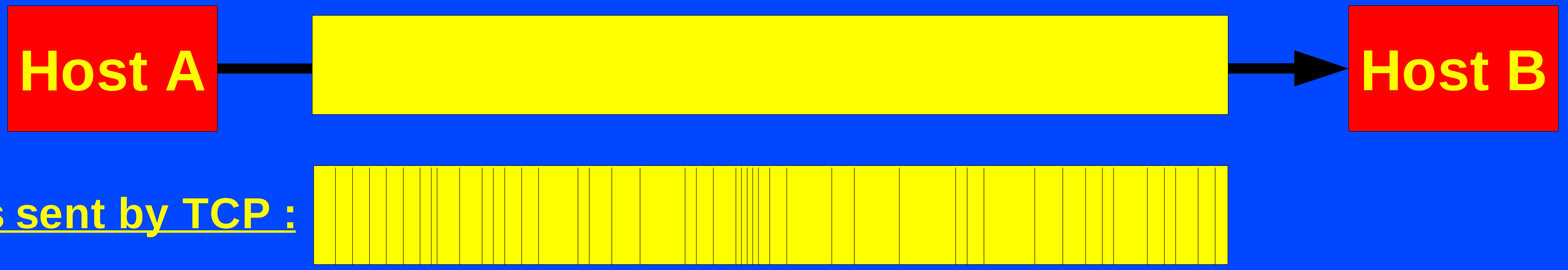
Step 3 :

Conservative cascades and TCP traffic

TCP :

- breaks the total mass of the data to be sent into segments

- application sending data, TCP state machine and random network conditions drive how data segments are broken and when they are sent over time



Segments sent by TCP :

Cascade parameters and TCP

H(q) : - describes cascade generator (1 step of the breaking of the data)

- represents how TCP distributes the mass of the TCP segments into large and small irregularities

n(2^j) : how many times the generator has been applied at timescale **j** (how many times the convolution operator has been applied)

⇒ **H()** and **n()** “ *summarize*” the numerical behavior of the cascade

Wavelet analysis

- process increments **d(j,k)** replaced by wavelet coefficients δ***(j,k)***

- Wavelet-based partition function :

$$\mathbf{S(q,j)} = \sum_k |2^{-j/2}\delta(j,k)|^q \qquad (\propto E|d(j,k)|^q)$$

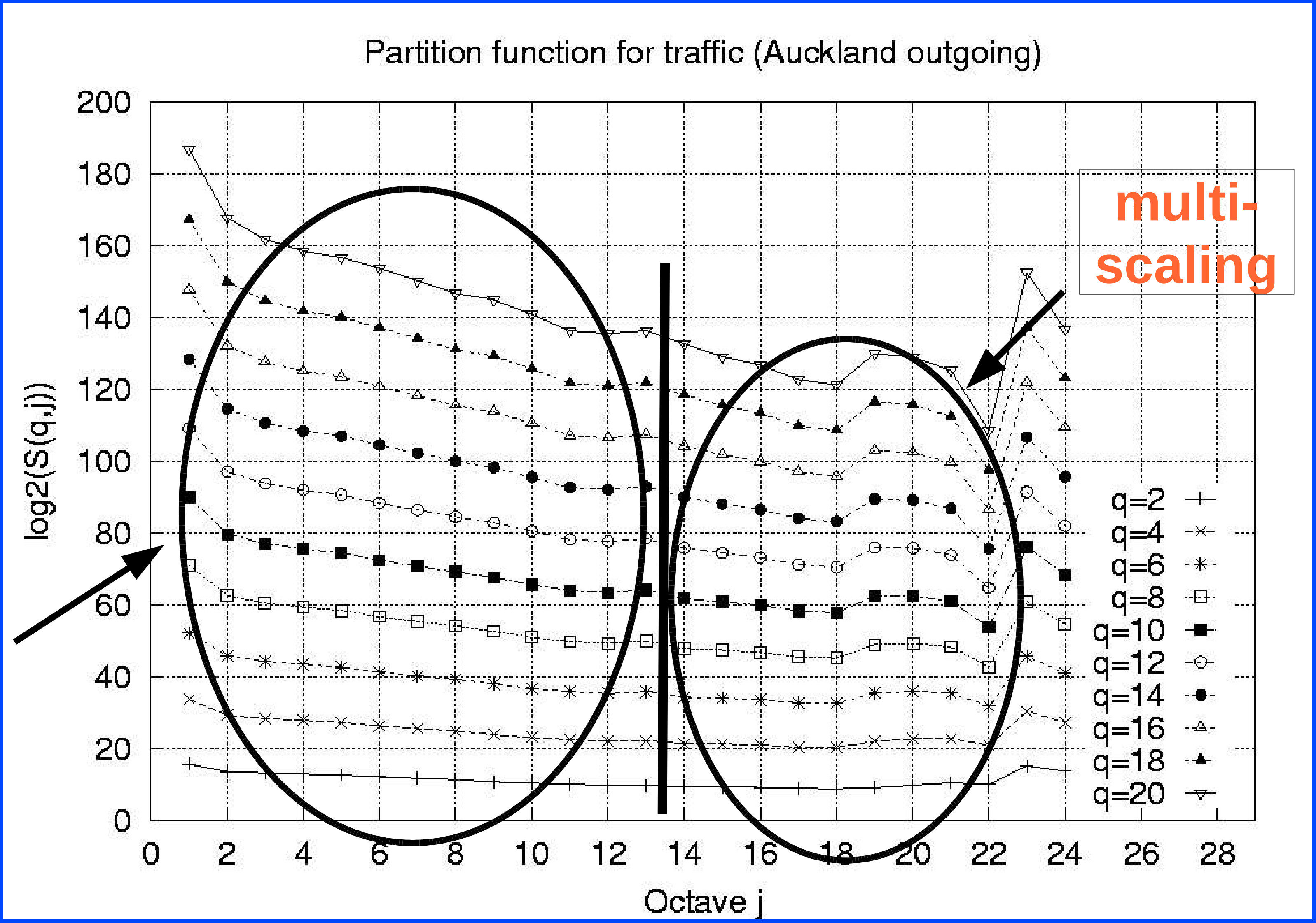
- Estimation of **H(q)** and **n(2^j)** via regression :

$$\ln(S(q,j)) \cong -H(q)\, n(2^j) + K_q$$

Internet traffic analysis

- **15** hours of IP packets from Auckland
- both traffic sent to the Internet and received from the Internet
- **1,629,069** incoming TCP flows, **1,613,976** outgoing
- more than **13** GBytes of incoming traffic, **10** GBytes outgoing
- 1 μs precision, 1 ms time granularity used.

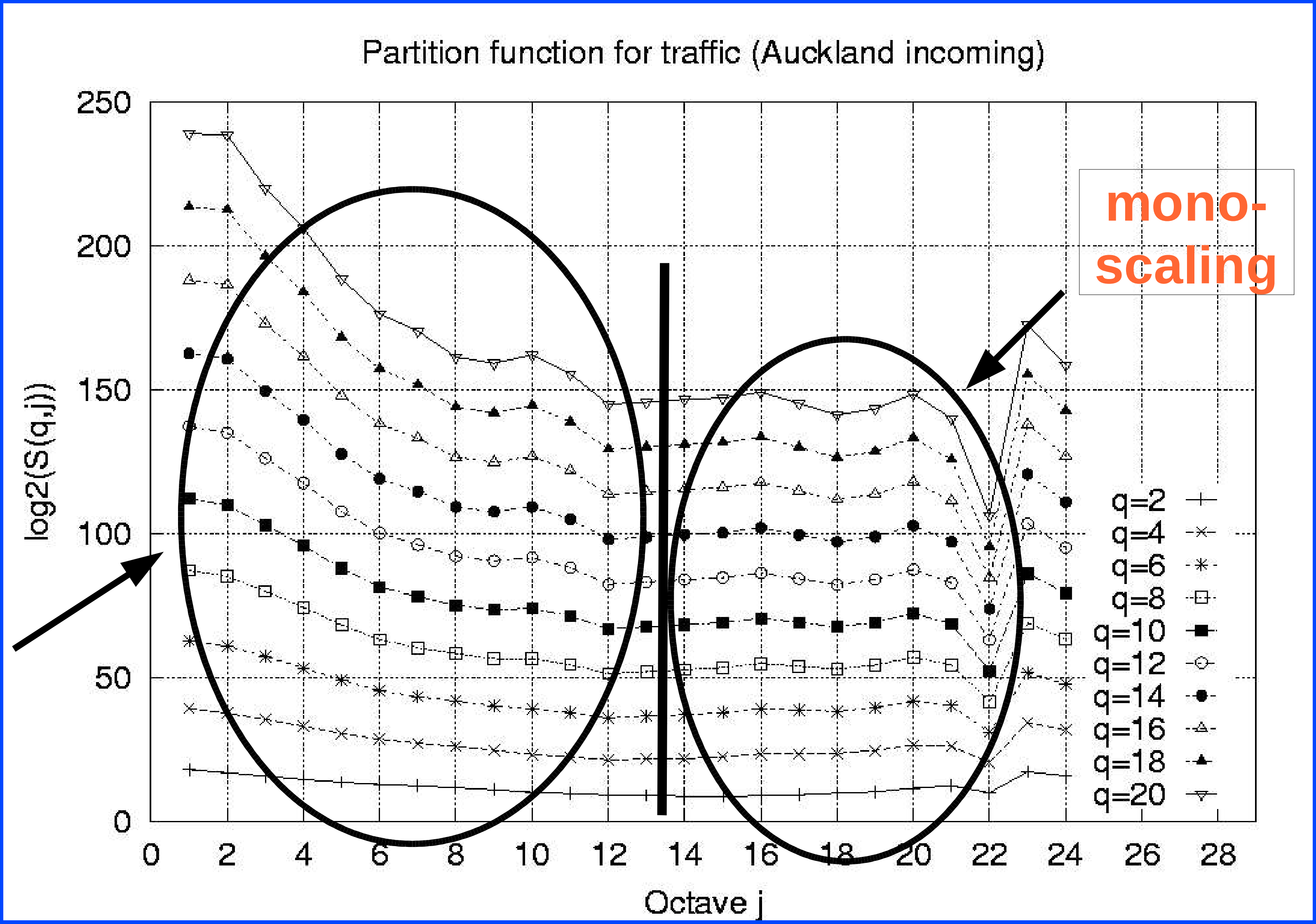
Multiscaling analysis (out)



multi-scaling

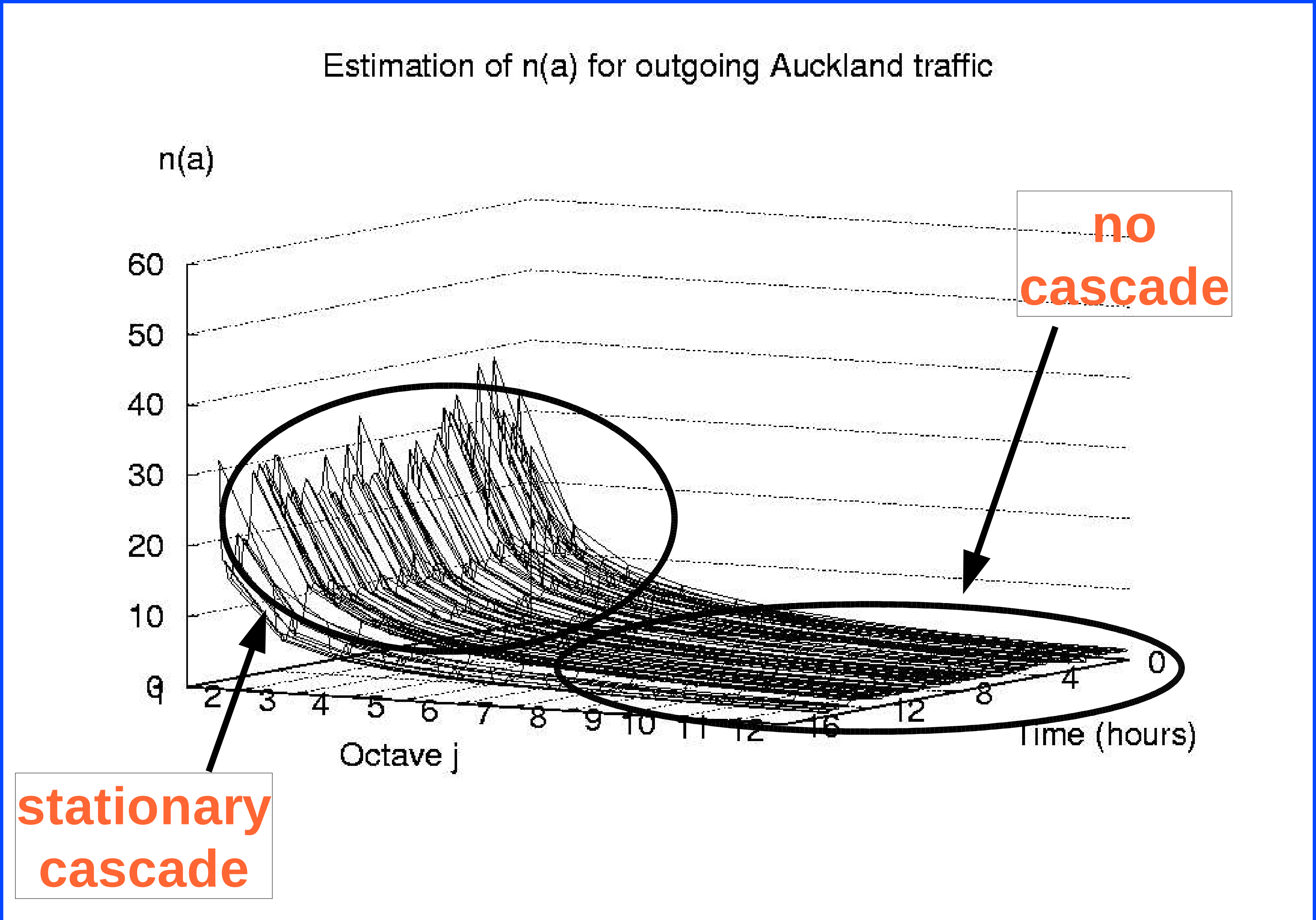
multi-scaling

Multiscaling analysis (in)

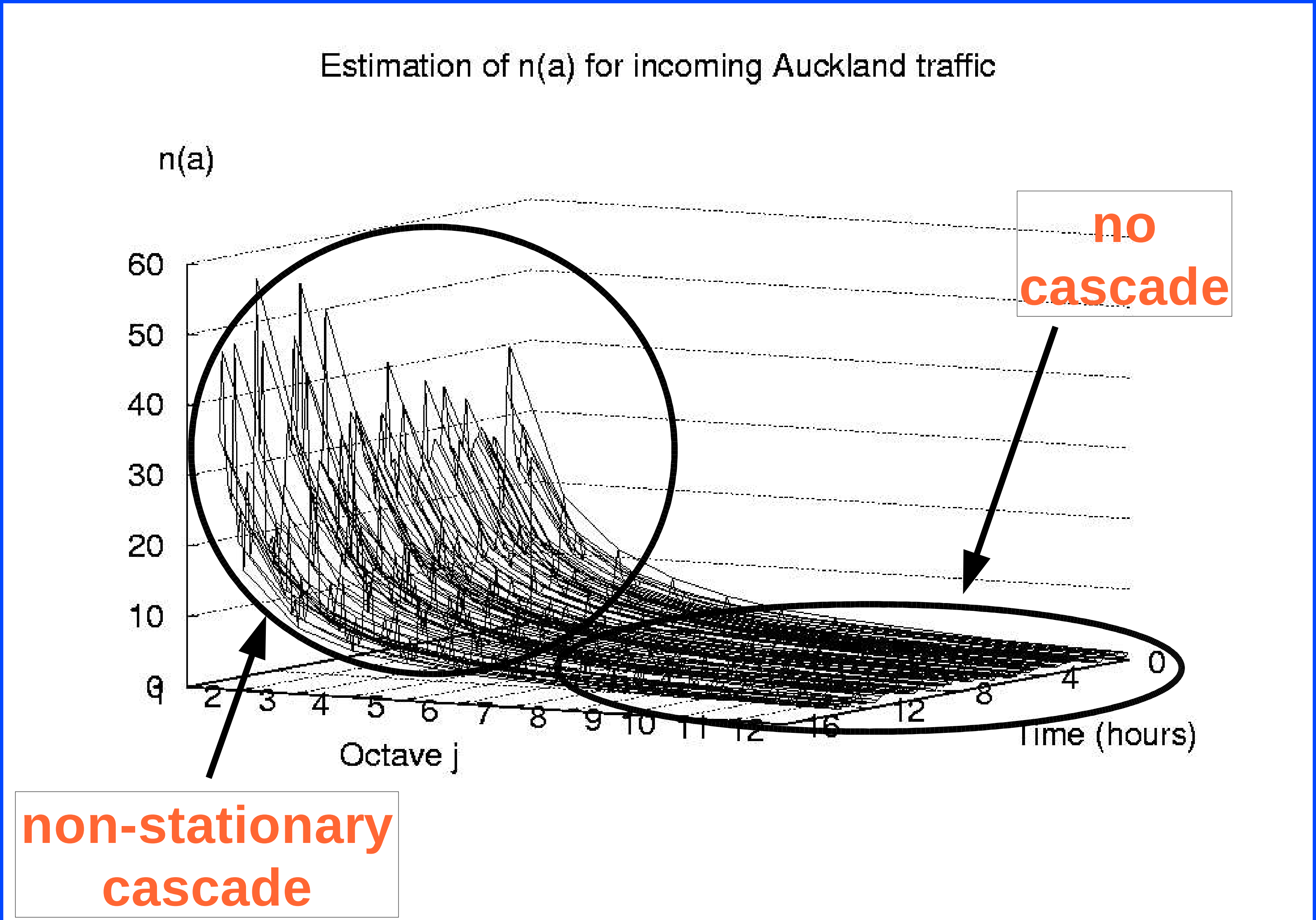


multi-scaling

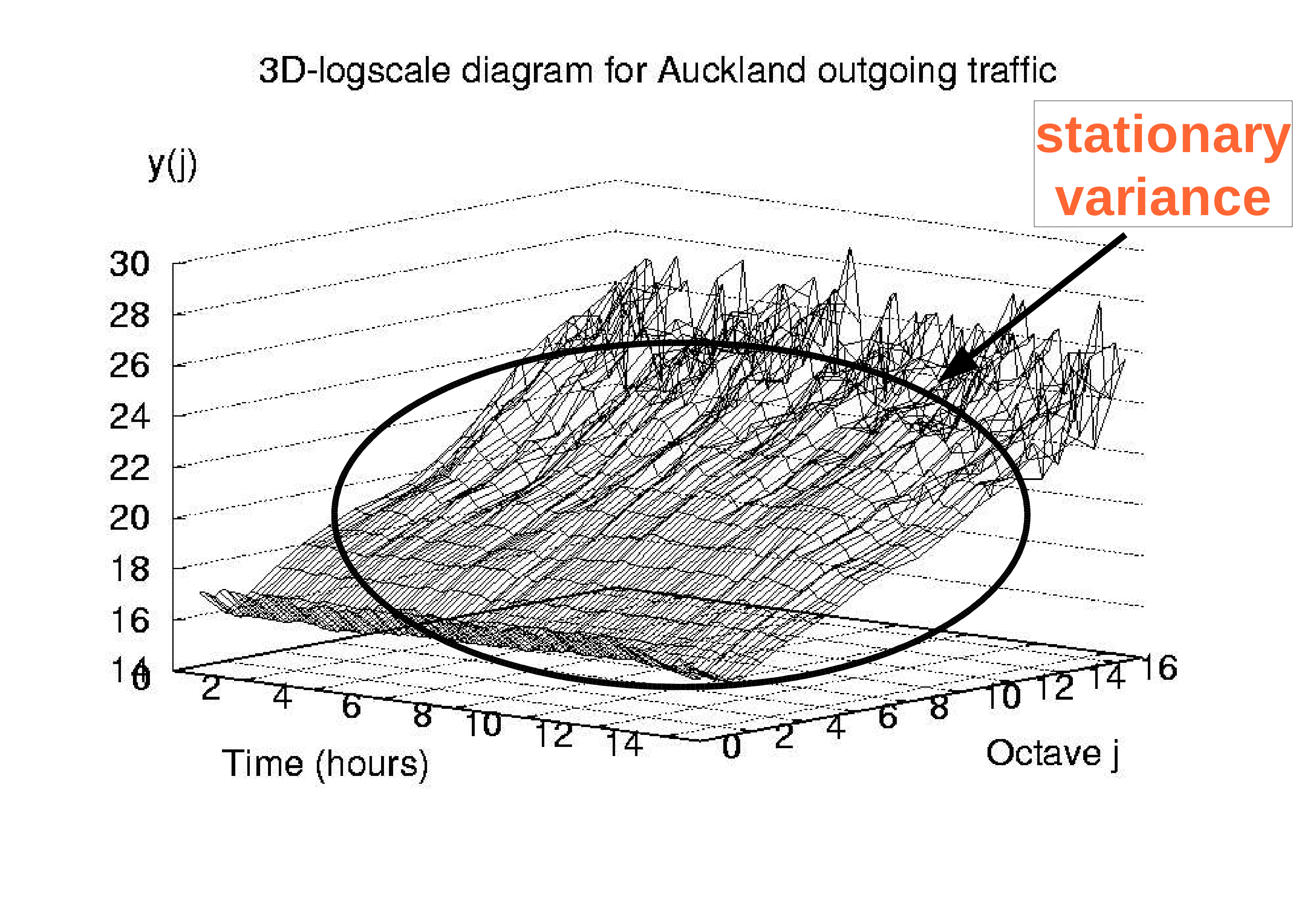
Cascade analysis (out)



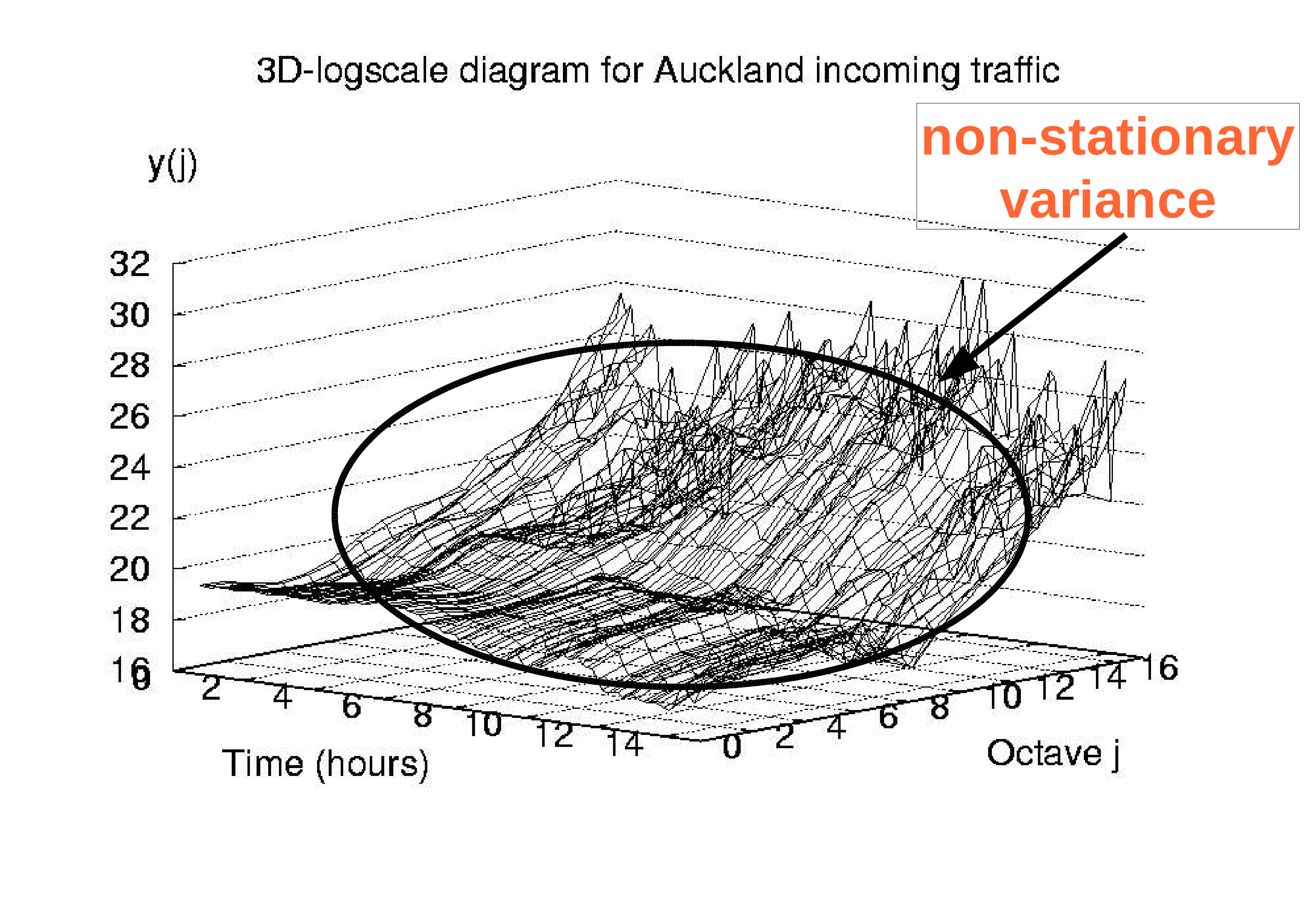
Cascade analysis (in)



Variance analysis (out)



Variance analysis (in)



Conclusions

- Cascade model captures well invariance in TCP behavior
- Cascade model does not fully capture traffic dynamics, only TCP traffic segmentation
- Cascade parameters do not show effect of cross-traffic, 2^{nd} order properties do

After-thoughts

- Cascade is OK to generate clean TCP traffic at timescales below seconds
- Cascade is not OK to simulate realistic Internet traffic
- More work necessary to understand role of the “ abstract” cascade parameters on the traffic behavior
- Can we distinguish between “ normal” and “ abnormal” TCP flows thanks to real-time monitoring of cascade parameters ?