

# An Overview of SAMSI Internet Program Activities

by J. S. Marron (& Darryl Veitch & Cheolwoo Park)

Statistical and Applied Mathematical Sciences Institute  
(<http://www.samsi.info/>)

&

Department of Statistics and Operations Research  
University of North Carolina

## Goals of SAMSI

- Catalyze research in important areas
- Bring together (new) teams of people  
(statisticians, applied mathematicians, domain scientists)
- Goals: serious and long-lasting collaborations
- “Catalytical, not necessarily Conclussional”

# Approaches of SAMSI

## Workshops & Meetings

- Aimed at starting engagement (not “just a meeting center”)

## Research programs

- 6 months to one year (usually one year)
- Continuing activities by:
  - Local participants
  - Long & Short term visitors

# SAMSI Workshops

Main Goal: Engage people in on-going program activities

Approaches:

- Theme Problems
- Small Group Discussions (with reports)
- 5 Minute Madness (Overview of new research + introduc'n)
- Pie in the Sky (what we WISH we could do)

# The Original Internet Program Research Plan

Major research thrusts in 4 directions:

- Measurement and Modeling
- Heavy Traffic - Congestion Control
- Internet Tomography
- Sensor Networks

## The Original Research Plan (cont.)

Opening Events: Workshops (1.5 days)

Sept. 18-19: Measurement and Modelling

Sept. 19-20: Heavy Traffic

Oct. 13-14: Internet Tomography

Oct. 14-15: Sensor Networks

# Change in Plans



Hurricane Isabel, Sept. 17

## Change in Plans (cont.)

Sept. 18-19: Measurement and Modeling Workshop

- Cancelled, research moving ahead anyway
- Since have local critical mass already
- Held “mini-meeting” Sept. 19 (calm after the storm!)
- Darryl Veitch, Sprint Labs, joined us despite Isabel
- Notes available at:

<http://www.samsi.info/200304/int/int-mmmeeting091903.html>

## Change in Plans (cont.)

Sept. 19-20: Heavy Traffic Workshop

- Workshop environment essential
- So rescheduled to Oct. 31 – Nov. 1
- Thanks for joining us!

# Internet Tomography

Workshop Oct. 13 – 14

Partly supported by:



- Theme Problem I: Testbed Validation
- Theme Problem II: Spatial – Temporal Data Analysis
  - Hot emerging topic (added 2 months before)
  - Motivation: joint monitoring, for network security
  - Nice Connection: a type of Sensory network

## Internet Tomography (cont.)

Theme Problem I: Testbed Validation

Original Ideas:

- Testbed is UNC network (much large scale than labs)
- Understood with *Avaya* analysis & local knowledge
- Use to compare methods when ground truth know
- Context of “Challenge Problems”

## Internet Tomography (cont.)

### Theme Problem I: Testbed Validation (cont.)

#### Lessons from Small Discussion Groups:

- Careful about technicalities  
(about scale, architecture and lack of congestion)
- Lack of interest in “Challenge Problems”
- Collaborative Research is preferred
- Need larger scale effort

# Sensor Networks

Workshop Oct. 14 – 15

Partly supported by:



- Idea: Area ripe for involvement by statisticians
- Traditional talk format (because gap too big)  
(e.g. discussion groups might be too much about language)
- Useful “non-expert discussants”
- Still did 5 minute madness & Pie in the Sky
- Useful introduction, hopes for future collaboration

# SAMSI Measurement and Modeling Research

Following Sept. 19 mini-meeting:

Organized into working groups.

Current status on web page:

<http://www.samsi.info/200304/int/int-project.html>

## SAMSI Measurement and Modeling Research (cont.)

Group discussion led to 2 main areas:

- “Model based” approaches
- “Method based” approaches

Group Discussion indicated interest in 10 topics

- So formed 10 working groups
- Now 8, since made sense to combine some
- Goal: joint research

# SAMSI Measurement and Modeling Research (cont.)

Look at: some typical pages

[Change Points and Extremes](#)

[SiZer and Wavelets](#)

Note: links to interesting data sets at bottom

(freely available, thanks to UNC DiRT Group)

## SAMSI Measurement and Modeling Research (cont.)

### Tools:

- **SiZer**, [Example](#) (non-White Noise structure)
- Dependent **SiZer**, [Example](#) (all FGN structure)
- Wavelet Spectra, [Toy E.g.](#) [Real E.G.](#) (bilinear structure)

(Examples here: series of 1 ms binned packet counts, UNC, 2002)

## SAMSI Measurement and Modeling Research (cont.)

Concept introduced by Darryl Veitch (at post-Isabel mini-meeting):

“Semi-experiments” ([Darryl's talk](#))

Idea: explore “causation” issues by:

- Modifying the data
- Viewing the result

Statistical Note:

Philosophy similar to permutation testing & bootstrapping

## Semi-Experiments

Nice Work by Cheolwoo Park: Explain the “Bump”

Starting Point: [Wavelet Spectrum](#) for Saturday, 13:00

- Strange “peak” at scale  $j = 11$
- Suggests “something happening at high frequency”???
- Happens at scale:  $2^{11} ms = 2048ms \approx 2sec$  ???
- Effect dominates somewhat coarser scales?
- Not typical for FGN or other common models.....

## Semi-Experiments (cont.)

### Dependent SiZer View ([Sat. 1300](#))

- Shows huge spike
- Not at all explainable using FGN model
- Which is reasonable in other time regions (magenta)....
- What is happening?
- Deeper look: Zoom in to part between red bars

## Semi-Experiments (cont.)

### Zoomed View:

- Looks mostly consistent with FGN (magenta)
- Significant valley in center?
- Average bin counts ~30
- Versus ~20 for “background”
- So really are “on top of plateau”
- Zoom in more?      To region between red bars

## Semi-Experiments (cont.)

### Second Zoomed View:

- Surprising periodic structure???
- Only at fine scales? Washes away at larger scales....
- Period ~ 3 sec.....
- Corresponds to scale of bump in [wavelet spectrum](#)!?!?
- Ande Broido explanation: IP & port # scan of UNC

## Semi-Experiments (cont.)

Semi-experimental investigation:

Eliminate bump, by “excising the spike”

i.e. the data between the **red bars**

Resulting [wavelet spectrum](#)

- Bump has disappeared
- Suggests was indeed caused by small time scale anomaly

## Semi-Experiments (cont.)

Note: Bump in spectrum was driven by “periodicities”,

*Not* “vertical shift”

[Related example](#) with “similar big shift” (downwards this time)

[Wavelet spectrum](#):

- No “bump” at a particular frequency
- Looks very standard
- Since no “periodic-like component”?

## Semi-Experiments (cont.)

### Other criteria besides Wavelet Spectrum

From working group led by Mike Devetsikiotis:

- Autocorrelation
- Fourier Spectrum
- Marginal Distributions
- SiZer
- Queueing Performance

## Semi-Experiments (cont.)

Why queueing performance?

Can see (important) differences not seen by others

Nice elucidation: presented by David Rolls,

At Suite of Models Working Group

Based on:

T. Taralp, M. Devetsikiotis, and I. Lambadaris (1999) In Search of Better Statistics for Traffic Characterization, *Journal of the Brazilian Computer Society*, 5, 5-13.

## Why queueing performance?

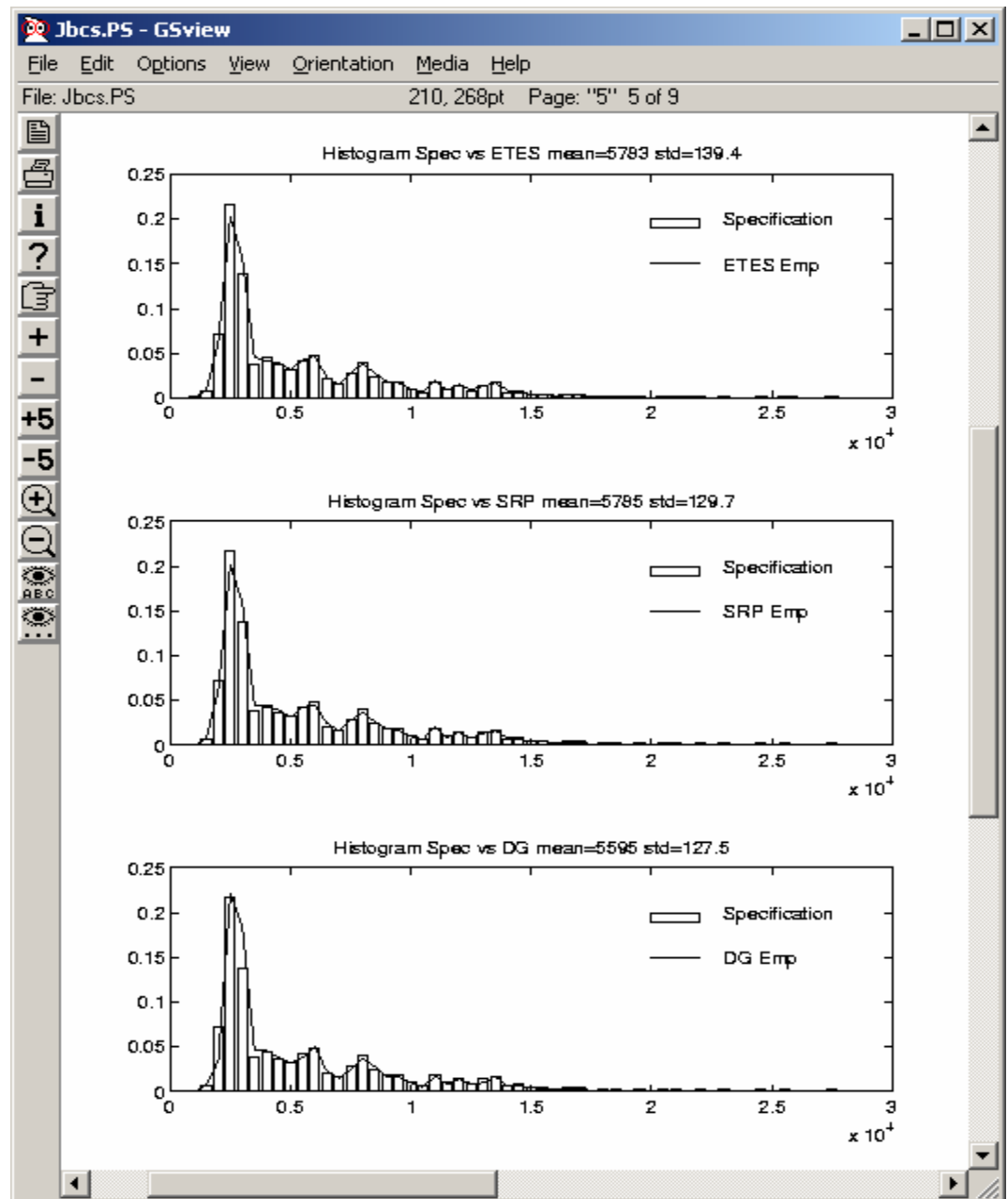
Experiment: 3 ways of generating traces

- Essentially same marginal distributions
- Essentially same autocovariance
- But *very* different queue length distributions

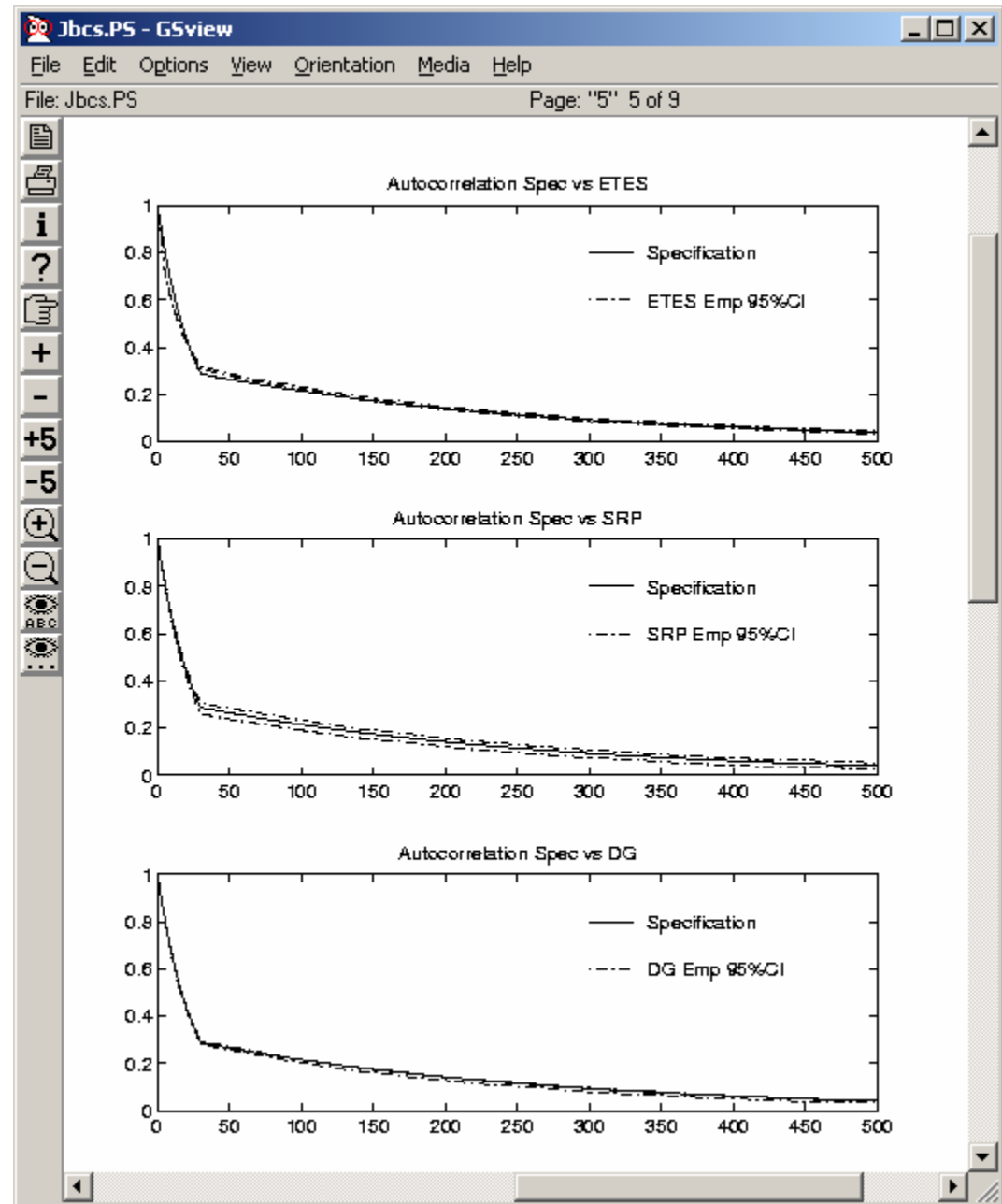
Shows queueing behavior provides “different window” on data

- Window of keen interest in internet world.....

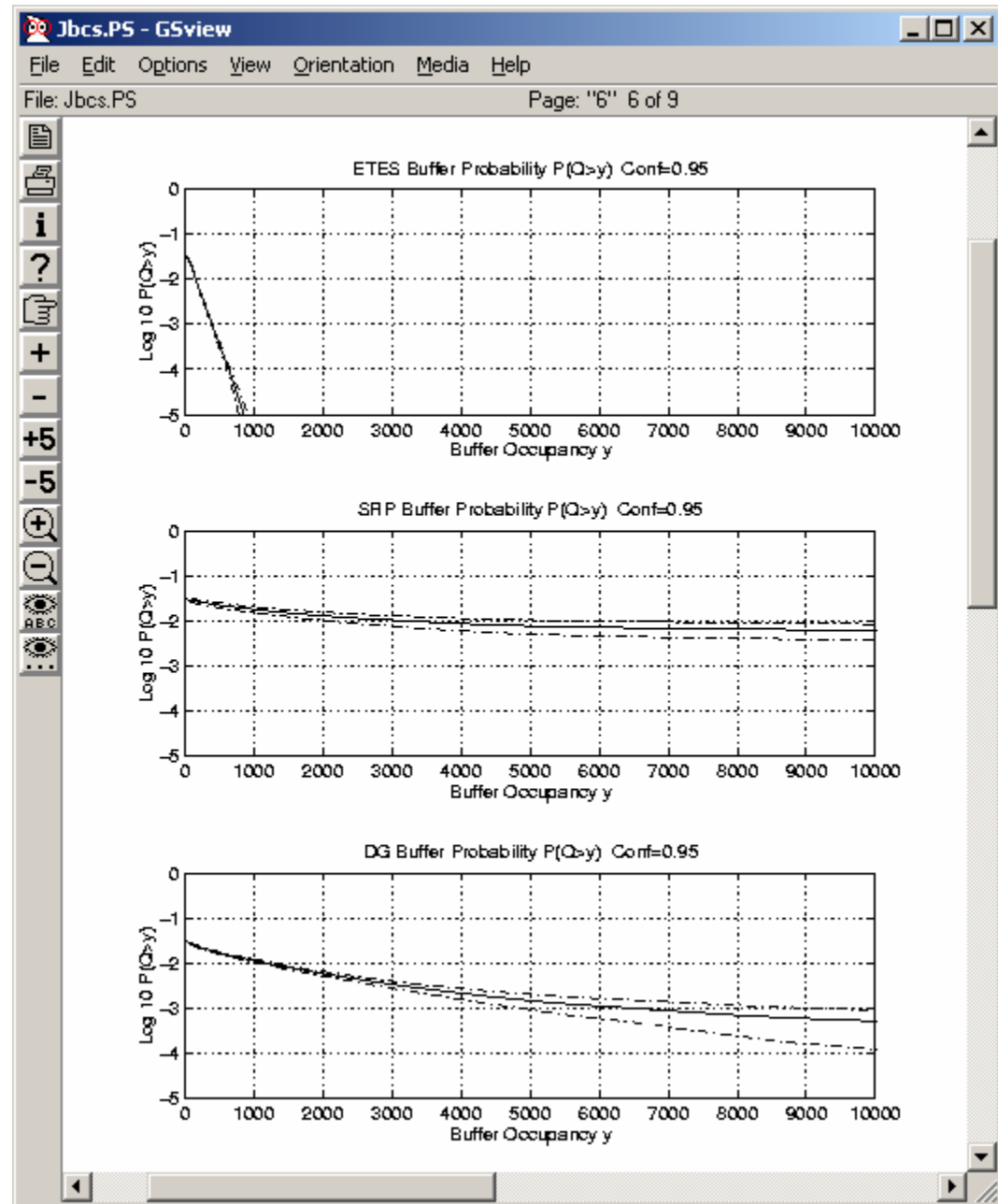
Same marginals:



# Same Autocorrelation:



# Different Queue Lengths



## Semi-Experiments (cont.)

Interesting upcoming work, by Felix Hernandez Campos

Semi-experiments in the lab?

Idea: vary things as in semi-experiments (real experiments?)

With goal of understanding “causal effects”

Big plus: are studying proper “closed loop” phenomena...