Comparison of Two Dimension-Reduction Methods for Network Simulation Models

image generated using http://www.wordle.net/

Kevin Mills (ANTD) & Jim Filliben (SED)
Complex Systems Study Group
NIST SED/ANTD Seminar 9/22/11 222/A326

millsjjfsedtalk092211.pptx

Comparison of Two Dimension-Reduction Methods for Network Simulation Models

image generated using http://www.wordle.net/

Kevin Mills (ANTD) & Jim Filliben (SED)
Complex Systems Study Group

NIST SED/ANTD Seminar 9/22/11   222/A326

millsjjfsedtalk092211.pptx

Motivation for this Talk

1. Generate useful internet MesoNet modeling conclusions & insight

2. Show stat framework/approach & methodology + beginning-to-"end" demo

3. Show dimension reduction dependency on Design of Experiment & Sensitivity Analysis

Outline

- CxS: Complex System IMS Project
- Goal – Problem – Solution
- Stat Framework
- Overview of Candidate *MesoNet* Factors & Responses
- Experiment Design
- Sensitivity Analysis
- Dimension Reduction
  - via Correlation Analysis with Clustering
  - via Principal Components Analysis
- Comparison of Dimension Reduction Techniques
- Conclusions

IMS Project: Measurement Science for Complex Information System

http://www.nist.gov/itl/antd/emergent_behavior.cfm

This project aims to develop and evaluate a coherent set of methods to understand behavior in complex information systems, such as the Internet, computational grids and computing clouds.

Such large distributed systems exhibit global behavior arising from independent decisions made by many simultaneous actors, which adapt their behavior based on local measurements of system state.

Actor adaptations shift the global system state, influencing subsequent measurements, leading to further adaptations.

This continuous cycle of measurement and adaptation drives a time-varying global behavior.

For this reason, proposed changes in actor decision algorithms must be examined/understood at large spatiotemporal scale in order to predict (and control) system behavior.

CxS Project

What is the problem? No one understands how to measure, predict or control macroscopic behavior in complex information systems: (1) threatening our nation's security and (2) costing billions of dollars.

"[Despite] *society's profound dependence on networks, fundamental knowledge about them is primitive.* [G]*lobal communication … networks have quite advanced technological implementations but their behavior under stress still cannot be predicted reliably.… There is no science today that offers the fundamental knowledge necessary to design large complex networks* [so] *that their behaviors can be predicted prior to building them*." (above quote from Network Science 2006, a National Research Council report)

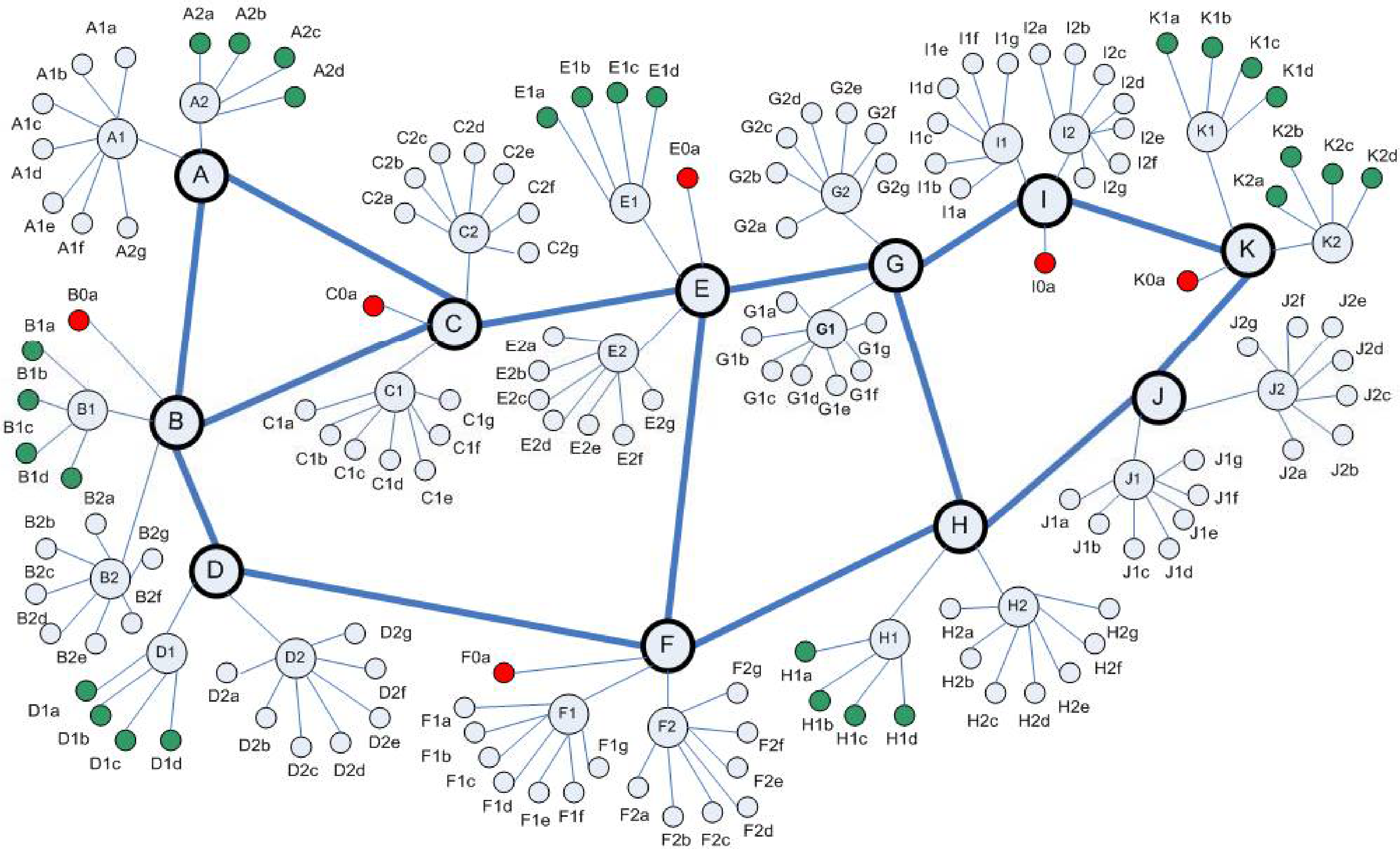
Project Goal – Problem – Solution

- **Goal** – understand internet congestion and compare proposed Internet congestion control algorithms under a wide range of controlled, repeatable conditions, as simulated by selecting combinations of parameter values for *MesoNet*, a 11- to 20-parameter network simulator.

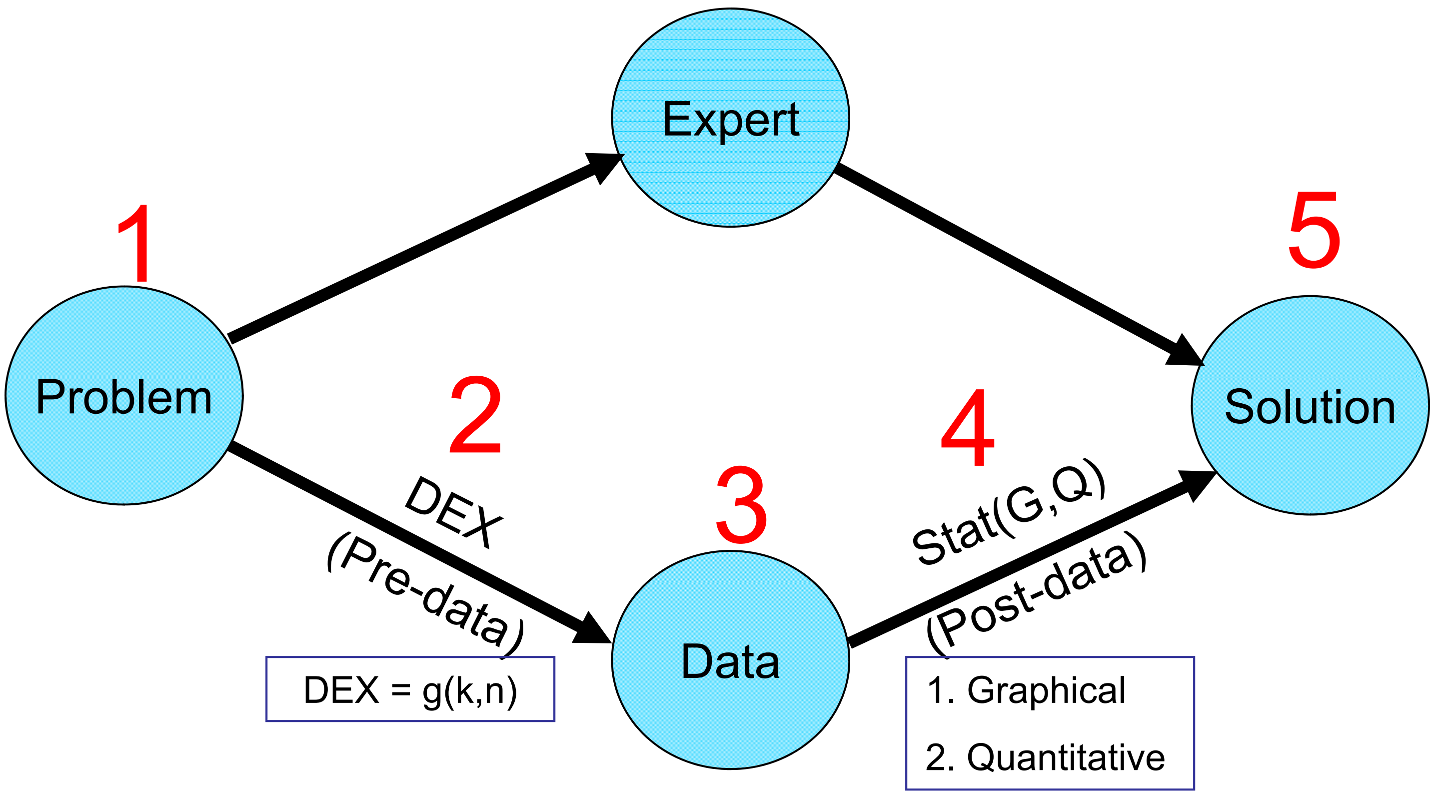
- **Problem** – how to determine which *MesoNet* core responses to analyze when characterizing model behavior.

- **Solution** – apply experiment design techniques to generate an affordable but representative data sample, and carry out the subsequent response variable evaluation via three data analysis approaches:
    1. sensitivity analysis
    2. correlation analysis with clustering &
    3. principal components analysis

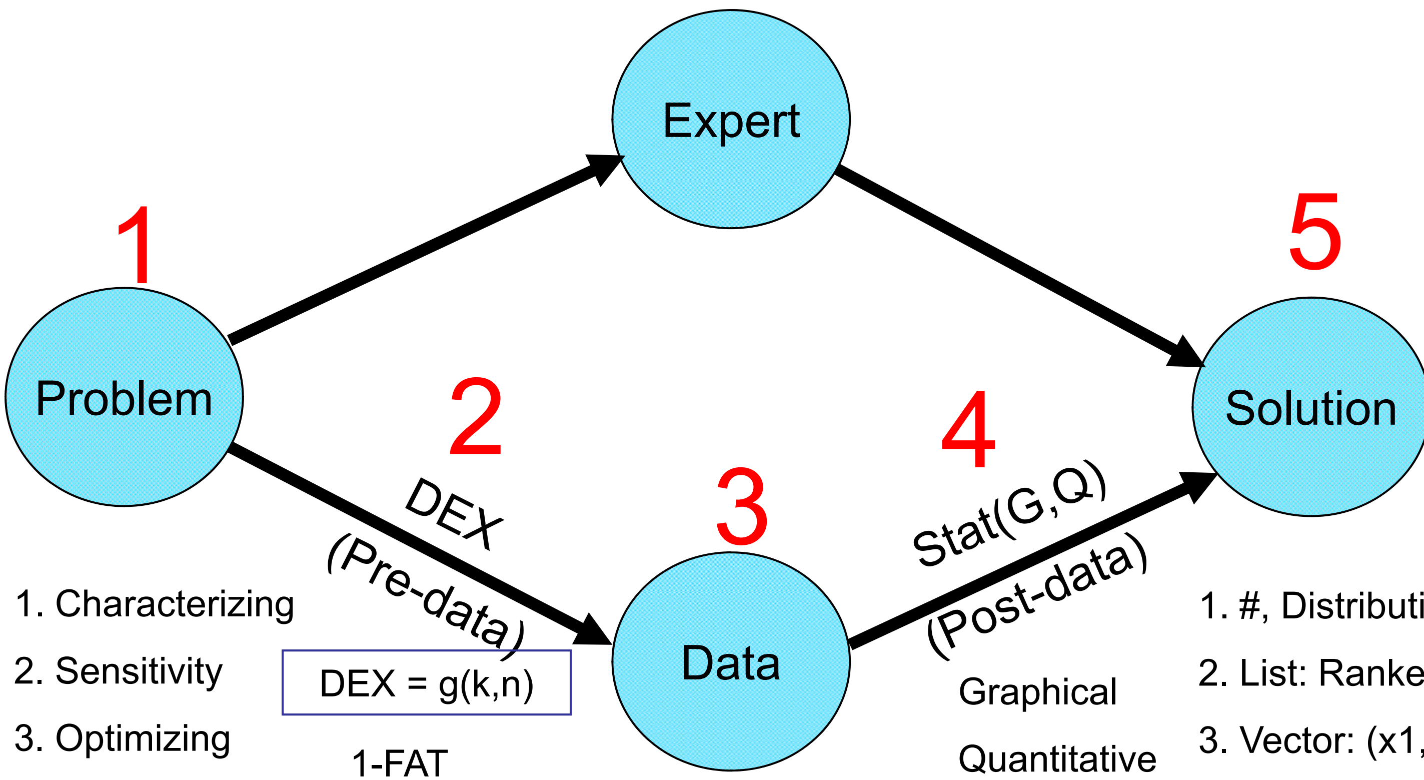
Abilene Network (3-Tier MesoNet Topology)



General Problem-Solving Framework



General Problem-Solving Framework



1. Characterizing
2. Sensitivity
3. Optimizing
4. Modeling
5. Comparing
6. Predicting
7. Uncertainty
8. Verifying
9. Validating

1-FAT

Monte Carlo

Latin HC

Orthogonal

Resp Surface

Reality

Lab

Computational

Graphical

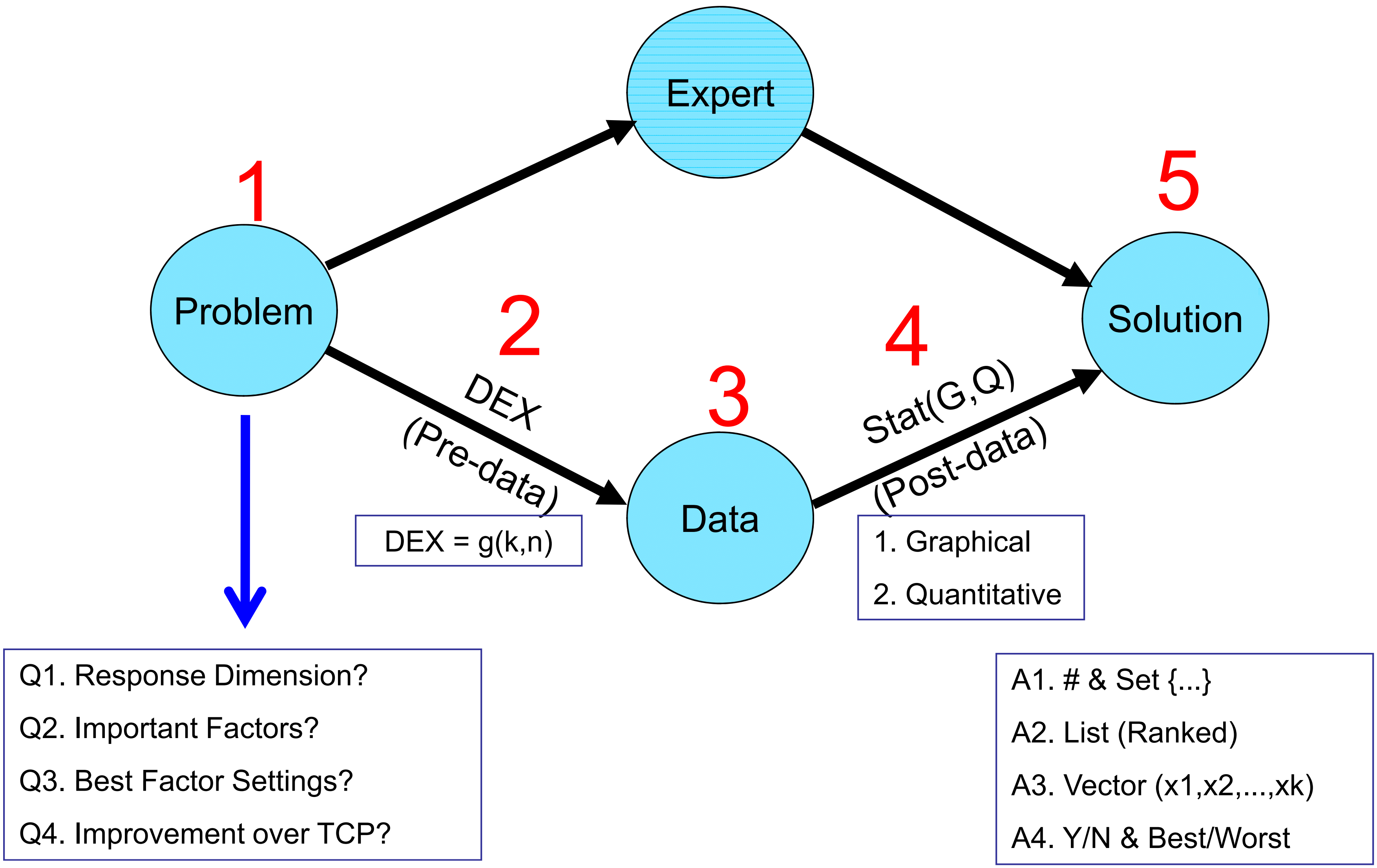
Quantitative

1. #, Distribution
2. List: Ranked Factors
3. Vector: (x1,…,xk)
4. f
5. Y/N

6 #

7. SD(#)
8. Y/N, Vector: (x1, …,xk)
9. Y/N, Vector: (x1,

General Problem-Solving Framework



The Starting Point: Generic Model

System Behavior Y = f(X1, X2, ..., Xk)

1. Y = f(X1, X2, ..., Xk)	Comparative
2. Y = f(X1, X2, ..., Xk)	Screening
3. Y = f(X1, X2, ..., Xk)	Regression
4. Y = f(X1, X2, ..., Xk)	Optimization
5. Y = f(X1, X2, ..., Xk) = c	Consensus
6. Y = f(X1, X2, ..., Xk)	Dimension Red.

The Starting Point: Generic Model

System Behavior Y = f(X1, X2, ..., Xk)

1. Y = f(X1, X2, ..., Xk)	Comparative
2. Y = f(X1, X2, ..., Xk) ✓	Screening
3. Y = f(X1, X2, ..., Xk)	Regression
4. Y = f(X1, X2, ..., Xk)	Optimization
5. Y = f(X1, X2, ..., Xk) = c	Consensus
6. Y = f(X1, X2, ..., Xk) ✓	Dimension Red.

The Starting Point: Generic Model (Part 2)

System Behavior $Y = f(X1, X2, ..., Xk)$

System Behavior $Y_i = f_i(X1, X2, ..., Xk)$ $(i = 1, 2, ..., m)$

$\boxed{\text{Unknowns: } (k=?, n=?, m=?)}$

Factor <u>Groups</u> Affecting System Behavior

1. Network Factors
2. User Factors
3. Source & Receiver Factors
4. Protocol Factors

Factors X_i Affecting System Behavior
$Y_i = f_i(X1, X2, ..., Xk)$

Network Factors	x1	Propagation delay
	x2	Network speed
	x3	Buffer sizing
User Factors	x4	Average file size for web pages
	x5	Average think time between web clicks
	x6	Probability a user opts to transfer a larger file
Source & Receiver Factors	x7	Probability a source or receiver is on a fast host
	x8	Scaling factor for number of sources & receivers
	x9	Distribution of sources
	x10	Distribution of receivers
Protocol Factors	x11	Initial TCP slow-start threshold

(k=11,n=?,m=?)

Affordable Number of Runs n = ?

n <= 100

4 Ways to Reduce DEX Full Factorial Design n:

1. Reduce # Factors (but scope reduced)
2. Reduce Number of Levels (=> 2?)
3. Reduce Number of Reps
4. Fractional Factorial Design

(k=11,n <= 100,m=?)

Affordable Number of Runs n = ?

n <= 100

4 Ways to Reduce DEX Full Factorial Design n:

1. Reduce # Factors (but scope reduced)
2. Reduce Number of Levels (2?)
3. Reduce Number or Reps
4. Fractional Factorial Design

(k=11,n <= 100,m=?)

Affordable Number of Runs n = ?

Additional Desirable Feature of the Design:

Good Estimates for (at least) the
Main Effects & 2-Term Interactions
(Resolution)

11 + 11-choose-2 = 11 + 55 = 66

(66+1) = 67 ➔ 64 ➔ 2^6 ➔ 2^{11-5}

Final Design: 2^{11-5} Orthogonal 2-Level Fractional Factorial Design (k=11,n=64)

$\boxed{(k=11,n=64,m=?)}$

MesoNet Factors (k=11) & Levels (2)

Category	Factor	Code	Definition	Level 1: -	Level 2: +
Network Factors	x1	PDM	Propagation delay	1	2
	x2	BRS (s)	Network speed	800 p/ms	400 p/ms
	x3	QSA	Buffer sizing	*RTT*x*C*/SQRT(*n*)	*RTT*x*C*
User Factors	x4	AvFSWO	Average file size for web pages	50 packets	100 packets
	x5	AvThT	Average think time between web clicks	2000 ms	5000 ms
	x6	PrLF	Probability a user opts to transfer a larger file	0.02	0.01
Source & Receiver Factors	x7	PrFH	Probability a source or receiver is on a fast host	0.4	0.2
	x8	SFSR	Scaling factor for number of sources & receivers	2	3
	x9	SDist	Distribution of sources	WEB	P2P
	x10	RDist	Distribution of receivers	WEB	P2P
Protocol Factors	x11	SST	Initial TCP slow-start threshold	43 packets	1.07×10^9 packets

2^{11-5} Orthogonal Fractional Factorial Design (k = 11, n = 64)

Generators:

X7 = X3*X4*X5
X8 = X1*X2*X3*X4
X9 = X1*X2*X6
X10 = X2*X4*X5*X6
X11 = X1*X4*X5*X6

Resolution IV

Reference: Box, Hunter, & Hunter, "Statistics for Experimenters", 2nd Edition, 2005, Wiley, p. 272

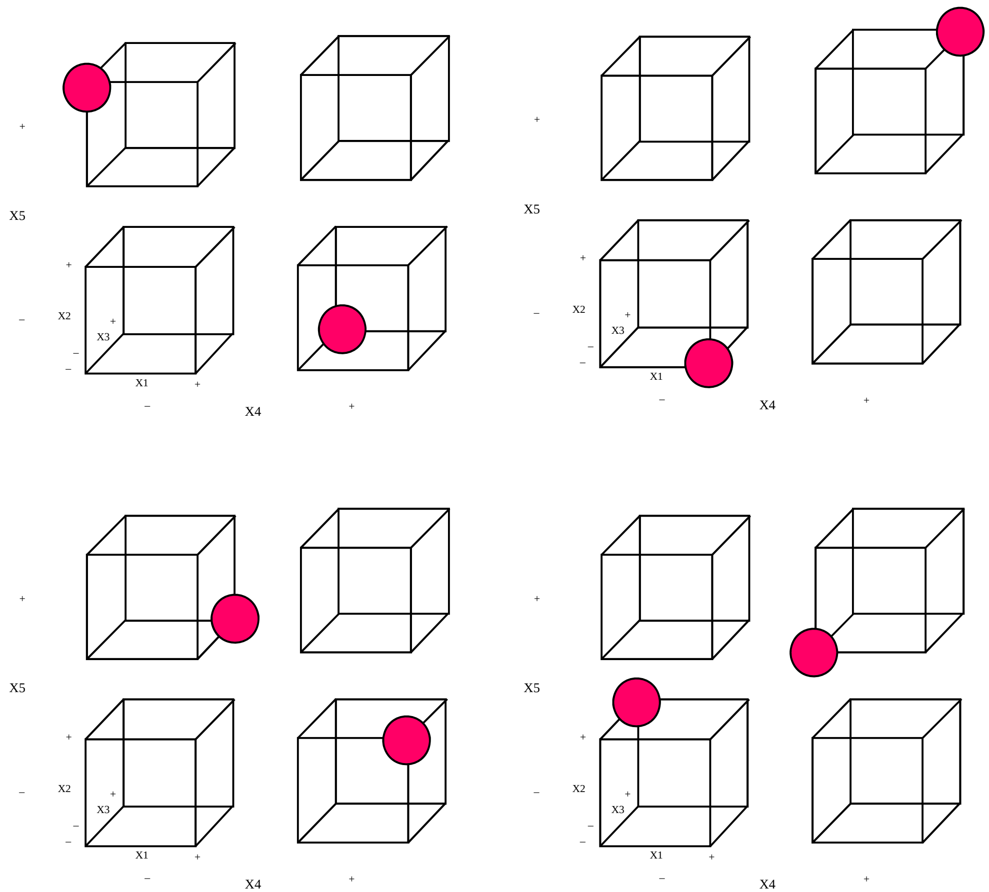
2^{11-5} Fractional Factorial Design (k=11,n=64)
(2to11m5.xls)

Index	*X1*	*X2*	*X3*	*X4*	*X5*	*X6*	*X7*	*X8*	*X9*	*X10*	*X11*
1	-1	-1	-1	-1	-1	-1	-1	+1	-1	+1	+1
2	+1	-1	-1	-1	-1	-1	-1	-1	+1	+1	-1
3	-1	+1	-1	-1	-1	-1	-1	-1	+1	-1	+1
4	+1	+1	-1	-1	-1	-1	-1	+1	-1	-1	-1
5	-1	-1	+1	-1	-1	-1	1	-1	-1	+1	+1
6	+1	-1	+1	-1	-1	-1	1	+1	+1	+1	-1
7	-1	+1	+1	-1	-1	-1	1	+1	+1	-1	+1
8	+1	+1	+1	-1	-1	-1	1	-1	-1	-1	-1
9	-1	-1	-1	+1	-1	-1	1	-1	-1	-1	-1
10	+1	-1	-1	+1	-1	-1	1	+1	+1	-1	+1
11	-1	+1	-1	+1	-1	-1	1	+1	+1	+1	-1
12	+1	+1	-1	+1	-1	-1	1	-1	-1	+1	+1
13	-1	-1	+1	+1	-1	-1	-1	+1	-1	-1	-1
14	+1	-1	+1	+1	-1	-1	-1	-1	+1	-1	+1
15	-1	+1	+1	+1	-1	-1	-1	-1	+1	+1	-1
16	+1	+1	+1	+1	-1	-1	-1	+1	-1	+1	+1
17	-1	-1	-1	-1	+1	-1	1	+1	-1	-1	-1
18	+1	-1	-1	-1	+1	-1	1	-1	+1	-1	+1
19	-1	+1	-1	-1	+1	-1	1	-1	+1	+1	-1
20	+1	+1	-1	-1	+1	-1	1	+1	-1	+1	+1
21	-1	-1	+1	-1	+1	-1	-1	-1	-1	-1	-1
22	+1	-1	+1	-1	+1	-1	-1	+1	+1	-1	+1
23	-1	+1	+1	-1	+1	-1	-1	+1	+1	+1	-1
24	+1	+1	+1	-1	+1	-1	-1	-1	-1	+1	+1
25	-1	-1	-1	+1	+1	-1	-1	-1	-1	+1	+1
26	+1	-1	-1	+1	+1	-1	-1	+1	+1	+1	-1
27	-1	+1	-1	+1	+1	-1	-1	+1	+1	-1	+1
28	+1	+1	-1	+1	+1	-1	-1	-1	-1	-1	-1
29	-1	-1	+1	+1	+1	-1	1	+1	-1	+1	+1
30	+1	-1	+1	+1	+1	-1	1	-1	+1	+1	-1
31	-1	+1	+1	+1	+1	-1	1	-1	+1	-1	+1
32	+1	+1	+1	+1	+1	-1	1	+1	-1	-1	-1

33	-1	-1	-1	-1	-1	+1	-1	+1	+1	-1	-1
34	+1	-1	-1	-1	-1	+1	-1	-1	-1	-1	+1
35	-1	+1	-1	-1	-1	+1	-1	-1	-1	+1	-1
36	+1	+1	-1	-1	-1	+1	-1	+1	+1	+1	+1
37	-1	-1	+1	-1	-1	+1	1	-1	+1	-1	-1
38	+1	-1	+1	-1	-1	+1	1	+1	-1	-1	+1
39	-1	+1	+1	-1	-1	+1	1	+1	-1	+1	-1
40	+1	+1	+1	-1	-1	+1	1	-1	+1	+1	+1
41	-1	-1	-1	+1	-1	+1	1	-1	+1	+1	+1
42	+1	-1	-1	+1	-1	+1	1	+1	-1	+1	-1
43	-1	+1	-1	+1	-1	+1	1	+1	-1	-1	+1
44	+1	+1	-1	+1	-1	+1	1	-1	+1	-1	-1
45	-1	-1	+1	+1	-1	+1	-1	+1	+1	+1	+1
46	+1	-1	+1	+1	-1	+1	-1	-1	-1	+1	-1
47	-1	+1	+1	+1	-1	+1	-1	-1	-1	-1	+1
48	+1	+1	+1	+1	-1	+1	-1	+1	+1	-1	-1
49	-1	-1	-1	-1	+1	+1	1	+1	+1	+1	+1
50	+1	-1	-1	-1	+1	+1	1	-1	-1	+1	-1
51	-1	+1	-1	-1	+1	+1	1	-1	-1	-1	+1
52	+1	+1	-1	-1	+1	+1	1	+1	+1	-1	-1
53	-1	-1	+1	-1	+1	+1	-1	-1	+1	+1	+1
54	+1	-1	+1	-1	+1	+1	-1	+1	-1	+1	-1
55	-1	+1	+1	-1	+1	+1	-1	+1	-1	-1	+1
56	+1	+1	+1	-1	+1	+1	-1	-1	+1	-1	-1
57	-1	-1	-1	+1	+1	+1	-1	-1	+1	-1	-1
58	+1	-1	-1	+1	+1	+1	-1	+1	-1	-1	+1
59	-1	+1	-1	+1	+1	+1	-1	+1	-1	+1	-1
60	+1	+1	-1	+1	+1	+1	-1	-1	+1	+1	+1
61	-1	-1	+1	+1	+1	+1	1	+1	+1	-1	-1
62	+1	-1	+1	+1	+1	+1	1	-1	-1	-1	+1
63	-1	+1	+1	+1	+1	+1	1	-1	-1	+1	-1
64	+1	+1	+1	+1	+1	+1	1	+1	+1	+1	+1
							345	1234	126	2456	1456

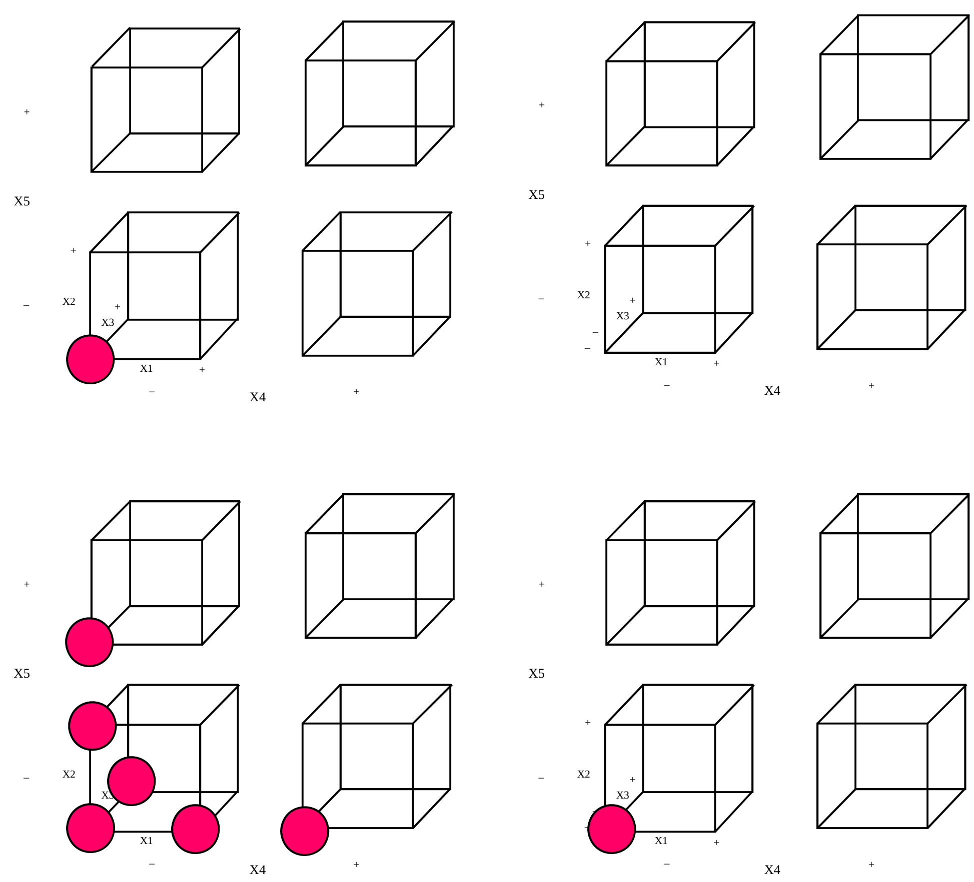
What does this design look like? Why use it?
(k=11,n=64) ➔ (k=7,n=8)

(k=7,n=8) 2^{7-4} Orthogonal Fractional Factorial Design



What does this design not look like?

(k=7,n=8) 1FAT Fractional Factorial Design



Measures of System Behavior (Response Variables)

$$Y_i = f_i(X1, X2, \ldots, Xk)$$

1. Characterizing Macroscopic Behavior

2. Characterizing Instantaneous Throughput for Active Flows by Flow Class (User)

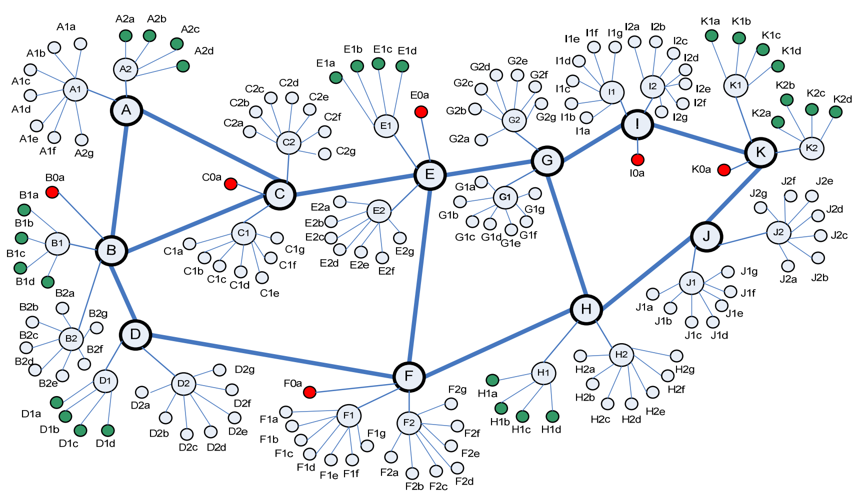
(k=11,n=64,m=?)

16 Responses Characterizing Macroscopic Behavior

Response	Definition
y1	Active Flows – flows attempting to transfer data
y2	Proportion of potential flows that were active: Active Flows/All Sources
y3	Data packets entering the network per measurement interval
y4	Data packets leaving the network per measurement interval
y5	Loss Rate: y4/(y3+y4)
y6	Flows Completed per measurement interval
y7	Flow-Completion Rate: y6/(y6+y1)
y8	Connection Failures per measurement interval
y9	Connection-Failure Rate: y8/(y8+y1)
y10	Retransmission Rate (ratio)
y11	Congestion Window per Flow (packets)
y12	Window Increases per Flow per measurement interval
y13	Negative Acknowledgments per Flow per measurement interval
y14	Timeouts per Flow per measurement interval
y15	Smoothed Round-Trip Time (ms)
y16	Relative queuing delay: y15/(x1x41)

6 Responses Characterizing Instantaneous Throughput for Active Flows by Flow Class

Response	Definition (Throughput in packets/second)
y17	Average Throughput for Active DD Flows
y18	Average Throughput for Active DF Flows
y19	Average Throughput for Active DN Flows
y20	Average Throughput for Active FF Flows
y21	Average Throughput for Active FN Flows
y22	Average Throughput for Active NN Flows



Router Type	Speed
Backbone	2*s*
PoP	25 % of *s*
D-class Access	25 % of *s*
F-class Access	5 % of *s*
N-class Access	2.5 % of *s*

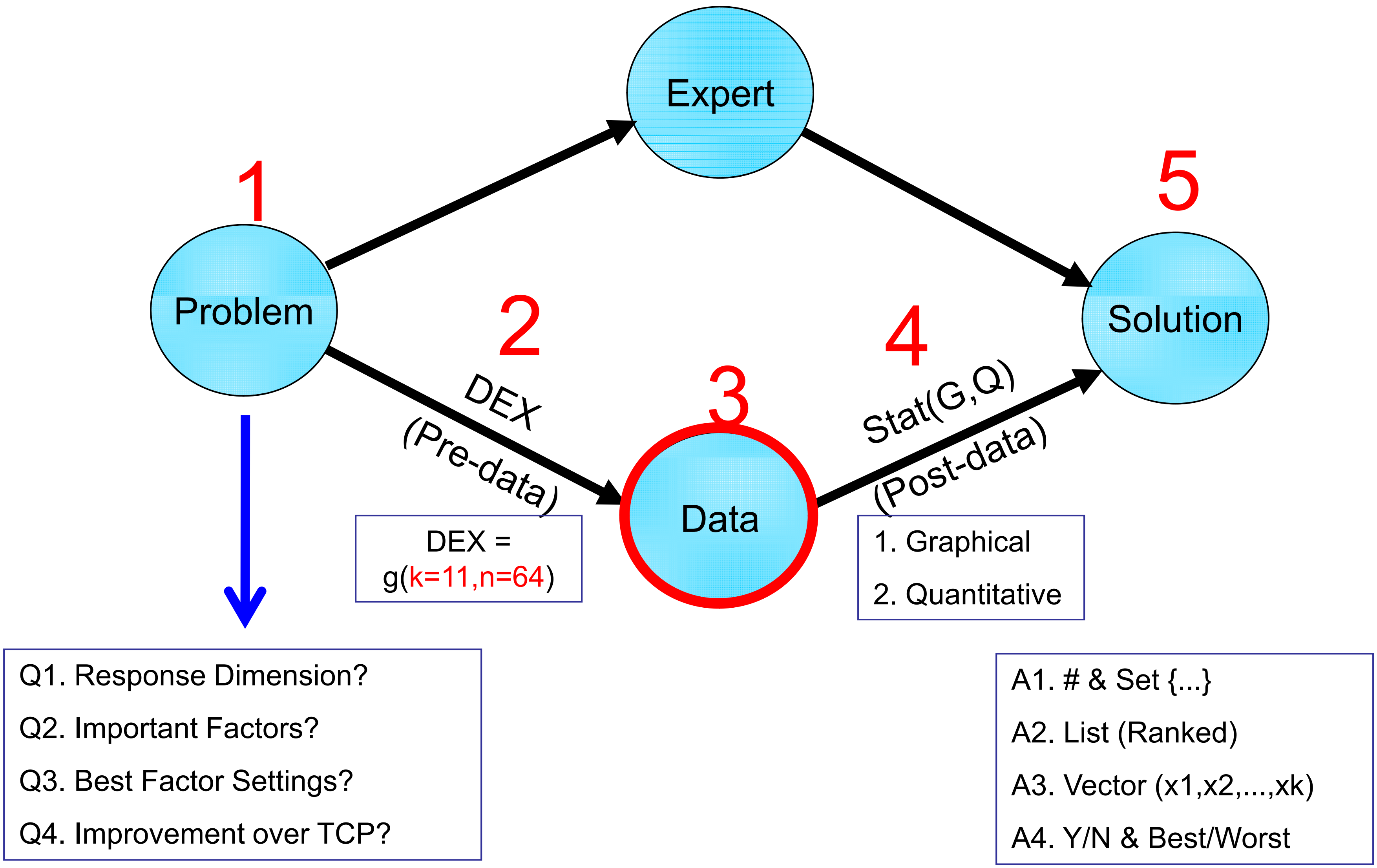
MesoNet 22 Responses: 16 Macro + 6 Throughput

Response	Definition
y1	Active Flows – flows attempting to transfer data
y2	Proportion of potential flows that were active: Active Flows/All Sources
y3	Data packets entering the network per measurement interval
y4	Data packets leaving the network per measurement interval
y5	Loss Rate: y4/(y3+y4)
y6	Flows Completed per measurement interval
y7	Flow-Completion Rate: y6/(y6+y1)
y8	Connection Failures per measurement interval
y9	Connection-Failure Rate: y8/(y8+y1)
y10	Retransmission Rate (ratio)
y11	Congestion Window per Flow (packets)
y12	Window Increases per Flow per measurement interval
y13	Negative Acknowledgments per Flow per measurement interval
y14	Timeouts per Flow per measurement interval
y15	Smoothed Round-Trip Time (ms)
y16	Relative queuing delay: y15/(x1x41)

y17	Average Throughput for Active DD Flows
y18	Average Throughput for Active DF Flows
y19	Average Throughput for Active DN Flows
y20	Average Throughput for Active FF Flows
y21	Average Throughput for Active FN Flows
y22	Average Throughput for Active NN Flows

(k=11,n=64,m=22)

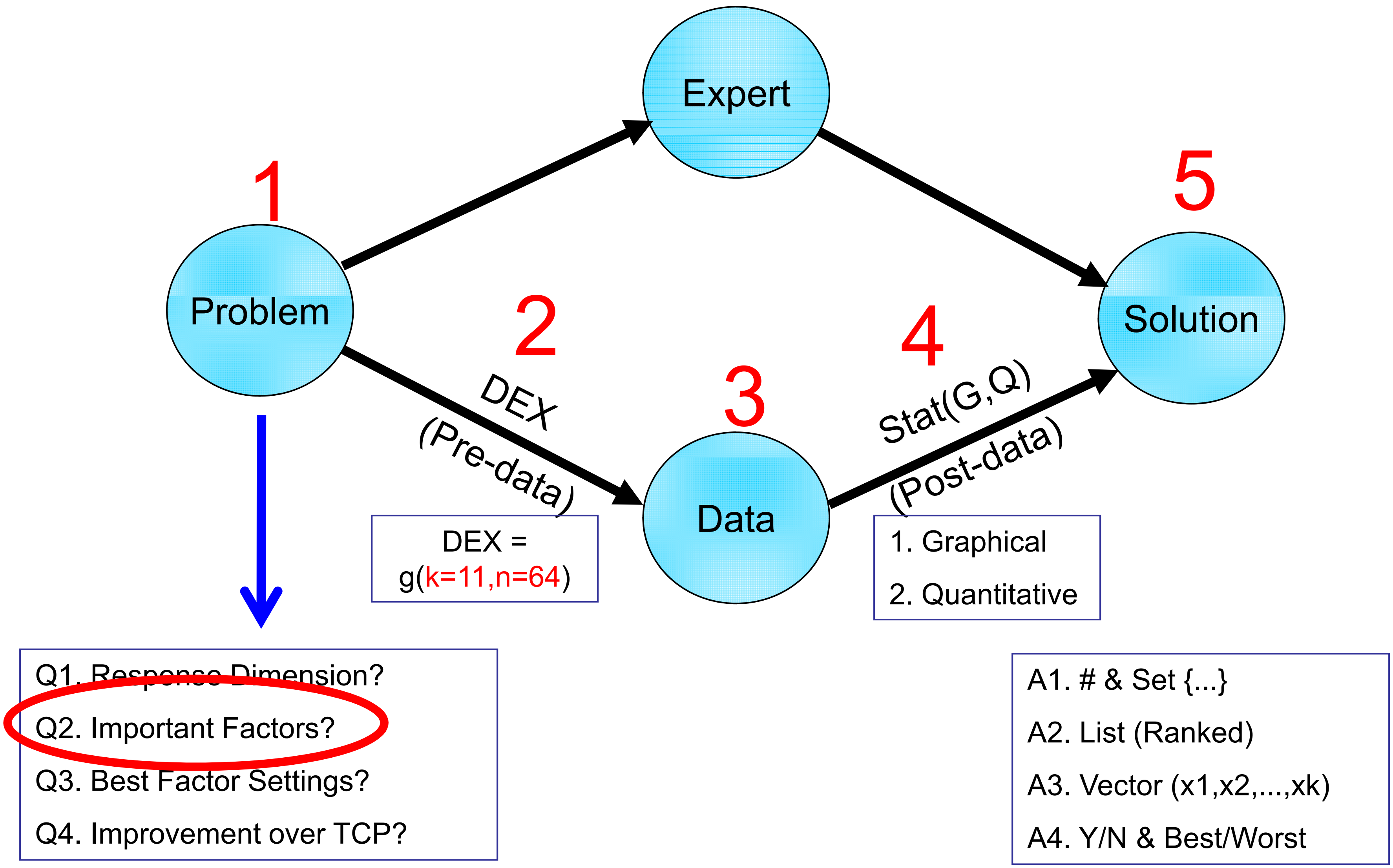
General Problem-Solving Framework



Data: 64 x 22 Multivariate Data Set Resulting from a 2^{11-5} Orthogonal Fractional Factorial Experiment Design

Run	y1	y2	…	y21	y22
1	4680.619	0.168126	…	92.034	89.785
2	6654.512	0.239371	…	72.596	57.738
3	9431.405	0.339259	…	29.569	13.963
4	11565.81	0.415439	…	23.427	19.882
…	…	…	…	…	…
61	10319.55	0.247471	…	87.969	41.573
62	1738.469	0.093668	…	159.298	161.602
63	1783.509	0.096094	…	148.395	161.36
64	21467.6	0.514811	…	26.159	9.981

General Problem-Solving Framework



Sensitivity Analysis

Sensitivity Analysis

Q1. Of the 11 factors, what are most/least important (including interactions)?

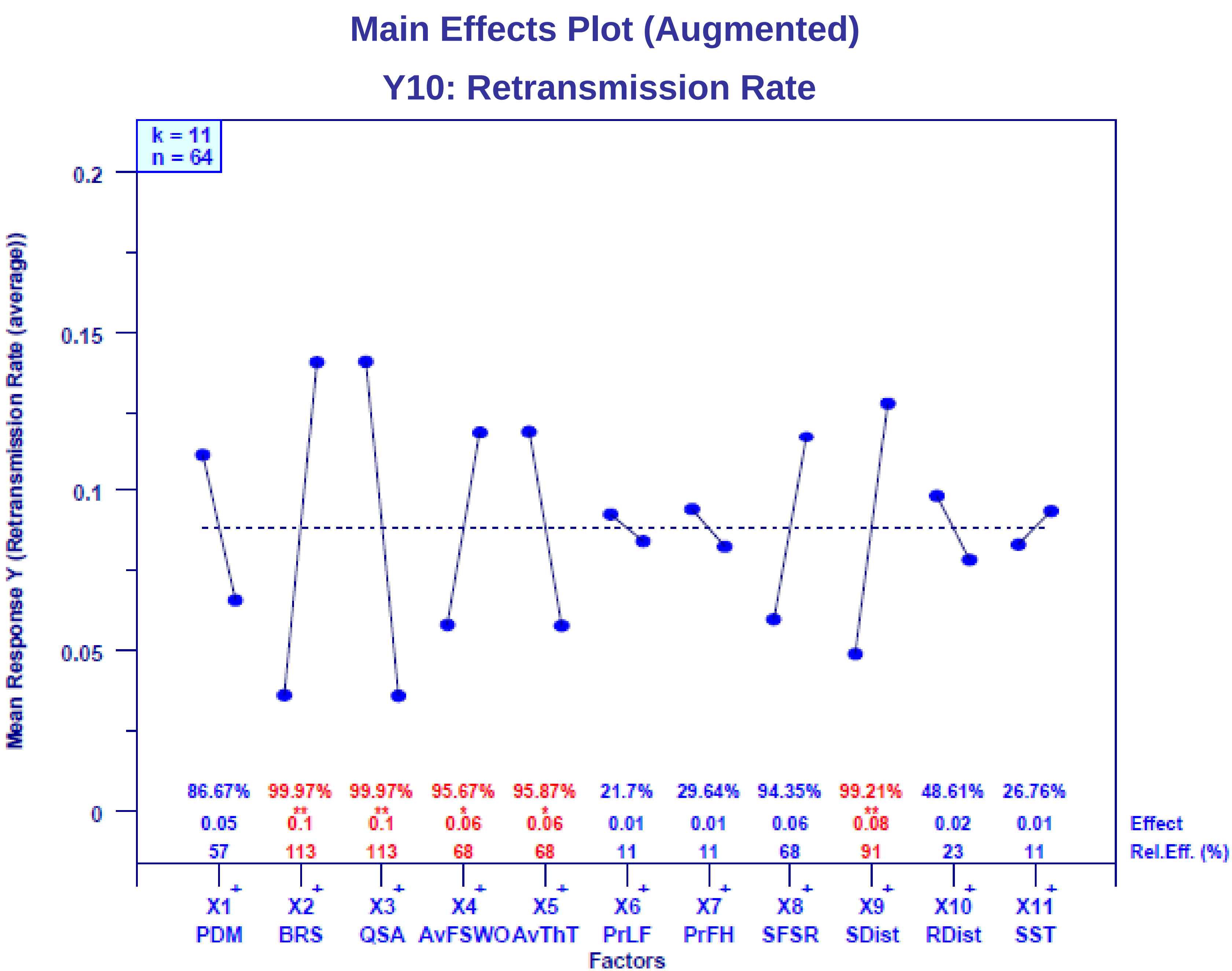
Q2. Robust over the 22 responses?

Analysis: For each of the 22 responses ...
Example 1: Y10 = Retransmission Rate

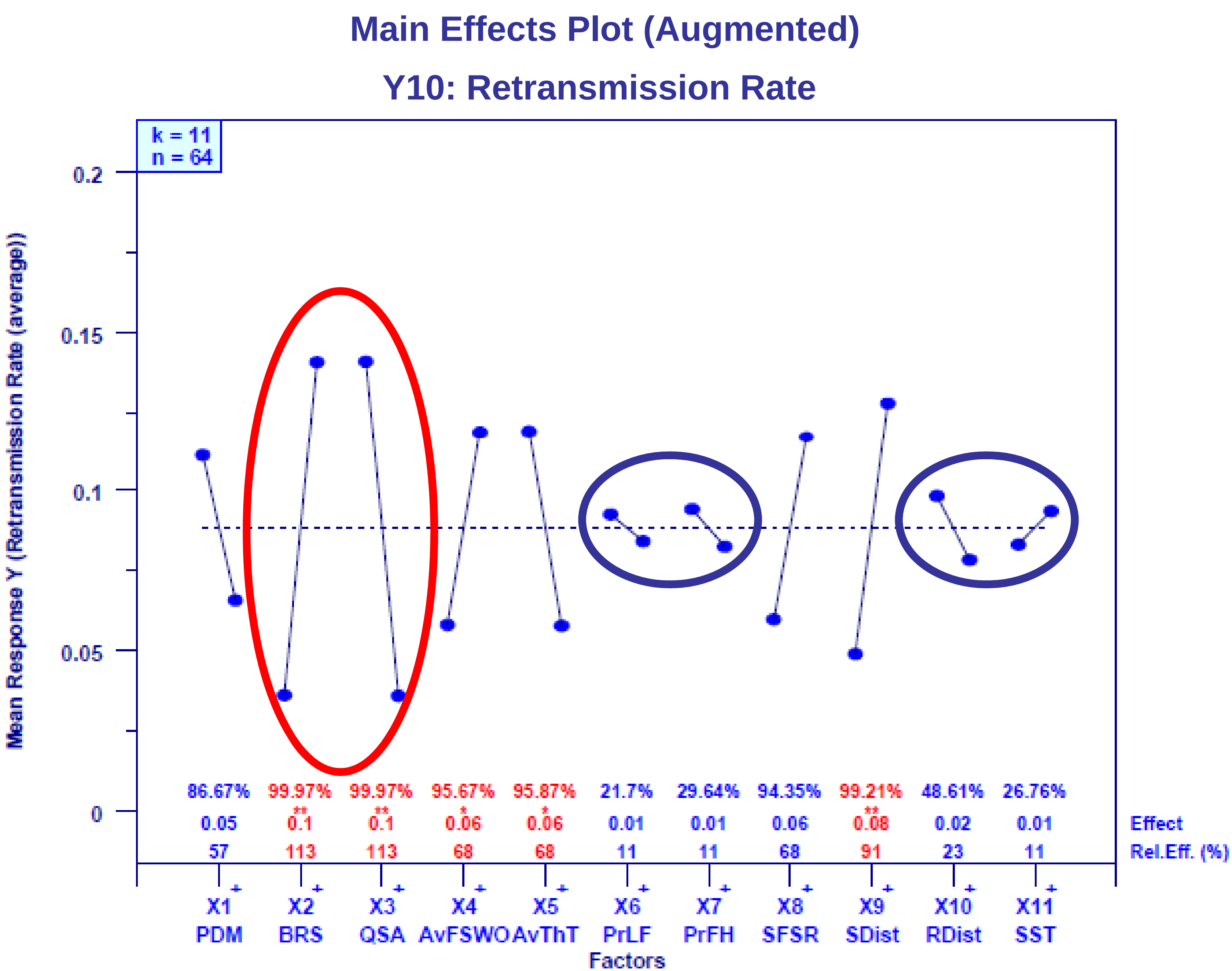
Response	Definition
y1	Active Flows – flows attempting to transfer data
y2	Proportion of potential flows that were active: Active Flows/All Sources
y3	Data packets entering the network per measurement interval
y4	Data packets leaving the network per measurement interval
y5	Loss Rate: y4/(y3+y4)
y6	Flows Completed per measurement interval
y7	Flow-Completion Rate: y6/(y6+y1)
y8	Connection Failures per measurement interval
y9	Connection-Failure Rate: y8/(y8+y1)
y10	Retransmission Rate (ratio)
y11	Congestion Window per Flow (packets)
y12	Window Increases per Flow per measurement interval
y13	Negative Acknowledgments per Flow per measurement interval
y14	Timeouts per Flow per measurement interval
y15	Smoothed Round-Trip Time (ms)
y16	Relative queuing delay: y15/(x1x41)

y17	Average Throughput for Active DD Flows
y18	Average Throughput for Active DF Flows
y19	Average Throughput for Active DN Flows
y20	Average Throughput for Active FF Flows
y21	Average Throughput for Active FN Flows
y22	Average Throughput for Active NN Flows

(k=11,n=64,m=22)



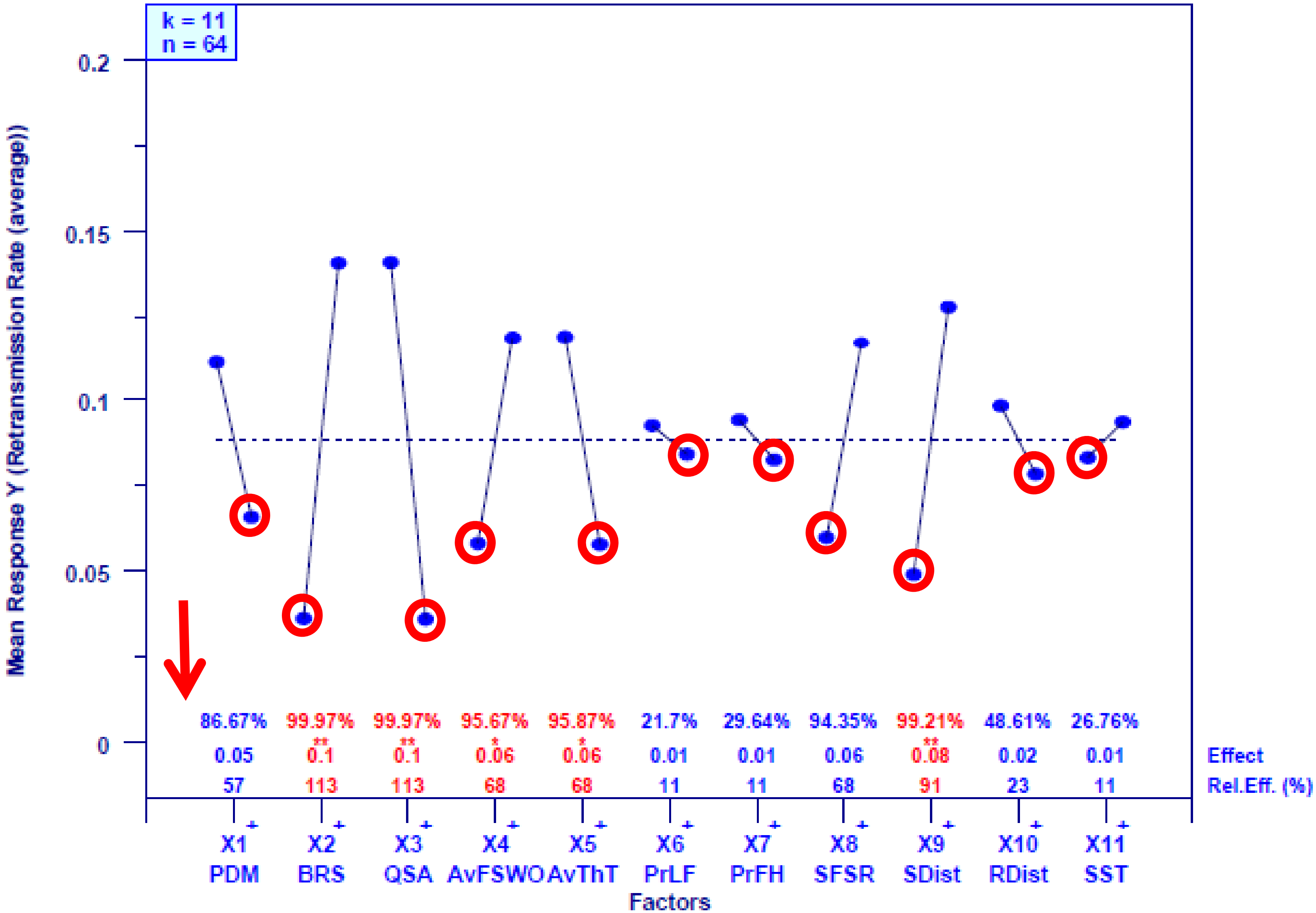
Main Effects Plot (Augmented)
Y10: Retransmission Rate
k = 11
n = 64
Mean Response Y (Retransmission Rate (average))
0.2
0.15
0.1
0.05
0
86.67% 99.97% 99.97% 95.67% 95.87% 21.7% 29.64% 94.35% 99.21% 48.61% 26.76%
0.05 0.1 0.1 0.06 0.06 0.01 0.01 0.06 0.08 0.02 0.01 Effect
57 113 113 68 68 11 11 68 91 23 11 Rel.Eff. (%)
X1 PDM
X2 BRS
X3 QSA
X4 AvFSWO
X5 AvThT
X6 PrLF
X7 PrFH
X8 SFSR
X9 SDist
X10 RDist
X11 SST
Factors



Main Effects Plot (Augmented)
Y10: Retransmission Rate
k = 11
n = 64
Mean Response Y (Retransmission Rate (average))
0.2
0.15
0.1
0.05
0
86.67% 99.97% 99.97% 95.67% 95.87% 21.7% 29.64% 94.35% 99.21% 48.61% 26.76%
0.05 0.1 0.1 0.06 0.06 0.01 0.01 0.06 0.08 0.02 0.01
57 113 113 68 68 11 11 68 91 23 11
Effect
Rel.Eff. (%)
X1 PDM
X2 BRS
X3 QSA
X4 AvFSWO
X5 AvThT
X6 PrLF
X7 PrFH
X8 SFSR
X9 SDist
X10 RDist
X11 SST
Factors

Main Effects Plot (Augmented)

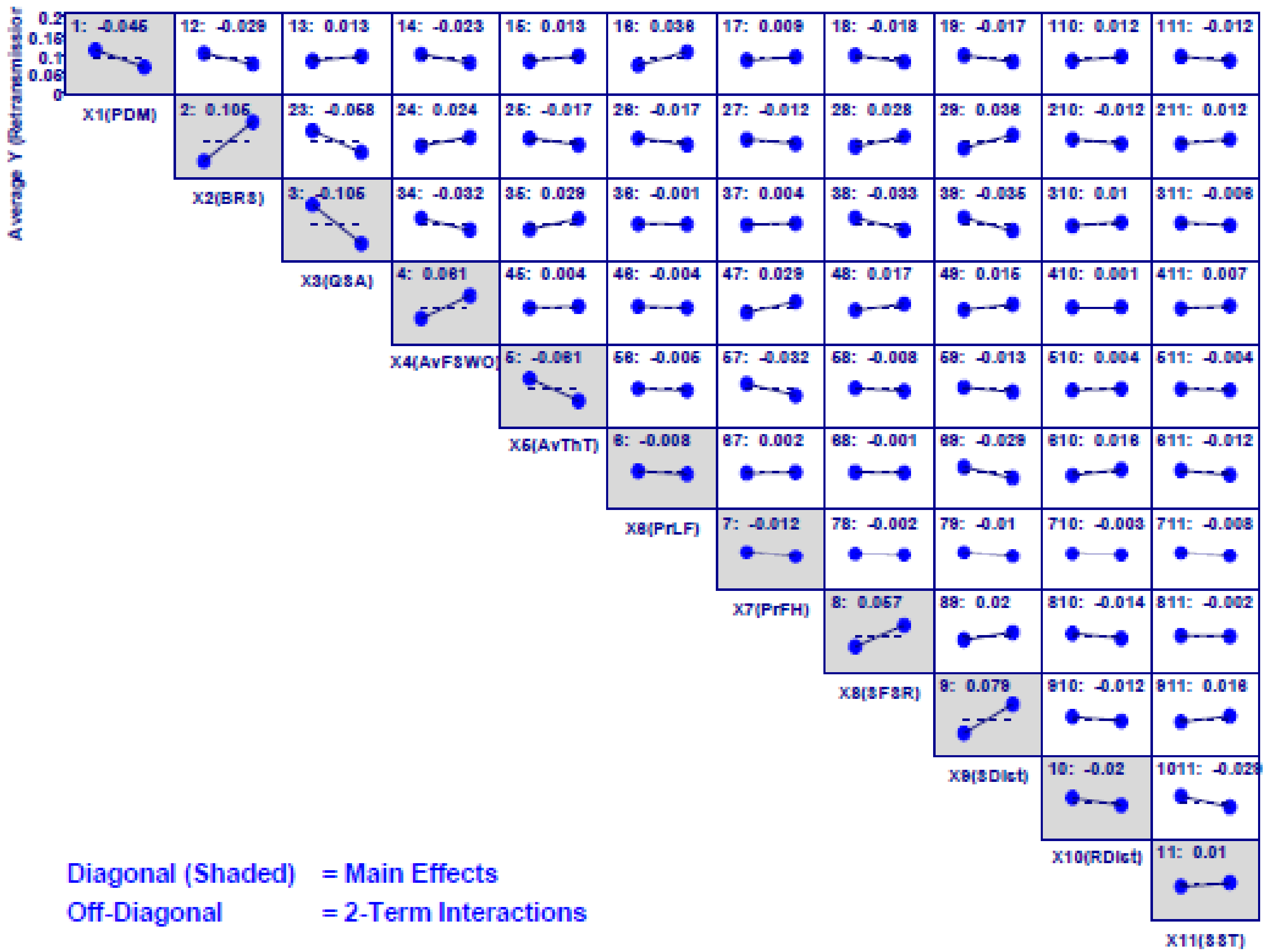
Y10: Retransmission Rate



Means: (+ - + - + + + - - + -)

Interaction Effects Matrix

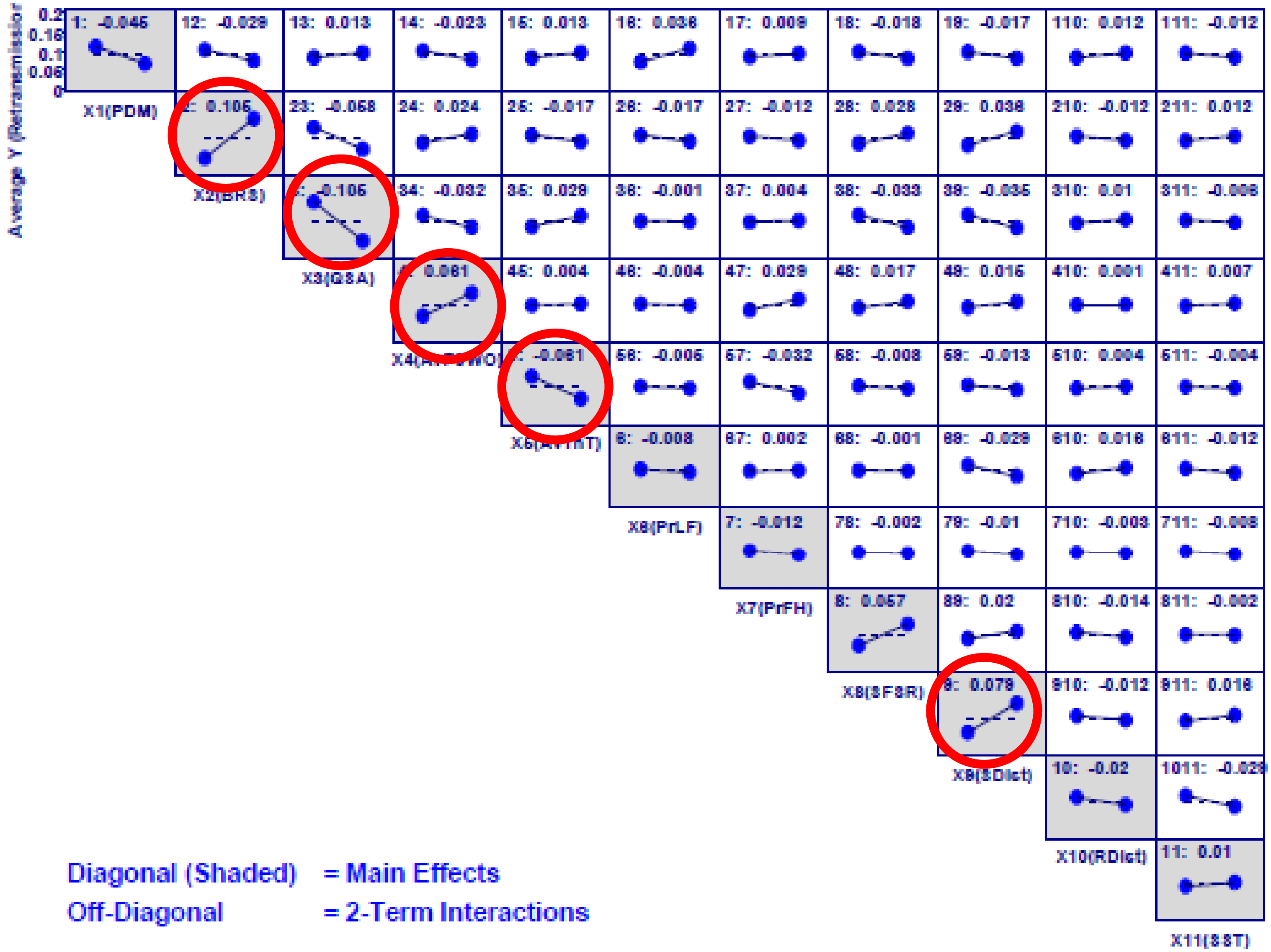
Y10: Retransmission Rate



Diagonal (Shaded) = Main Effects
Off-Diagonal = 2-Term Interactions

Interaction Effects Matrix

Y10: Retransmission Rate

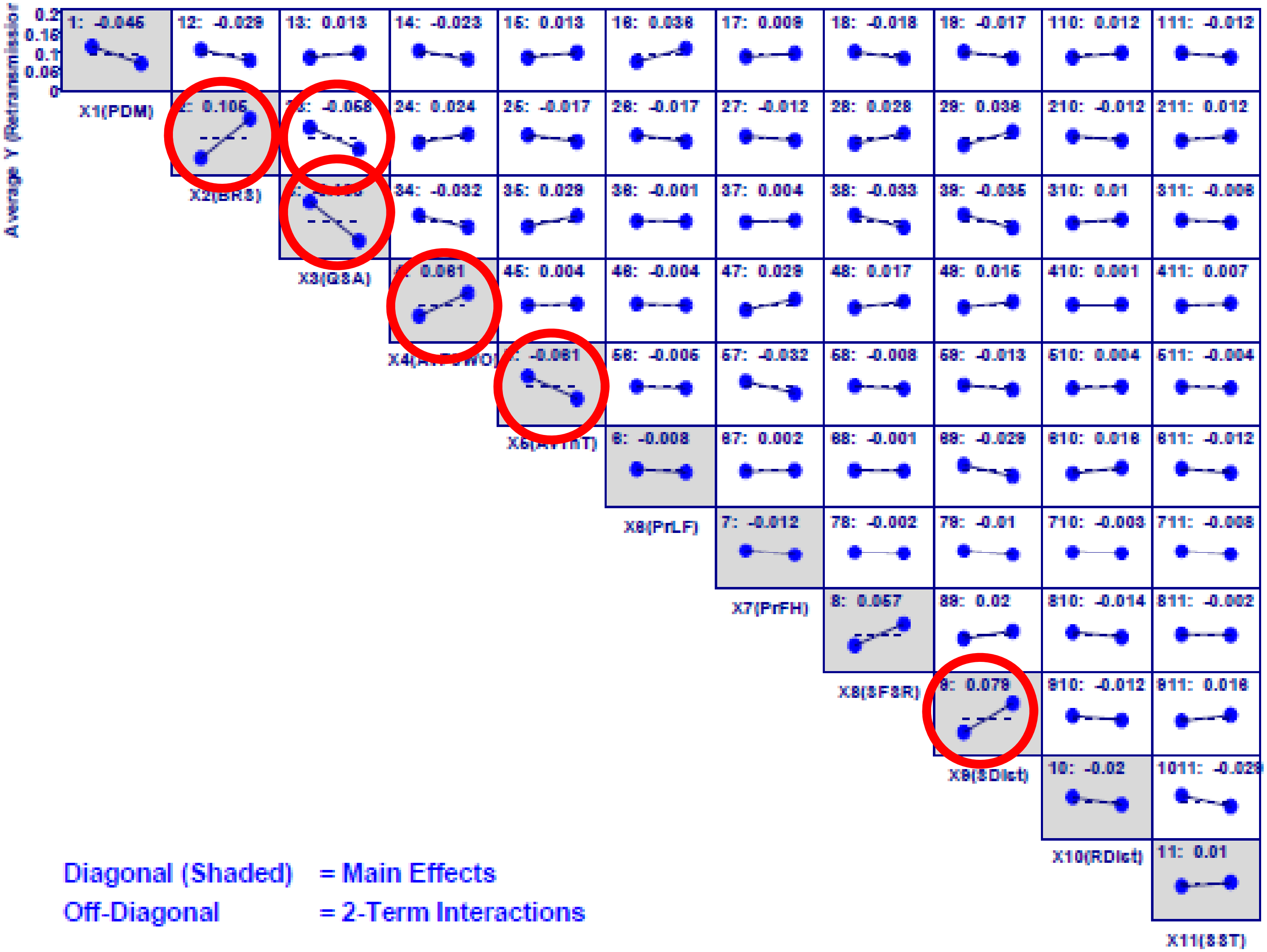


Diagonal (Shaded) = Main Effects

Off-Diagonal = 2-Term Interactions

Interaction Effects Matrix

Y10: Retransmission Rate



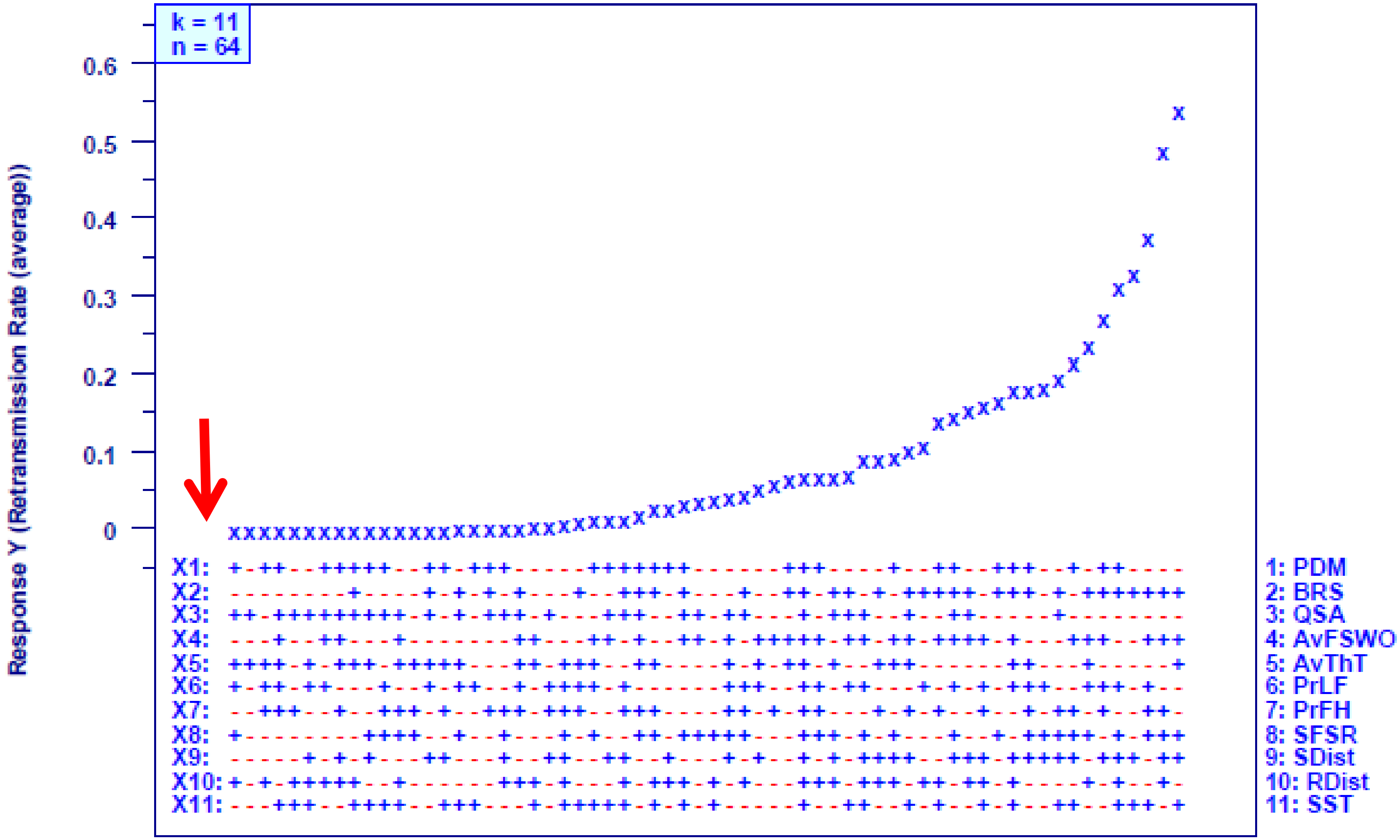
Diagonal (Shaded) = Main Effects

Off-Diagonal = 2-Term Interactions

http://www.itl.nist.gov/div898/handbook/pri/section5/pri59.htm

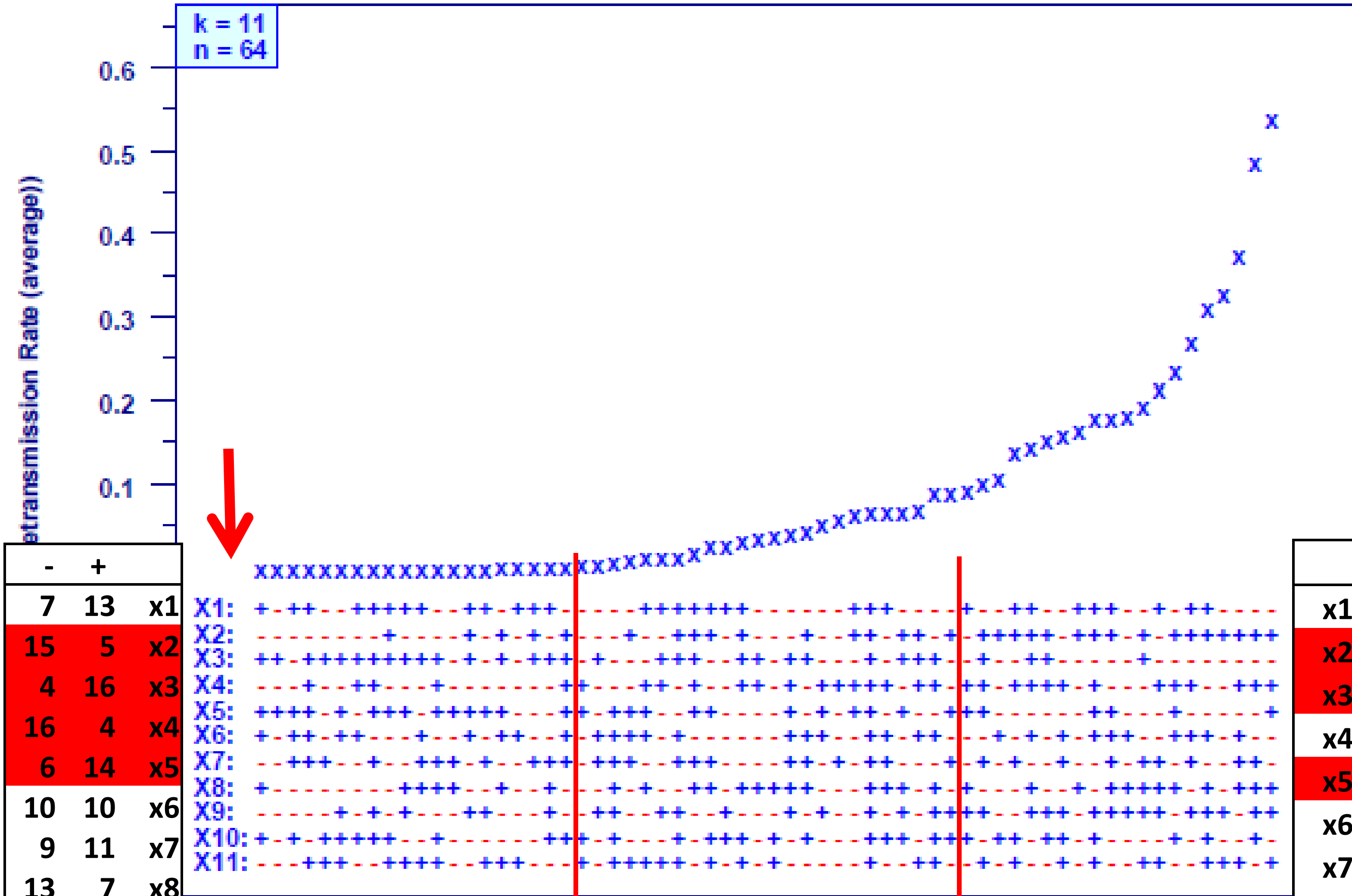
Ordered Data Plot

Y10: Retransmission Rate



Ordered Data Plot

Y10: Retransmission Rate

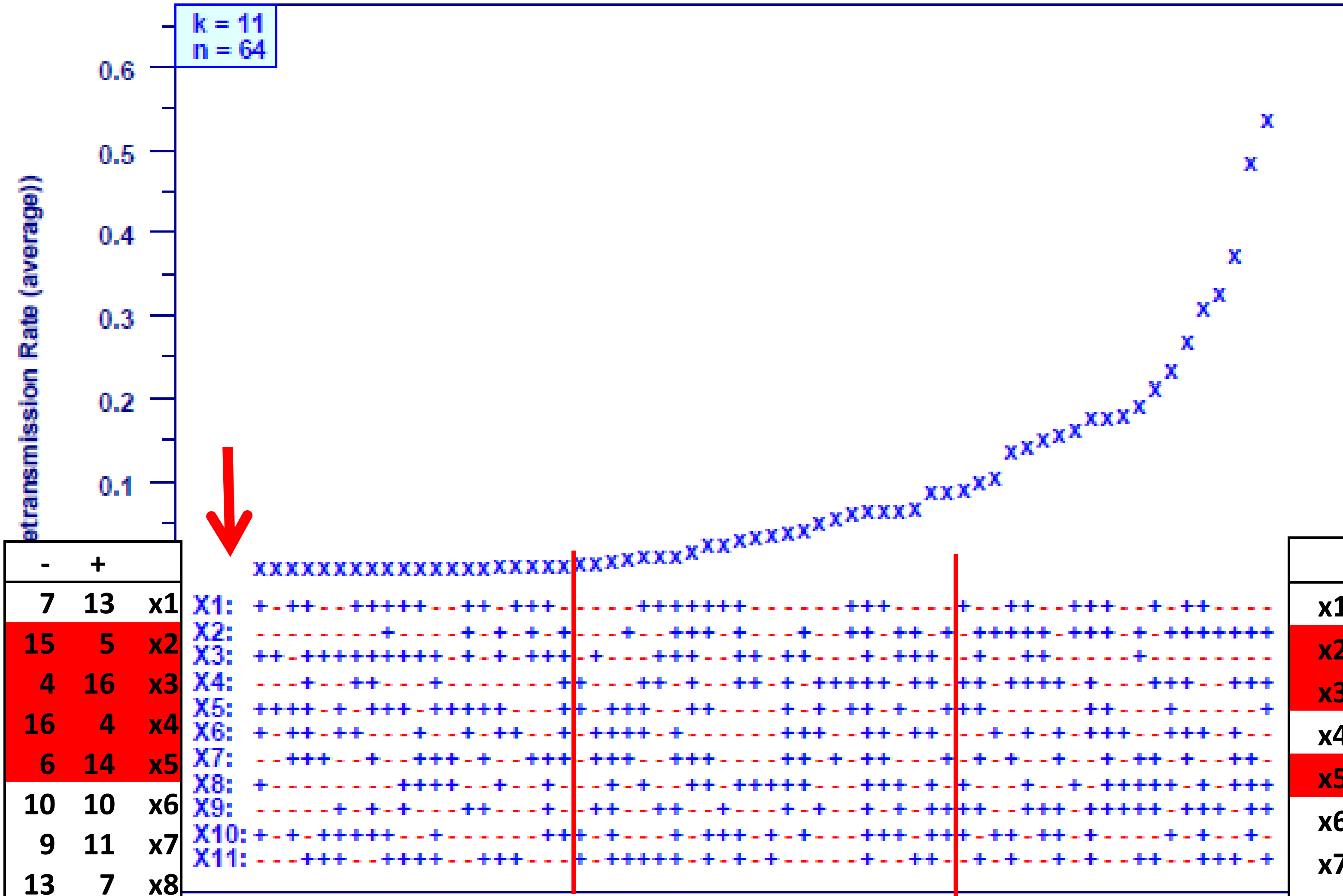


-	+	
7	13	x1
15	5	x2
4	16	x3
16	4	x4
6	14	x5
10	10	x6
9	11	x7
13	7	x8
14	6	x9
10	10	x10
10	10	x11

	-	+
x1	11	9
x2	4	16
x3	16	4
x4	7	13
x5	14	6
x6	10	10
x7	11	9
x8	8	12
x9	5	15
x10	11	9
x11	10	10

Ordered Data Plot

Y10: Retransmission Rate



-	+	
7	13	x1
15	5	x2
4	16	x3
16	4	x4
6	14	x5
10	10	x6
9	11	x7
13	7	x8
14	6	x9
10	10	x10
10	10	x11

	-	+
x1	11	9
x2	4	16
x3	16	4
x4	7	13
x5	14	6
x6	10	10
x7	11	9
x8	8	12
x9	5	15
x10	11	9
x11	10	10

Left:(+ - + - + . + - - . .)

Right:(+ - + - + . + - - + .)

Means:(+ - + - + + + - - + -)

Example 2: Y17 = Ave. TP for Active DD Flows

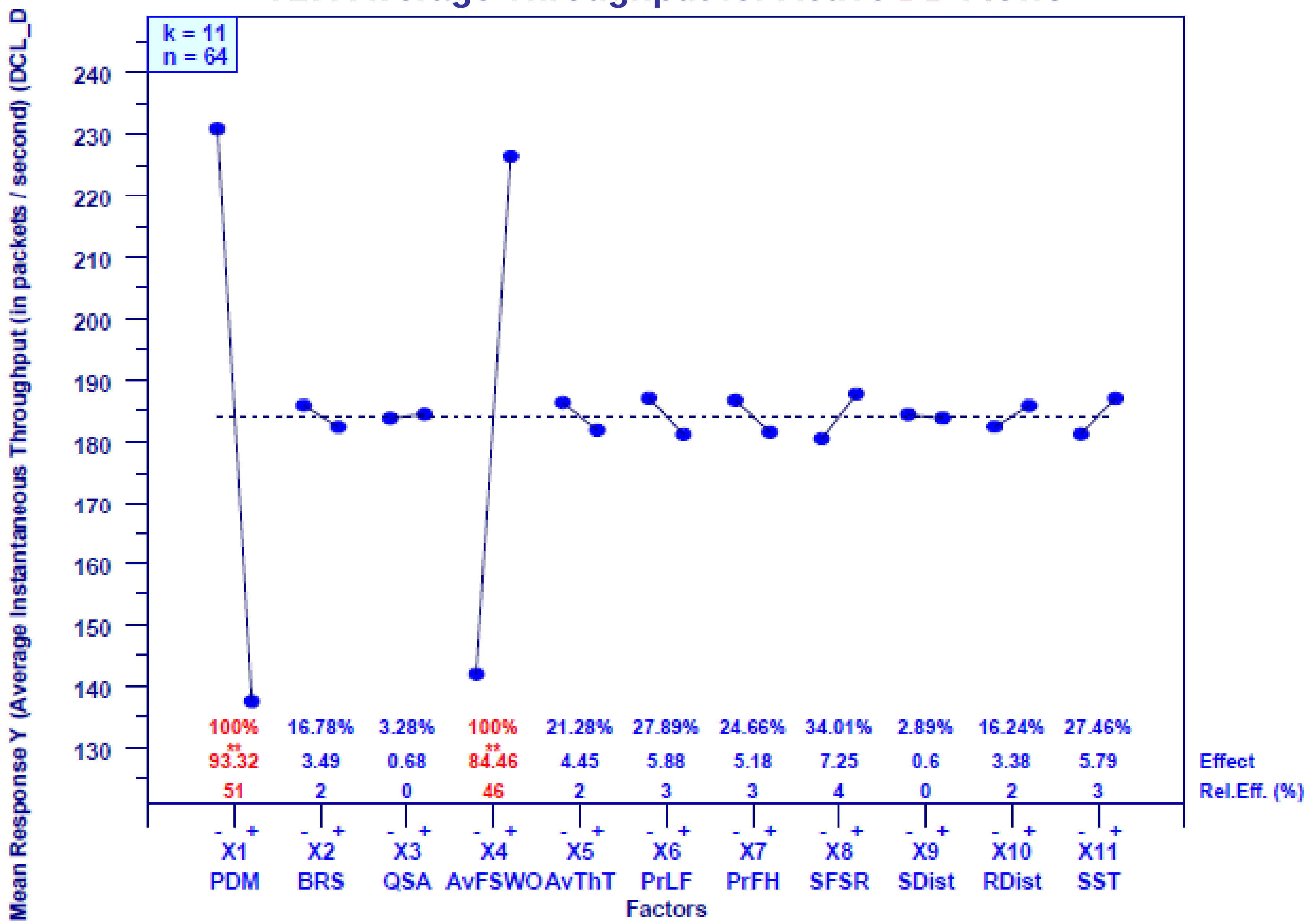
Response	Definition
y1	Active Flows – flows attempting to transfer data
y2	Proportion of potential flows that were active: Active Flows/All Sources
y3	Data packets entering the network per measurement interval
y4	Data packets leaving the network per measurement interval
y5	Loss Rate: y4/(y3+y4)
y6	Flows Completed per measurement interval
y7	Flow-Completion Rate: y6/(y6+y1)
y8	Connection Failures per measurement interval
y9	Connection-Failure Rate: y8/(y8+y1)
y10	Retransmission Rate (ratio)
y11	Congestion Window per Flow (packets)
y12	Window Increases per Flow per measurement interval
y13	Negative Acknowledgments per Flow per measurement interval
y14	Timeouts per Flow per measurement interval
y15	Smoothed Round-Trip Time (ms)
y16	Relative queuing delay: y15/(x1x41)

y17	Average Throughput for Active DD Flows
y18	Average Throughput for Active DF Flows
y19	Average Throughput for Active DN Flows
y20	Average Throughput for Active FF Flows
y21	Average Throughput for Active FN Flows
y22	Average Throughput for Active NN Flows

(k=11,n=64,m=22)

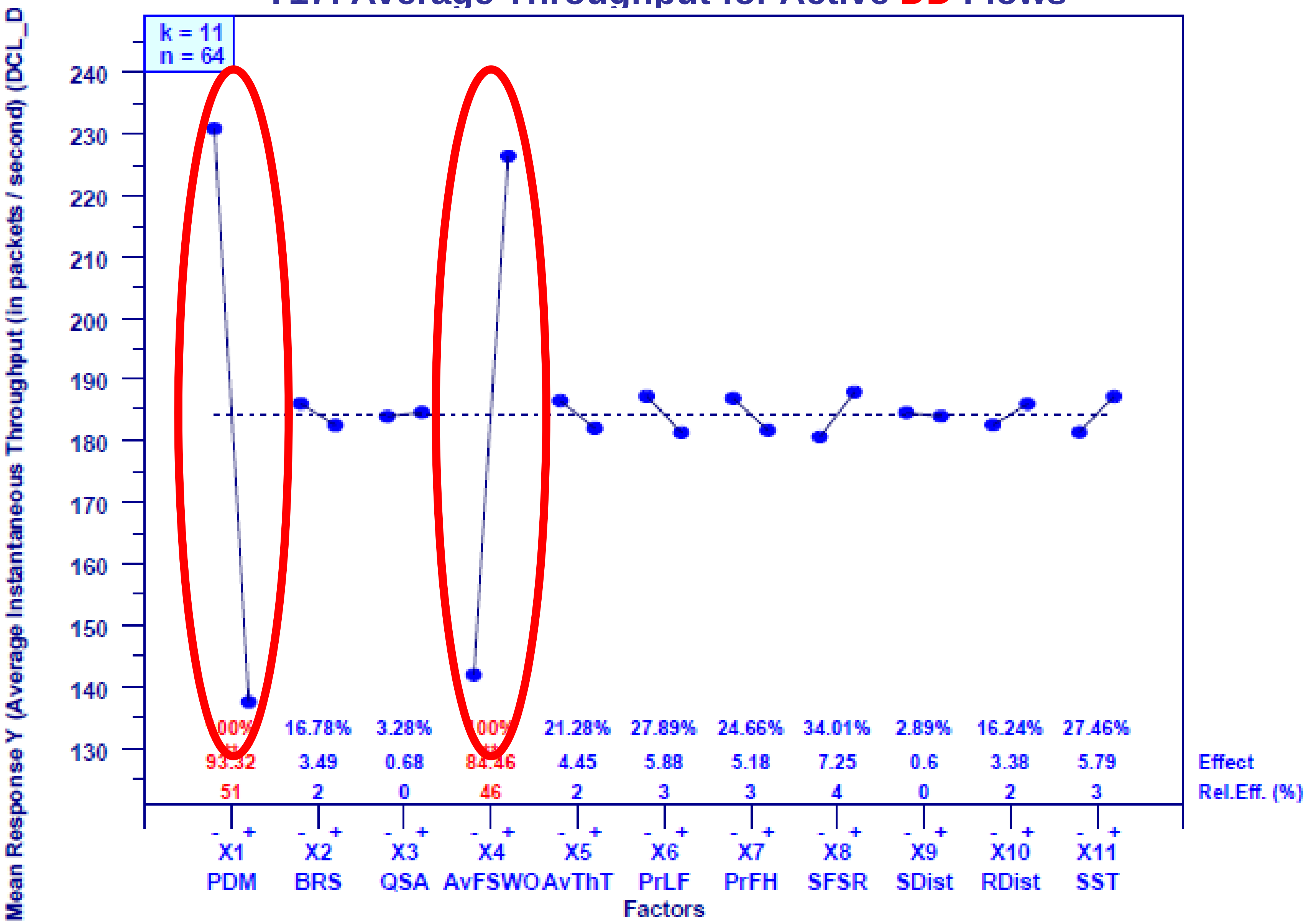
Main Effects Plot (Augmented)

Y17: Average Throughput for Active DD Flows



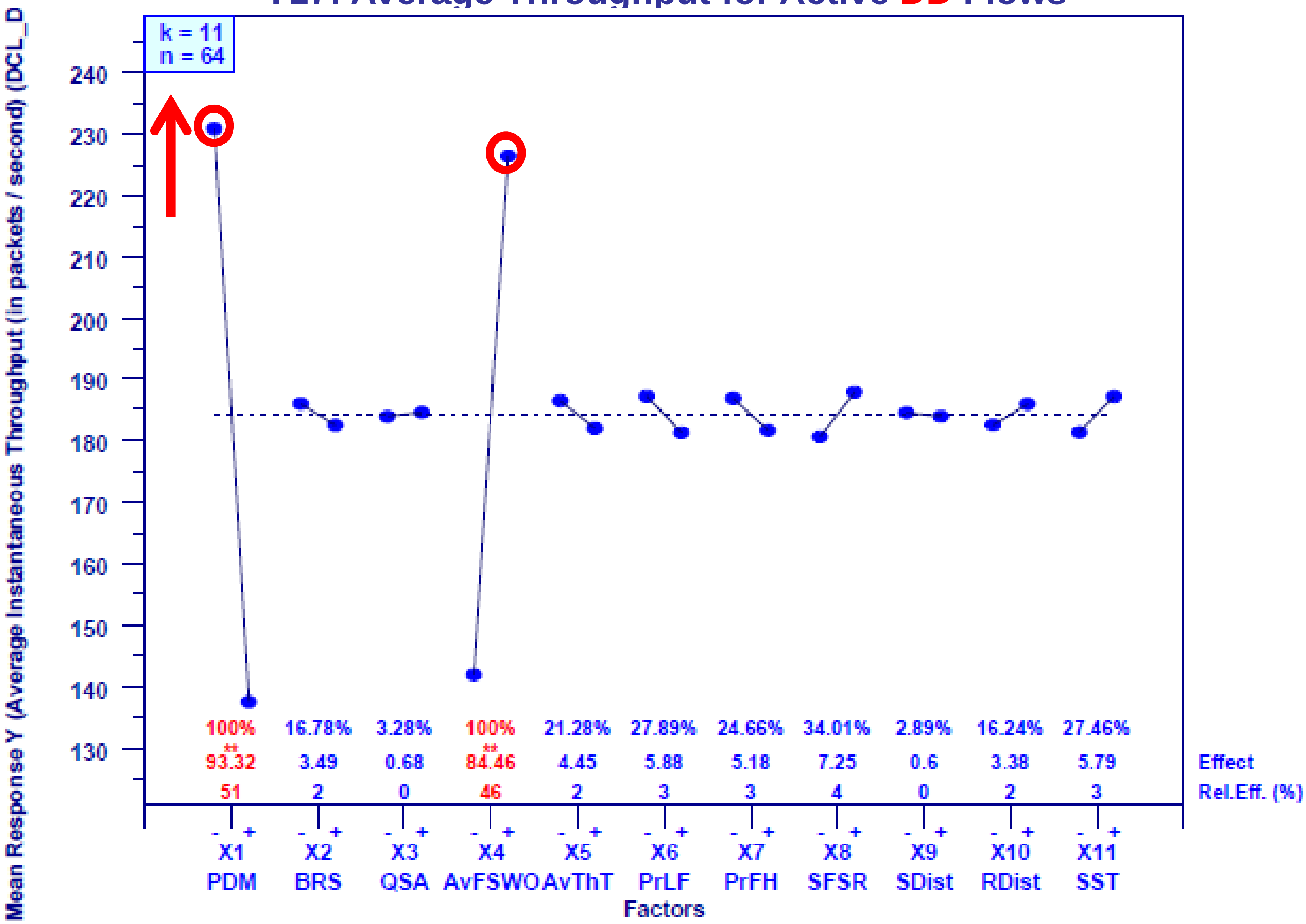
Main Effects Plot (Augmented)

Y17: Average Throughput for Active DD Flows



Main Effects Plot (Augmented)

Y17: Average Throughput for Active DD Flows



Means: (- - + + - - - + . + +)

Interaction Effects Matrix

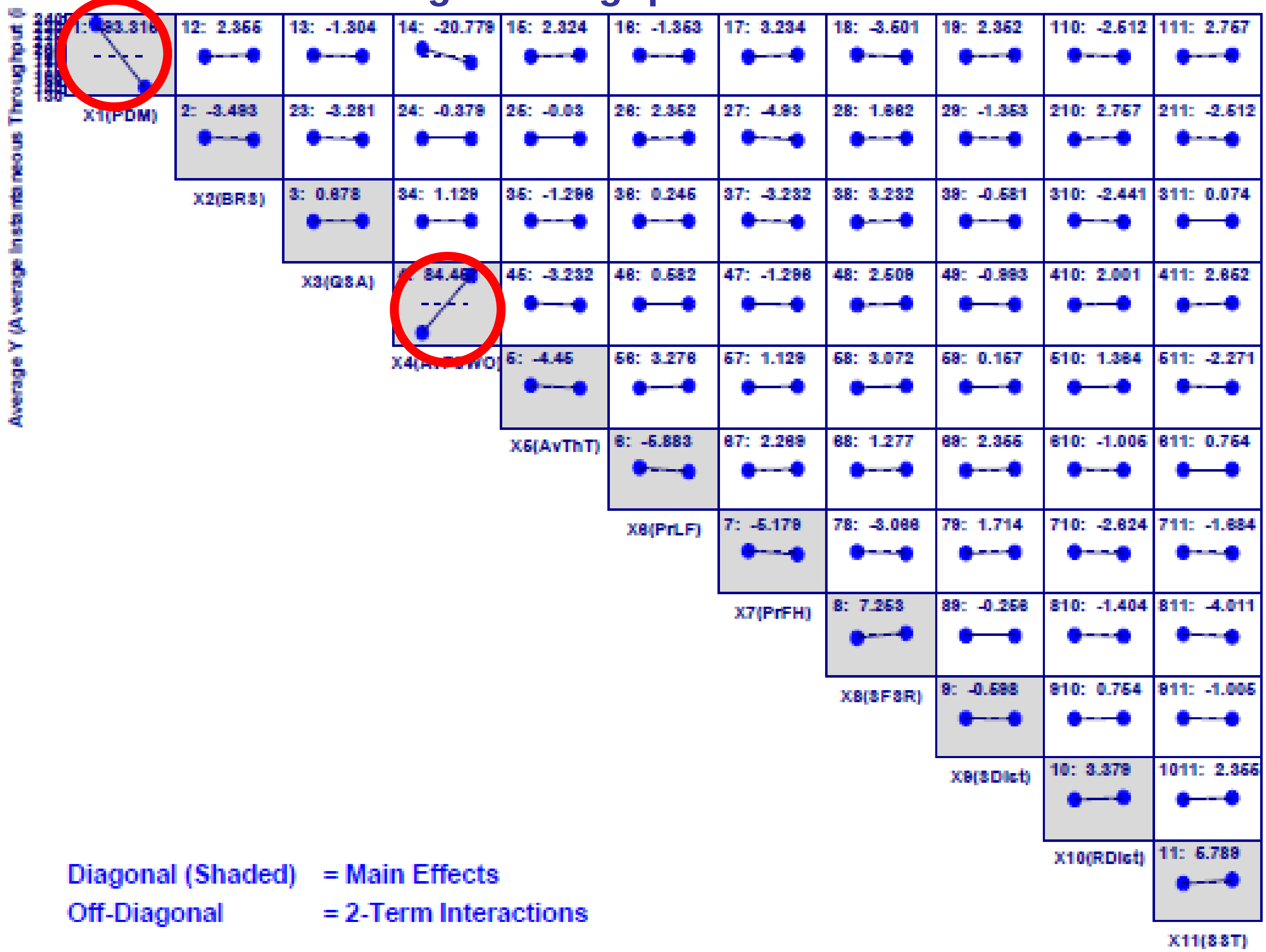
Y17: Average Throughput for Active DD Flows

Diagonal (Shaded) = Main Effects

Off-Diagonal = 2-Term Interactions

Interaction Effects Matrix

Y17: Average Throughput for Active DD Flows

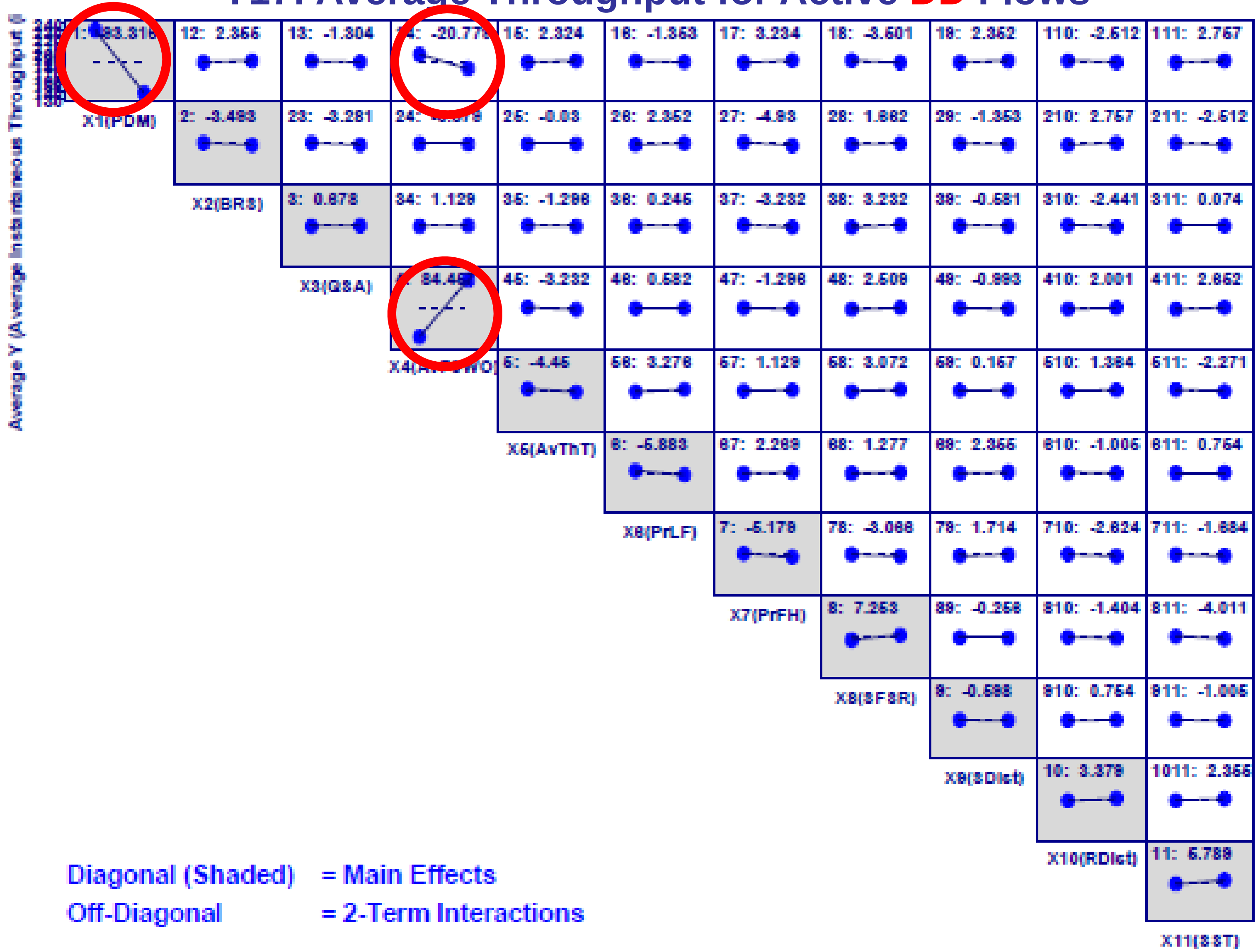


Diagonal (Shaded) = Main Effects

Off-Diagonal = 2-Term Interactions

Interaction Effects Matrix

Y17: Average Throughput for Active DD Flows

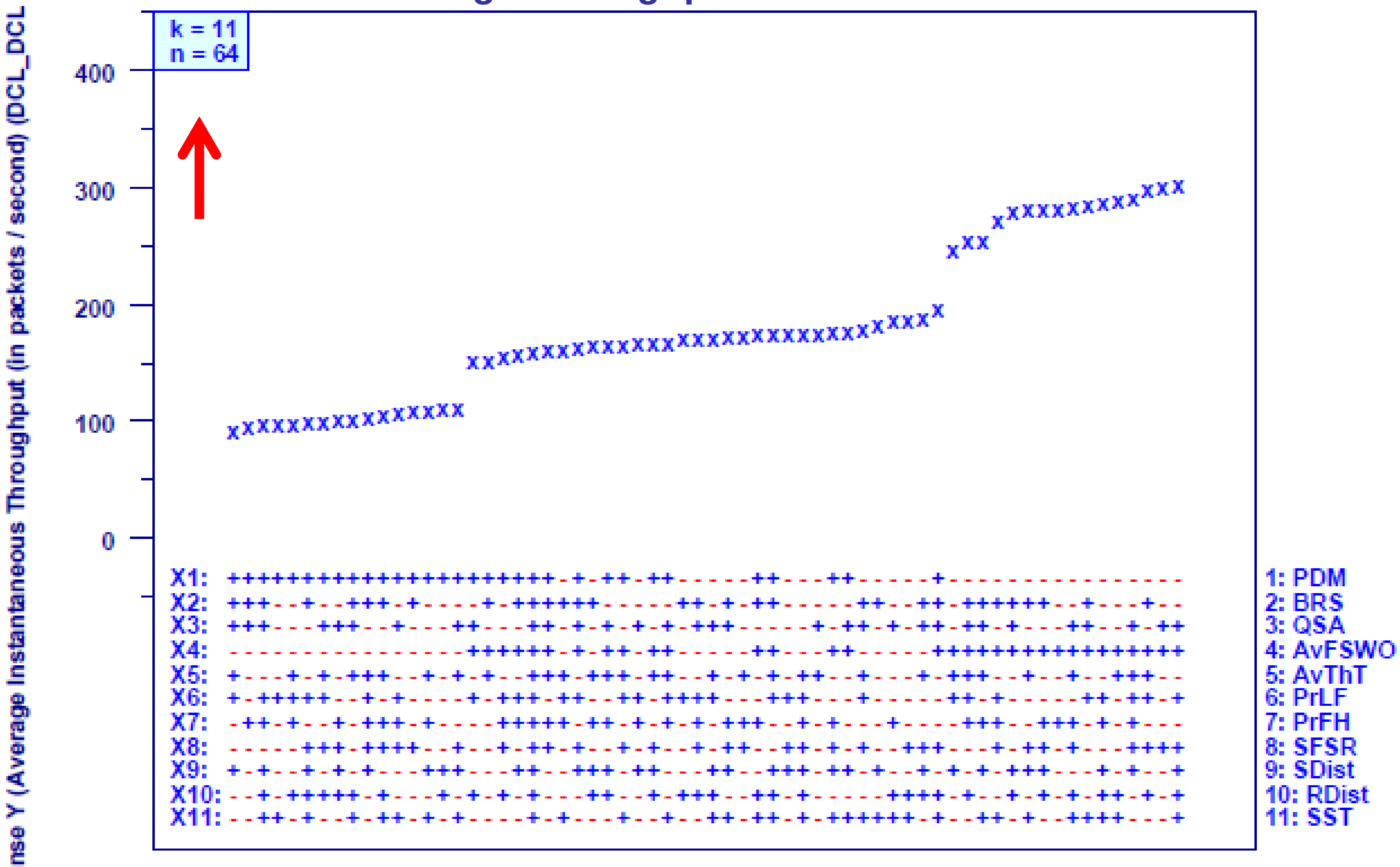


Diagonal (Shaded) = Main Effects

Off-Diagonal = 2-Term Interactions

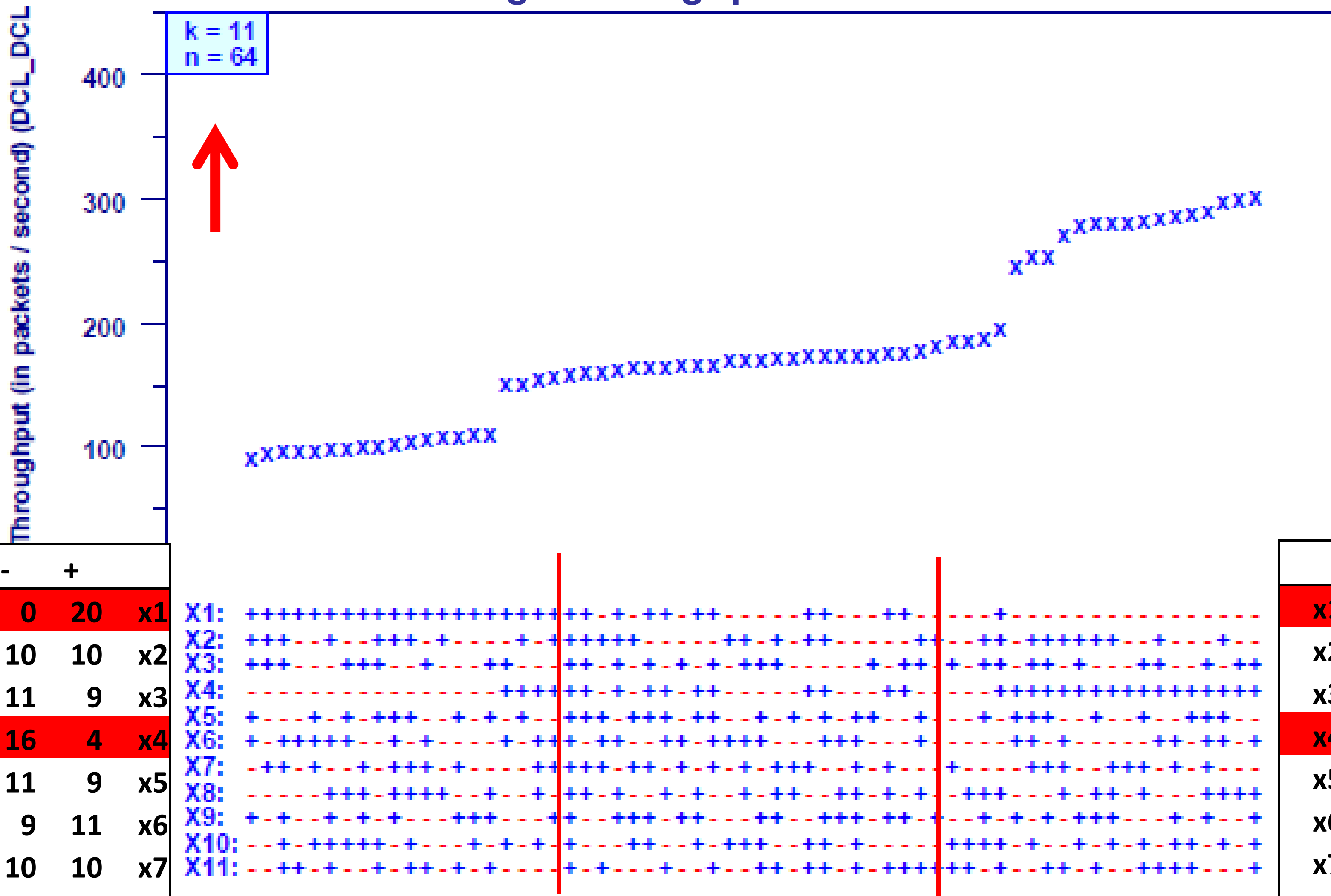
Ordered Data Plot

Y17: Average Throughput for Active DD Flows



Ordered Data Plot

Y17: Average Throughput for Active DD Flows

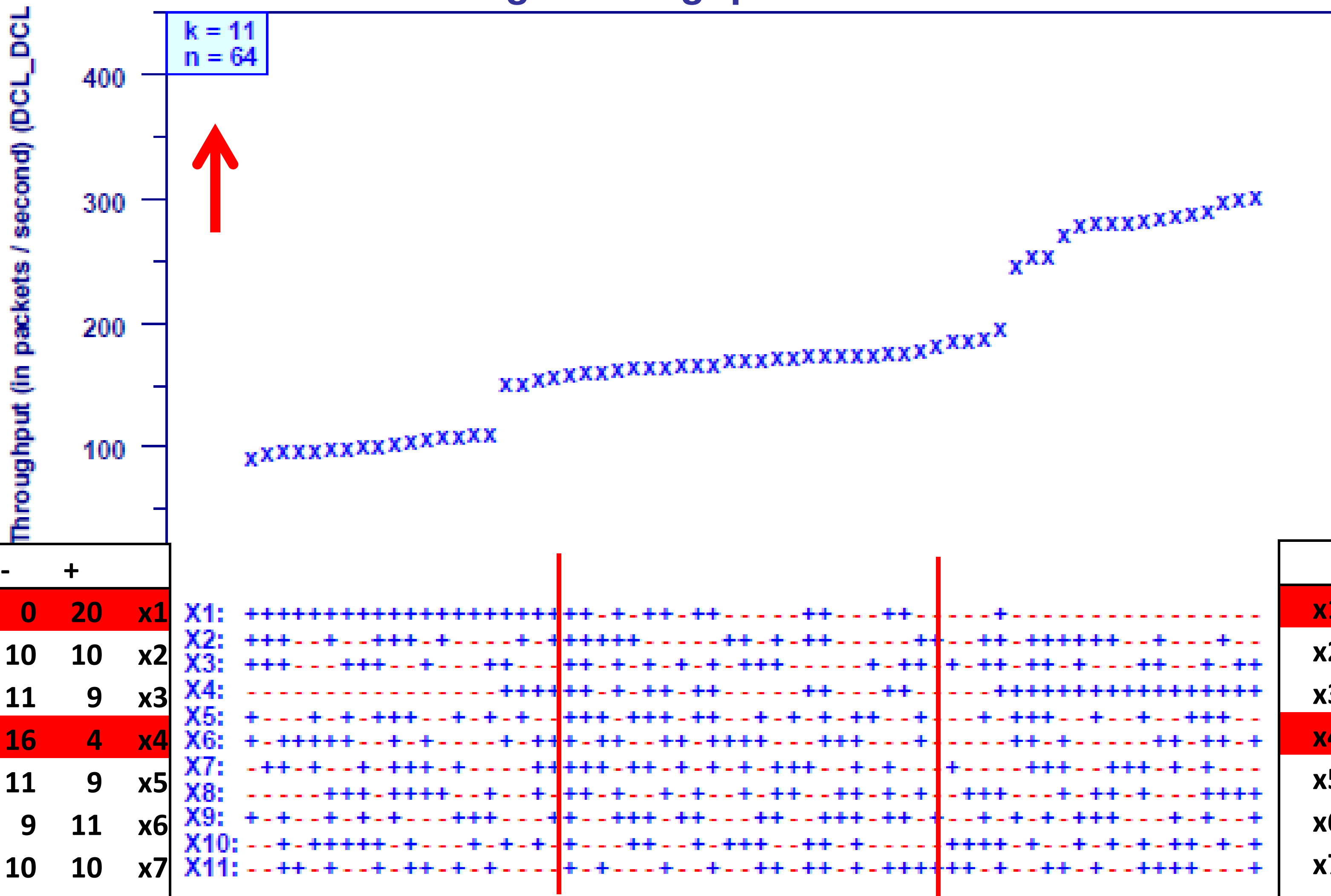


-	+	
0	20	x1
10	10	x2
11	9	x3
16	4	x4
11	9	x5
9	11	x6
10	10	x7
11	9	x8
11	9	x9
10	10	x10
12	8	x11

	-	+
x1	19	1
x2	10	10
x3	9	11
x4	3	17
x5	11	9
x6	12	8
x7	11	9
x8	9	11
x9	11	9
x10	8	12
x11	9	11

Ordered Data Plot

Y17: Average Throughput for Active DD Flows



-	+	
0	20	x1
10	10	x2
11	9	x3
16	4	x4
11	9	x5
9	11	x6
10	10	x7
11	9	x8
11	9	x9
10	10	x10
12	8	x11

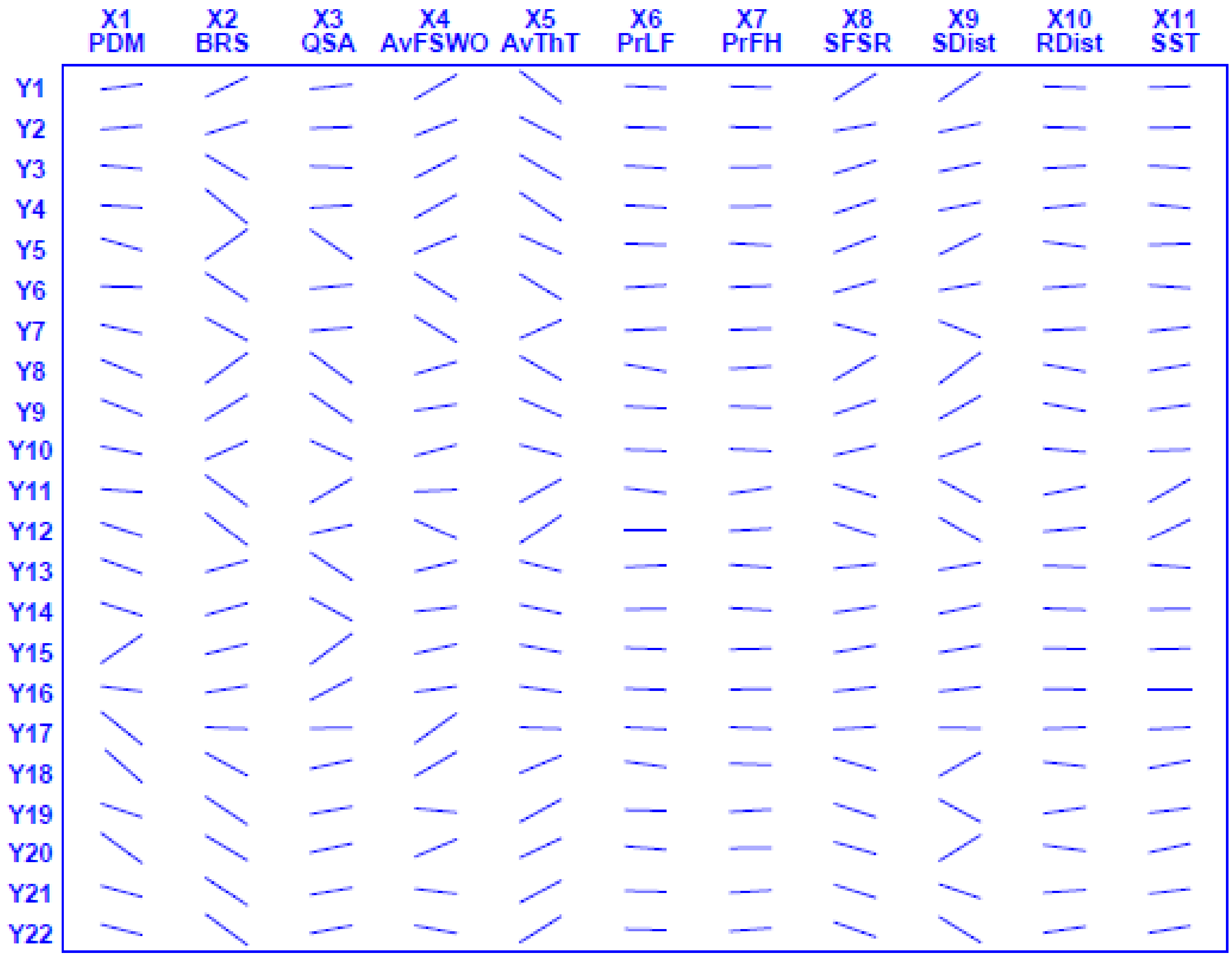
Left:(- . . + ... +)

Right:(- . . + . - . . . + .)

Means:(- - + + - - - + . + +)

	-	+
x1	19	1
x2	10	10
x3	9	11
x4	3	17
x5	11	9
x6	12	8
x7	11	9
x8	9	11
x9	11	9
x10	8	12
x11	9	11

Robustness Assessment: Stacked Main Effects Plot



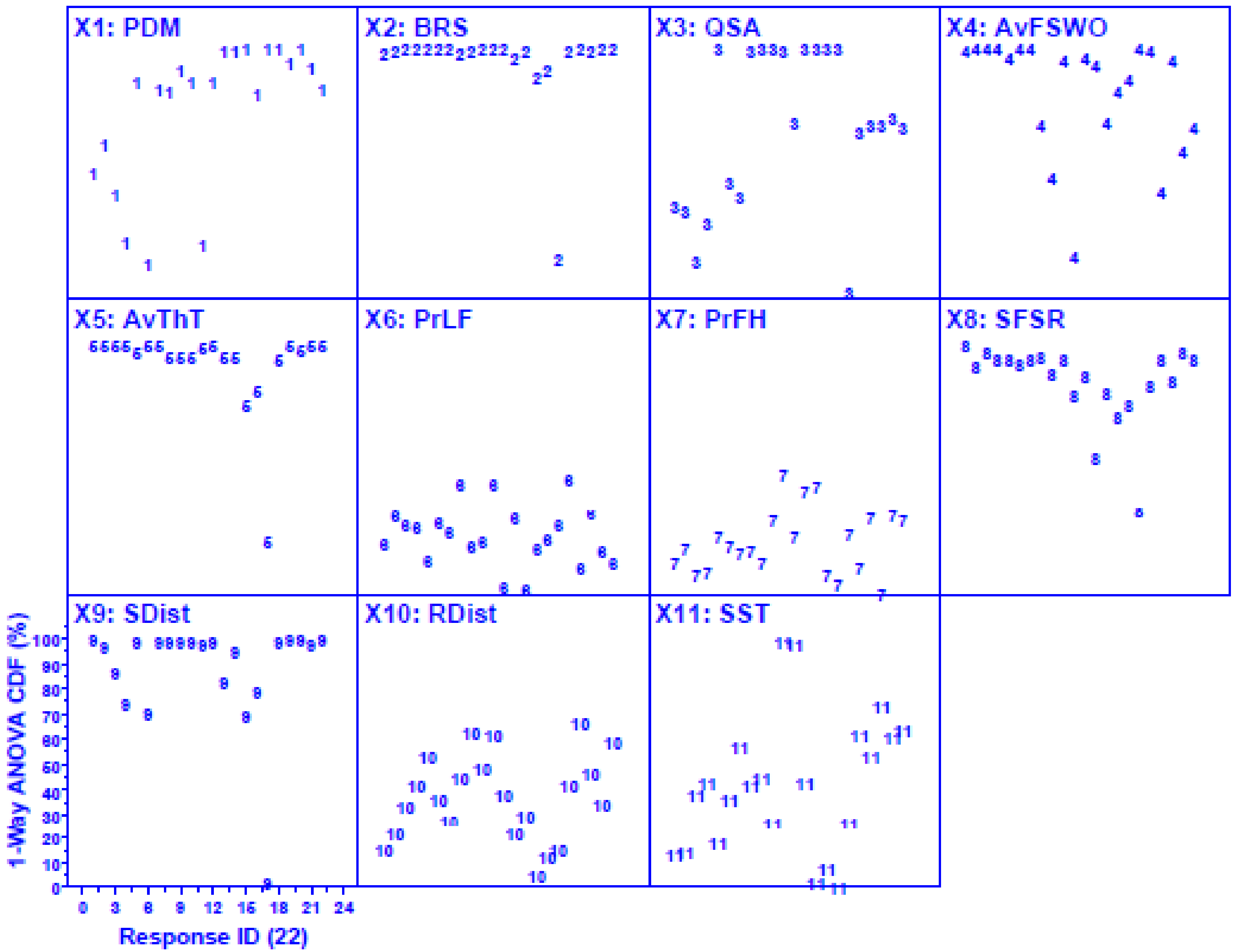
Robustness Assessment: (1-Way) ANOVA CDF Values (unordered)

	X1 PDM	X2 BRS	X3 QSA	X4 AvFSW	X5 AVThT	X6 PrLF	X7 PrFH	X8 SFSR	X9 SDist	X10 RDist	X11 SST
Y1	51.16	98.66	37.27	99.83	100	20.6	13.47	99.93	99.98	15.78	14.39
Y2	62.1	99.84	35.47	99.99	100	31.51	18.87	91.42	96.85	22.06	14.69
Y3	42.48	100	15.96	99.97	100	28.33	7.84	97.49	87.44	33.63	38.25
Y4	23.87	100	30.78	99.88	99.99	27.37	9.48	94.56	74.98	42.21	42.83
Y5	86.91	99.99	99.98	96.28	97.66	13.56	23.51	94.83	98.87	54.01	18.52
Y6	14.99	100	47.49	99.99	99.99	29.52	19.27	92.44	71.04	36.43	36.74
Y7	84.55	99.99	41.43	100	99.82	24.79	16.99	94.31	99.37	27.9	57.22
Y8	83.44	98.98	99.06	70.34	95.79	45.54	18.13	95.79	99.25	44.83	42.18
Y9	91.84	99.57	99.89	49.3	95.69	20.05	13.19	88.83	99.21	62.88	45.21
Y10	86.67	99.97	99.97	95.67	95.87	21.7	29.64	94.35	99.21	48.61	26.76
Y11	22.45	99.94	99.09	17.5	98.91	45.27	48.81	80.41	98.37	62.46	98.93
Y12	87.12	99.99	71.44	96.85	99.91	3.49	23.44	87.87	99.4	38.95	98.02
Y13	99.47	96.76	100	93.93	95.28	31.3	42.08	55.53	83.6	22.11	43.18
Y14	99.68	99.32	100	70.85	95.1	2.42	44.48	81.68	95.31	30.49	2.75
Y15	100	88.52	100	83.64	76	18.17	8.34	71.77	69.49	5.28	8.59
Y16	81.89	91.56	100	87.83	82.66	22.07	4.41	76.31	79.34	13.34	0.82
Y17	100	16.78	3.28	100	21.28	27.89	24.66	34.01	2.89	16.24	27.46
Y18	100	99.09	67.06	99.45	94.98	47.51	11.33	84.16	99.41	42.36	62.51
Y19	95.05	100	70.38	43.16	99.94	10.71	30.51	95.02	99.94	66.59	53.33
Y20	99.98	99.71	70.05	95.48	98.11	33.15	0.96	85.65	99.85	47.06	73.11
Y21	93	100	73.21	59.53	99.98	17.79	32.17	97.03	98.34	34.56	61.62
Y22	83.79	100	69.13	69.1	99.96	12.49	30.32	95.01	99.95	59.86	63.94
Sum	7	19	9	13	18	0	0	11	15	0	2

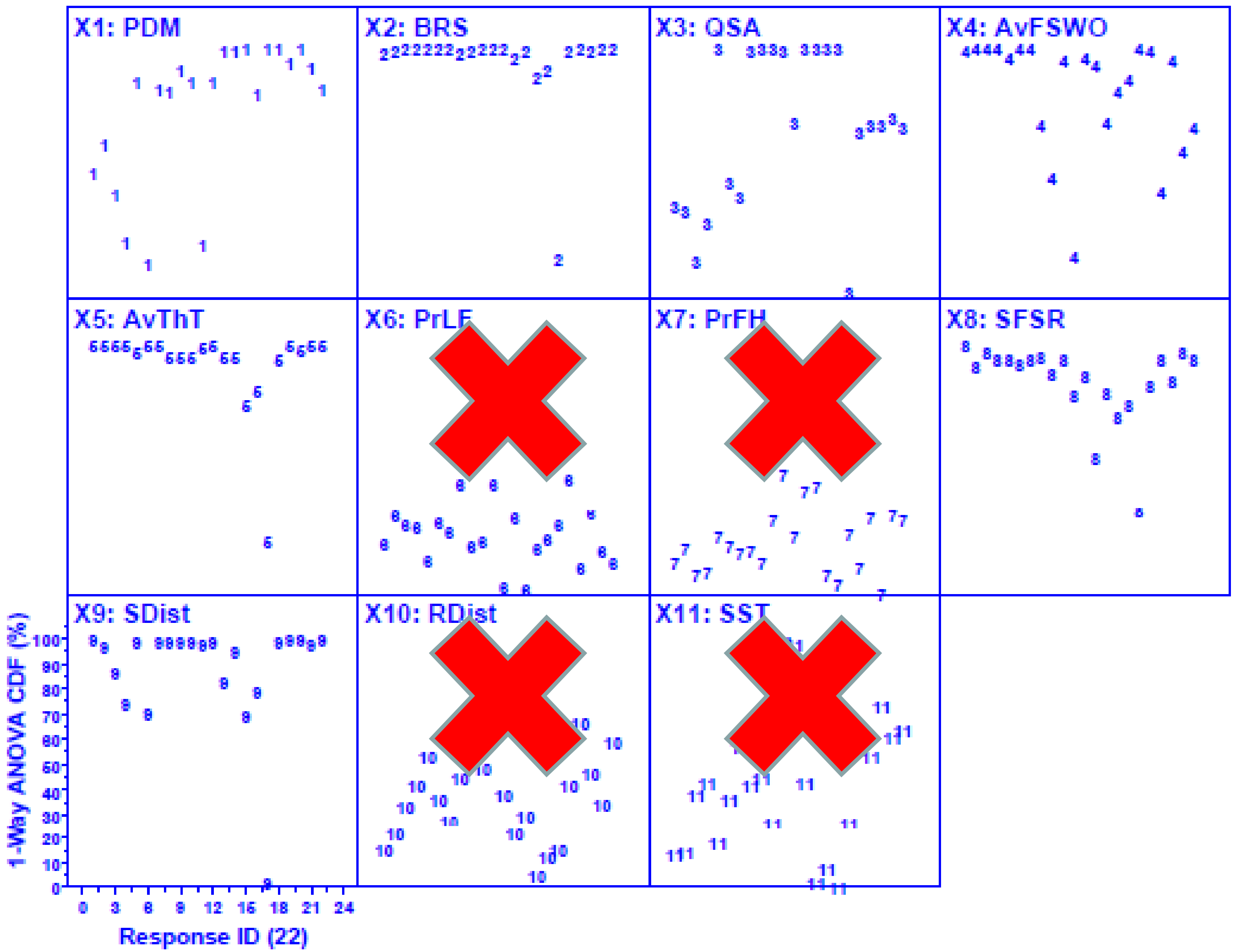
Robustness Assessment: (1-Way) ANOVA CDF Values (ordered)

	X2 BRS	X5 AVThT	X9 SDist	X4 AvFSW	X8 SFSR	X3 QSA	X1 PDM	X11 SST	X10 RDist	X7 PrFH	X6 PrLF
Y1	98.66	100	99.98	99.83	99.93	37.27	51.16	14.39	15.78	13.47	20.6
Y2	99.84	100	96.85	99.99	91.42	35.47	62.1	14.69	22.06	18.87	31.51
Y3	100	100	87.44	99.97	97.49	15.96	42.48	38.25	33.63	7.84	28.33
Y4	100	99.99	74.98	99.88	94.56	30.78	23.87	42.83	42.21	9.48	27.37
Y5	99.99	97.66	98.87	96.28	94.83	99.98	86.91	18.52	54.01	23.51	13.56
Y6	100	99.99	71.04	99.99	92.44	47.49	14.99	36.74	36.43	19.27	29.52
Y7	99.99	99.82	99.37	100	94.31	41.43	84.55	57.22	27.9	16.99	24.79
Y8	98.98	95.79	99.25	70.34	95.79	99.06	83.44	42.18	44.83	18.13	45.54
Y9	99.57	95.69	99.21	49.3	88.83	99.89	91.84	45.21	62.88	13.19	20.05
Y10	99.97	95.87	99.21	95.67	94.35	99.97	86.67	26.76	48.61	29.64	21.7
Y11	99.94	98.91	98.37	17.5	80.41	99.09	22.45	98.93	62.46	48.81	45.27
Y12	99.99	99.91	99.4	96.85	87.87	71.44	87.12	98.02	38.95	23.44	3.49
Y13	96.76	95.28	83.6	93.93	55.53	100	99.47	43.18	22.11	42.08	31.3
Y14	99.32	95.1	95.31	70.85	81.68	100	99.68	2.75	30.49	44.48	2.42
Y15	88.52	76	69.49	83.64	71.77	100	100	8.59	5.28	8.34	18.17
Y16	91.56	82.66	79.34	87.83	76.31	100	81.89	0.82	13.34	4.41	22.07
Y17	16.78	21.28	2.89	100	34.01	3.28	100	27.46	16.24	24.66	27.89
Y18	99.09	94.98	99.41	99.45	84.16	67.06	100	62.51	42.36	11.33	47.51
Y19	100	99.94	99.94	43.16	95.02	70.38	95.05	53.33	66.59	30.51	10.71
Y20	99.71	98.11	99.85	95.48	85.65	70.05	99.98	73.11	47.06	0.96	33.15
Y21	100	99.98	98.34	59.53	97.03	73.21	93	61.62	34.56	32.17	17.79
Y22	100	99.96	99.95	69.1	95.01	69.13	83.79	63.94	59.86	30.32	12.49
Sum	19	18	15	13	11	9	7	2	0	0	0

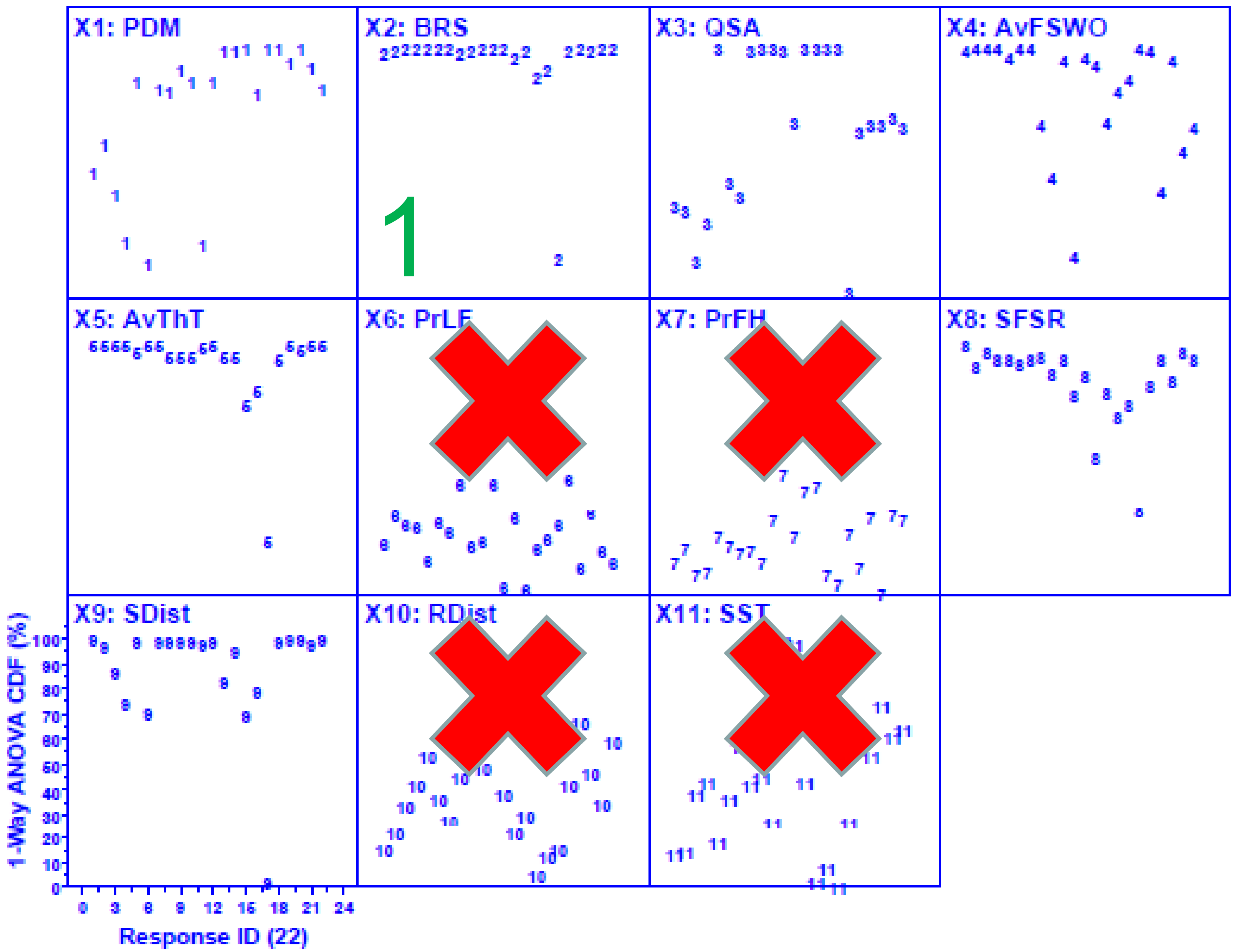
Robustness Assessment: Multiplot of (1-Way) ANOVA CDF Values



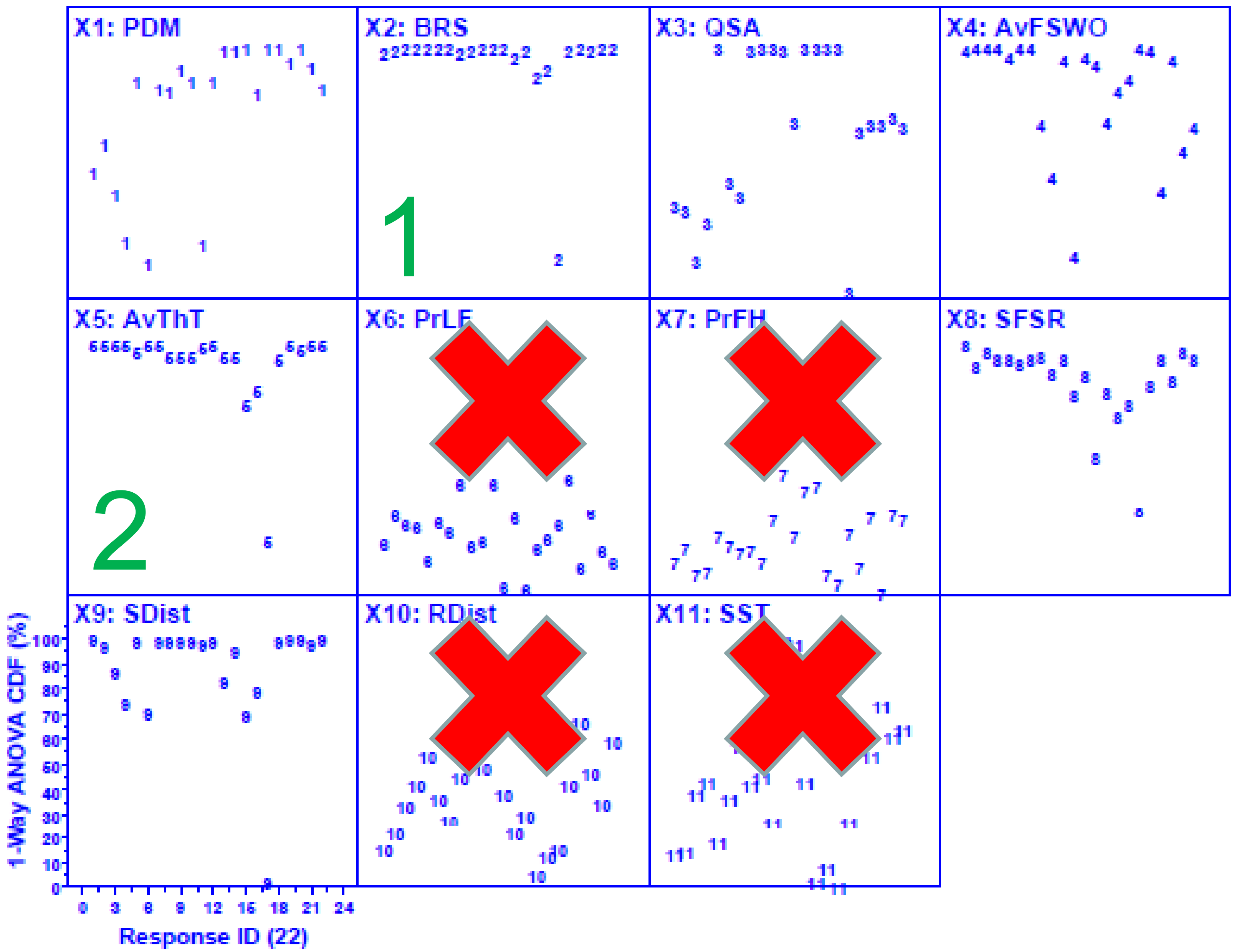
Robustness Assessment: Multiplot of (1-Way) ANOVA CDF Values



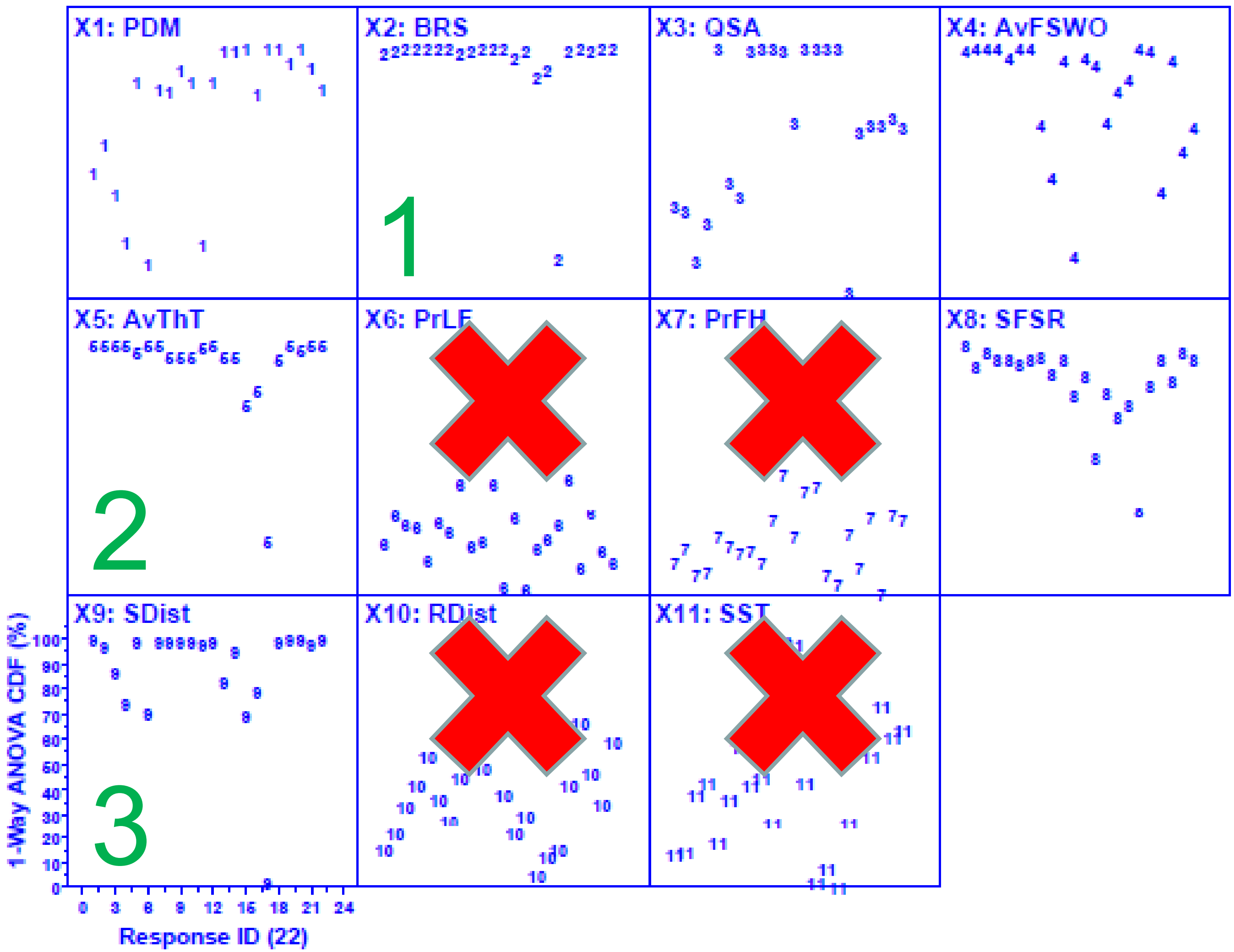
Robustness Assessment: Multiplot of (1-Way) ANOVA CDF Values



Robustness Assessment: Multiplot of (1-Way) ANOVA CDF Values



Robustness Assessment: Multiplot of (1-Way) ANOVA CDF Values



Robust Sensitivity Analysis Ranking (Criterion 1)

Major Factors (ordered) influencing MesoNet behavior:

- **X2: Network Speed**
- **X5: Think Time**
- **X9: Distribution of Sources**
- **X4: File Size**
- **X8: Number of Sources**

Minor Factor influencing MesoNet behavior:

- **X3: Buffer Size – small buffer sizes reduces delay variability & large buffer size has greater effect under high network speed**
- **X1: Propagation Delay**

Non-Factors

- **X11: Initial TCP Slow-Start Threshold**
- **X10: Distribution of Receivers**
- **X7: Probability a Source or Receiver is on a Fast Host**
- **X6: Probability a User Opts to Transfer a Larger File**

Robust Sensitivity Analysis Ranking (Criterion 2)

Major Factors (ordered) influencing MesoNet behavior:

- **X2: Network Speed**
- **X4: File Size**
- **X5: Think Time**
- **X8: Number of Sources**
- **X1: Propagation Delay**
- **X9: Distribution of Sources**

Minor Factor influencing MesoNet behavior:

- **X3: Buffer Size – small buffer sizes reduces delay variability & large buffer size has greater effect under high network speed**

Non-Factors

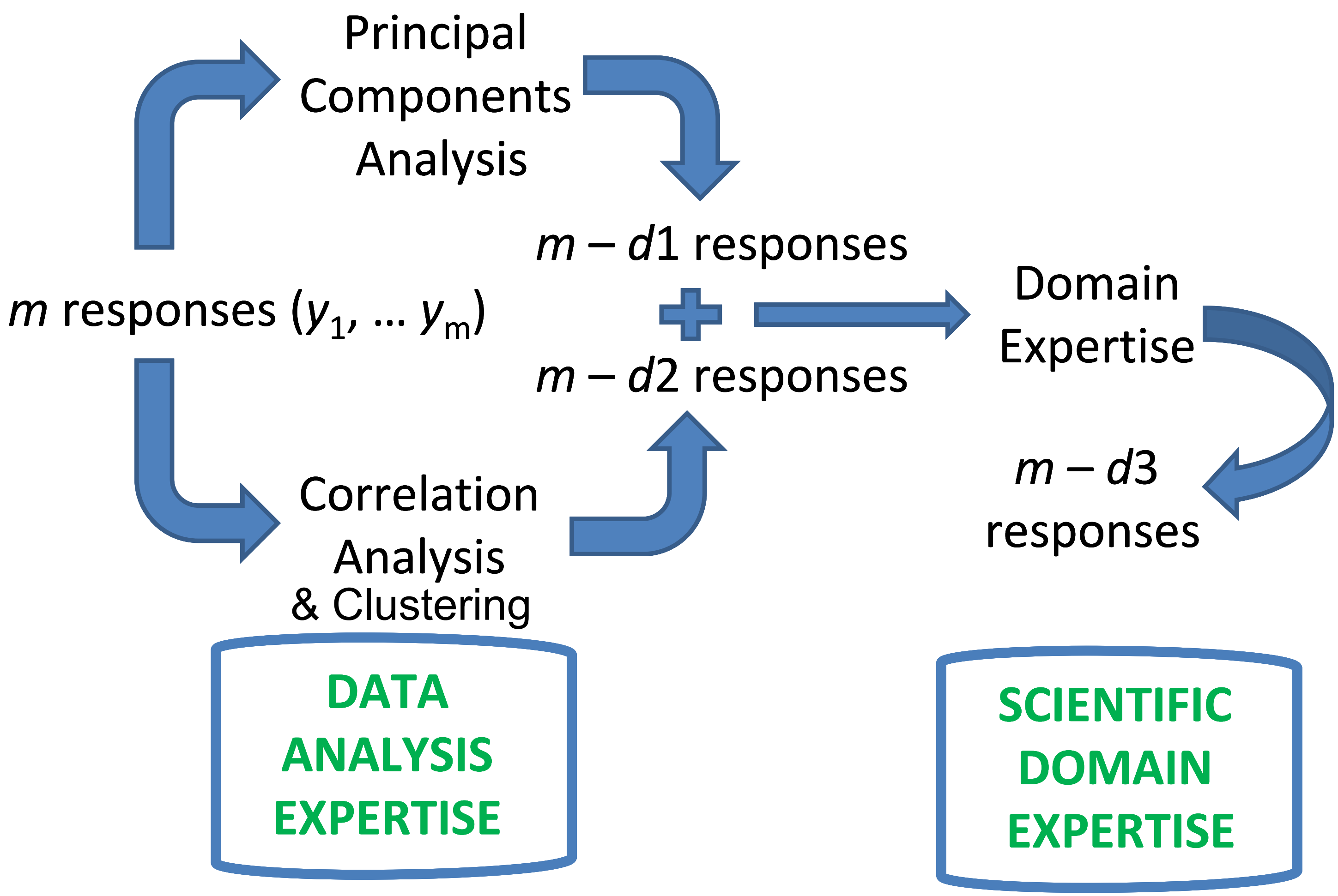
- **X11: Initial TCP Slow-Start Threshold**
- **X10: Distribution of Receivers**
- **X7: Probability a Source or Receiver is on a Fast Host**
- **X6: Probability a User Opts to Transfer a Larger File**

Robust Sensitivity Analysis Ranking (Criterion 2)

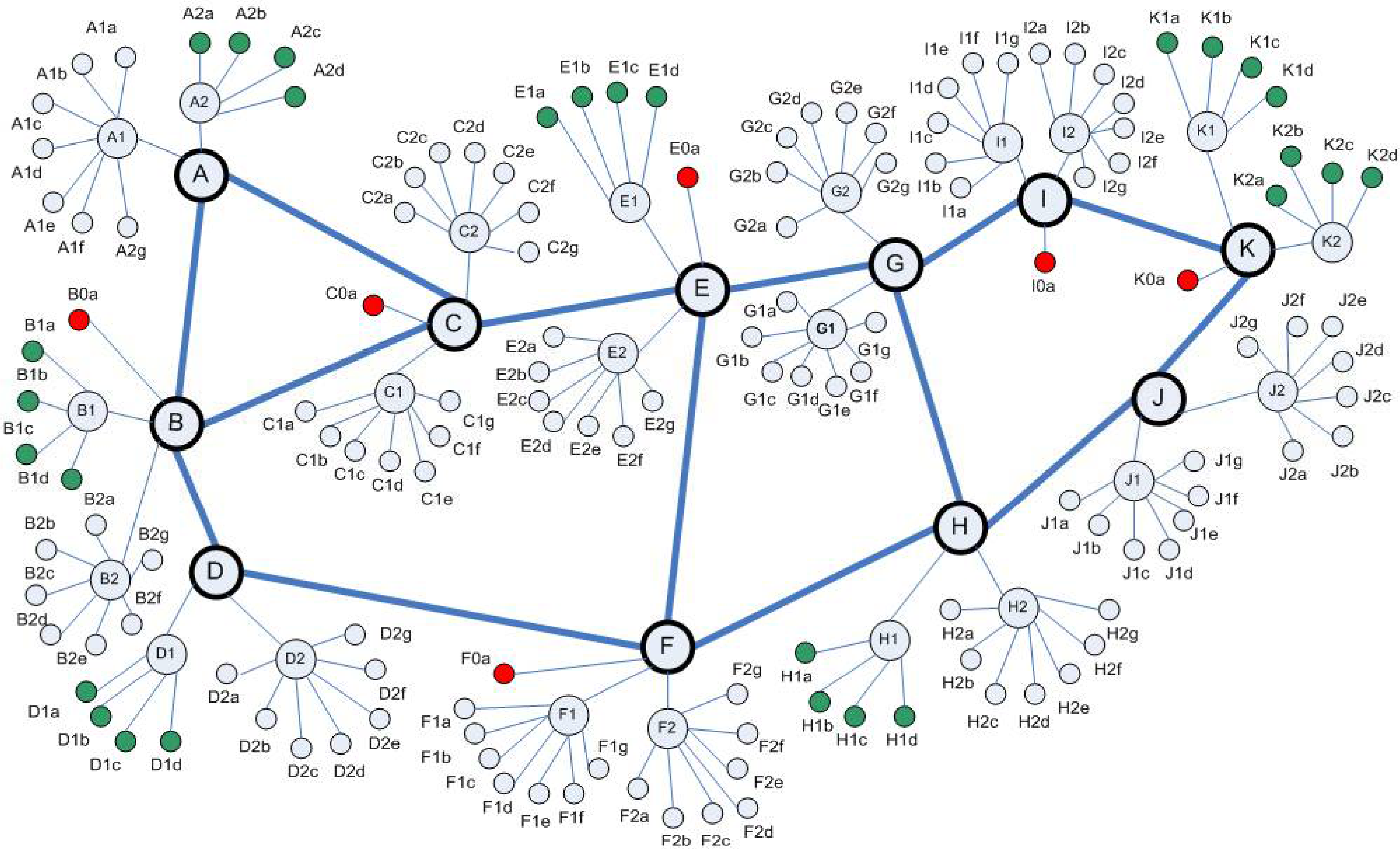
Category	Factor	Code	Definition	Level 1: -	Level 2: +
Network Factors	x1	PDM	Propagation delay	1	2
	x2	BRS (s)	Network speed	800 p/ms	400 p/ms
	x3	QSA	Buffer sizing	*RTT*x*C*/SQRT(*n*)	*RTT*x*C*
User Factors	x4	AvFSWO	Average file size for web pages	50 packets	100 packets
	x5	AvThT	Average think time between web clicks	2000 ms	5000 ms
	x6	PrLF	Probability a user opts to transfer a larger file	0.02	0.01
Source & Receiver Factors	x7	PrFH	Probability a source or receiver is on a fast host	0.4	0.2
	x8	SFSR	Scaling factor for number of sources & receivers	2	3
	x9	SDist	Distribution of sources	WEB	P2P
	x10	RDist	Distribution of receivers	WEB	P2P
Protocol Factors	x11	SST	Initial TCP slow-start threshold	43 packets	1.07×10^9 packets

Dimension Reduction Analysis

We Applied Two Different Techniques



Abilene Network (3-Tier MesoNet Topology)



22 Responses: 16 Macro + 6 Throughput

Response	Definition
y1	Active Flows – flows attempting to transfer data
y2	Proportion of potential flows that were active: Active Flows/All Sources
y3	Data packets entering the network per measurement interval
y4	Data packets leaving the network per measurement interval
y5	Loss Rate: y4/(y3+y4)
y6	Flows Completed per measurement interval
y7	Flow-Completion Rate: y6/(y6+y1)
y8	Connection Failures per measurement interval
y9	Connection-Failure Rate: y8/(y8+y1)
y10	Retransmission Rate (ratio)
y11	Congestion Window per Flow (packets)
y12	Window Increases per Flow per measurement interval
y13	Negative Acknowledgments per Flow per measurement interval
y14	Timeouts per Flow per measurement interval
y15	Smoothed Round-Trip Time (ms)
y16	Relative queuing delay: y15/(x1x41)

y17	Average Throughput for Active DD Flows
y18	Average Throughput for Active DF Flows
y19	Average Throughput for Active DN Flows
y20	Average Throughput for Active FF Flows
y21	Average Throughput for Active FN Flows
y22	Average Throughput for Active NN Flows

(k=11,n=64,m=22)

Data: 64 x 22 Multivariate Data Set Resulting from a 2^{11-5} Orthogonal Fractional Factorial Experiment Design

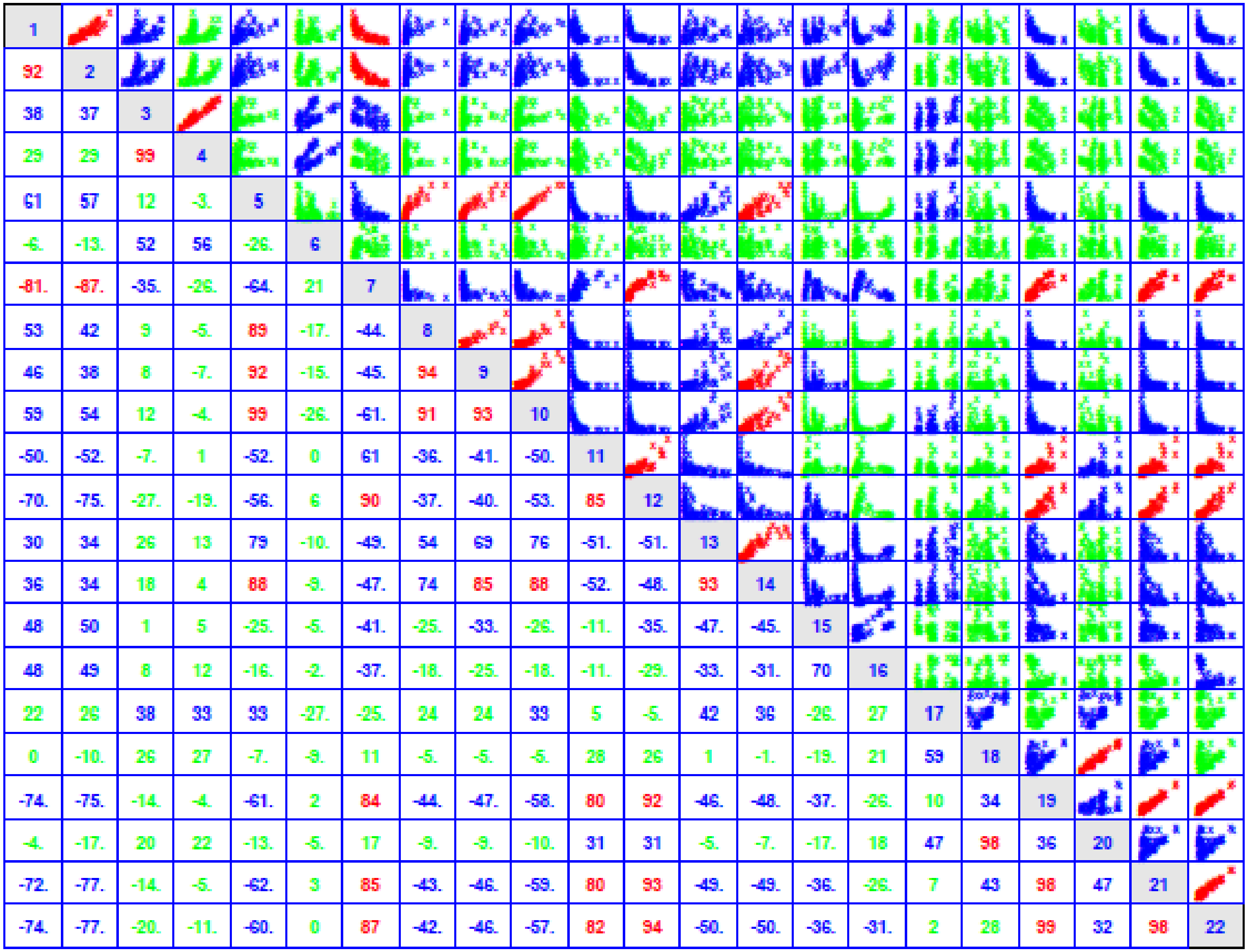
Run	y1	y2	…	y21	y22
1	4680.619	0.168126	…	92.034	89.785
2	6654.512	0.239371	…	72.596	57.738
3	9431.405	0.339259	…	29.569	13.963
4	11565.81	0.415439	…	23.427	19.882
…	…	…	…	…	…
61	10319.55	0.247471	…	87.969	41.573
62	1738.469	0.093668	…	159.298	161.602
63	1783.509	0.096094	…	148.395	161.36
64	21467.6	0.514811	…	26.159	9.981

Method 1: Correlation Analysis & Clustering

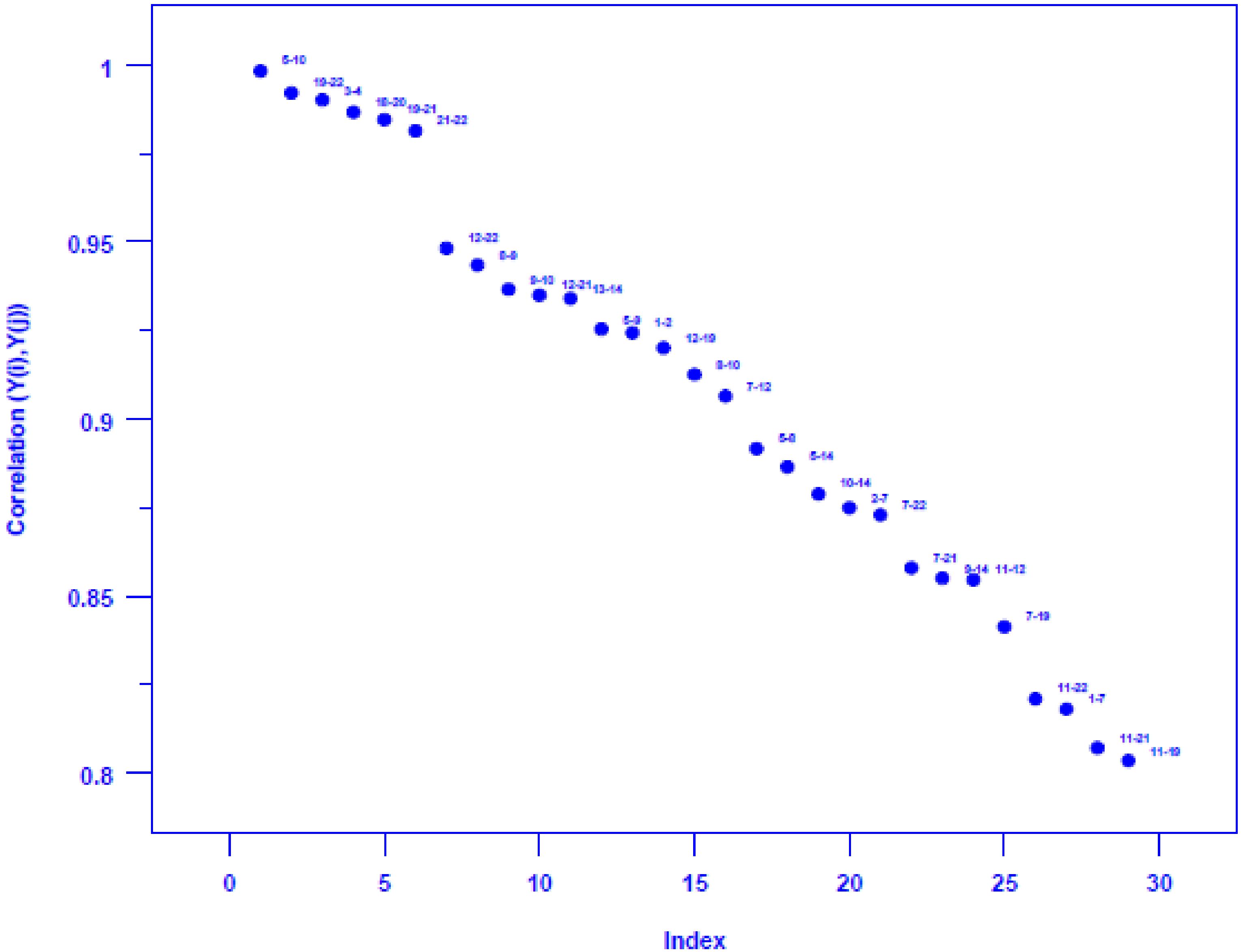
Abilene Network PCA Analysis (Standardized Responses) (Kevin Mills) Exp. 3 (2to(11-5))

Q. Are Any Pairs of the Raw Variables Correlated?

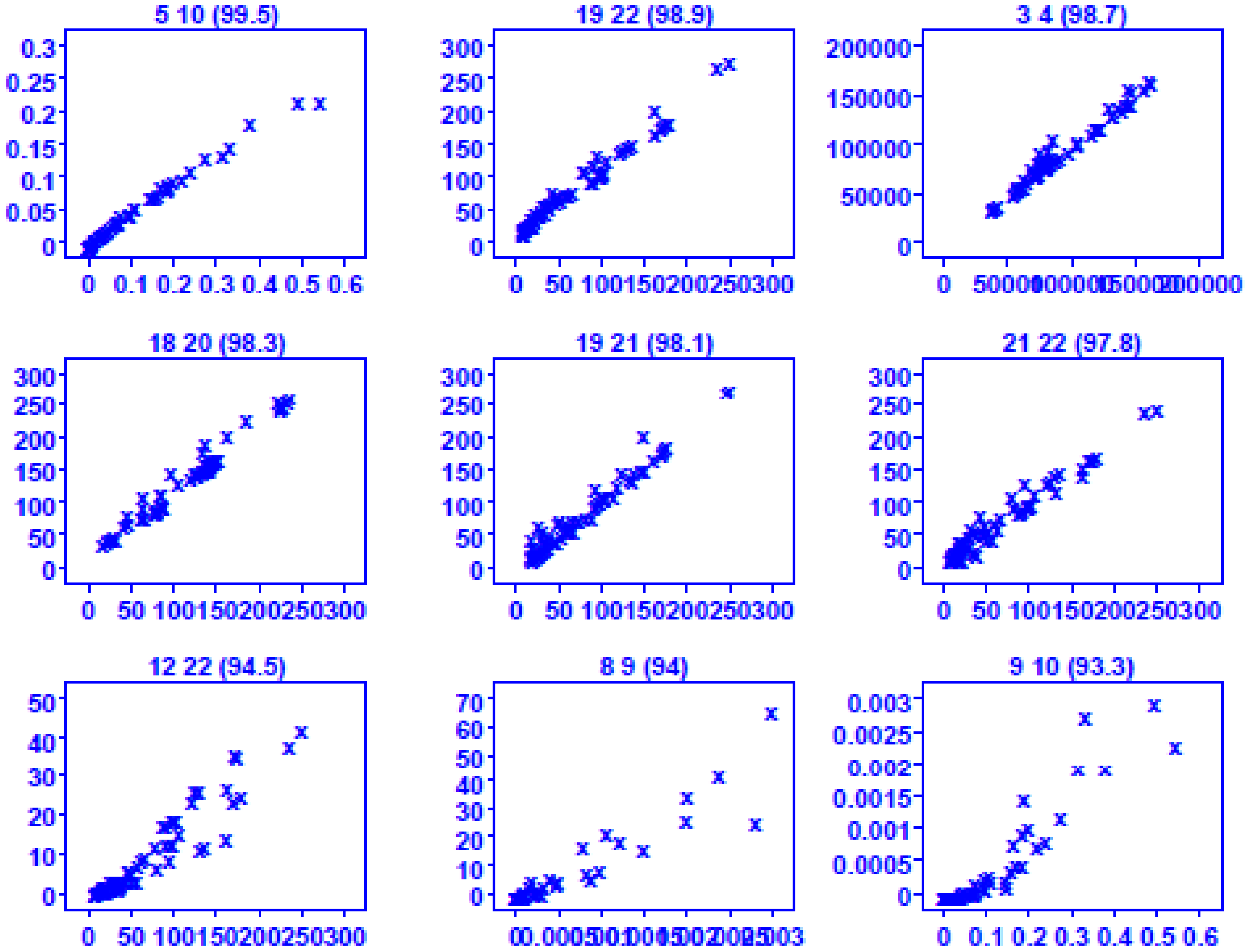
Scatter Plot Matrix of Raw Responses



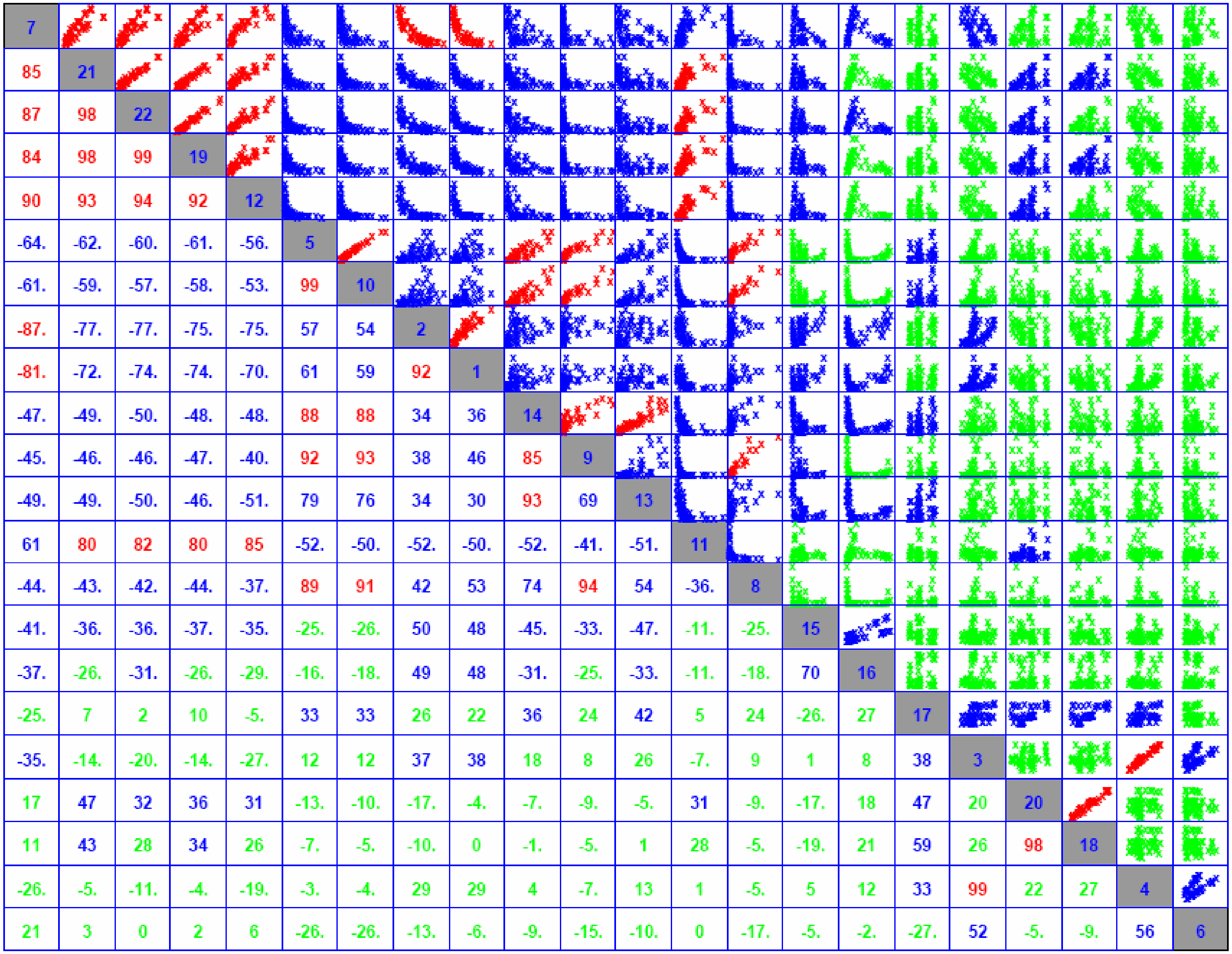
Sorted Correlations



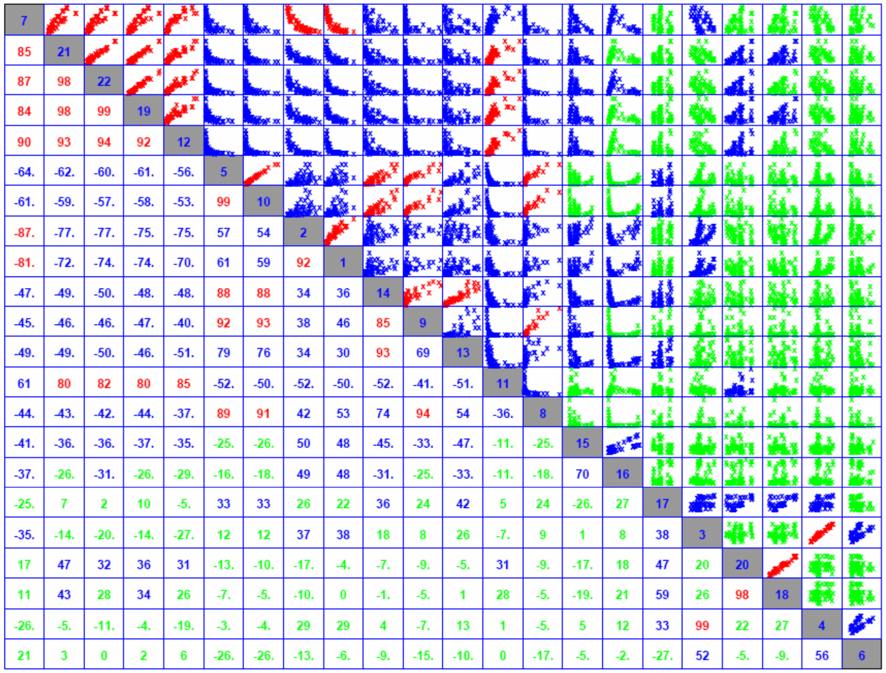
Sorted Correlations



Matrix of Pair-wise Scatter Plots & Correlation Coefficients (Ordered)

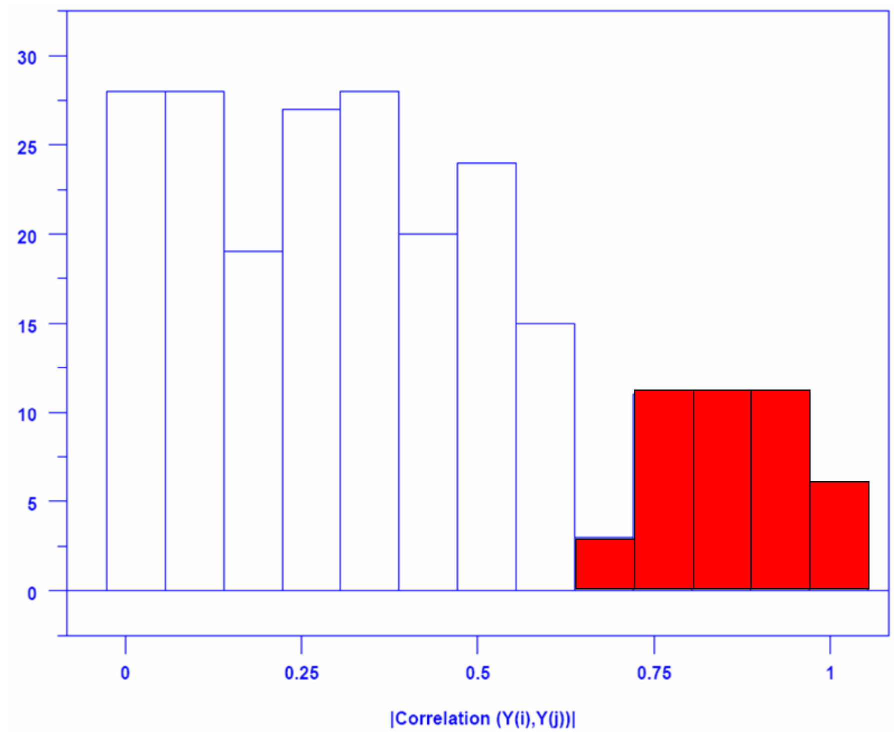


Red 80 ≥ |r|x100 ≤ 100 **Blue** 30 ≥ |r|x100 < 80 **Green** |r|x100 < 30



Red 80 ≥ |r|x100 ≤ 100 Blue 30 ≥ |r|x100 < 80 Green |r|x100 < 30

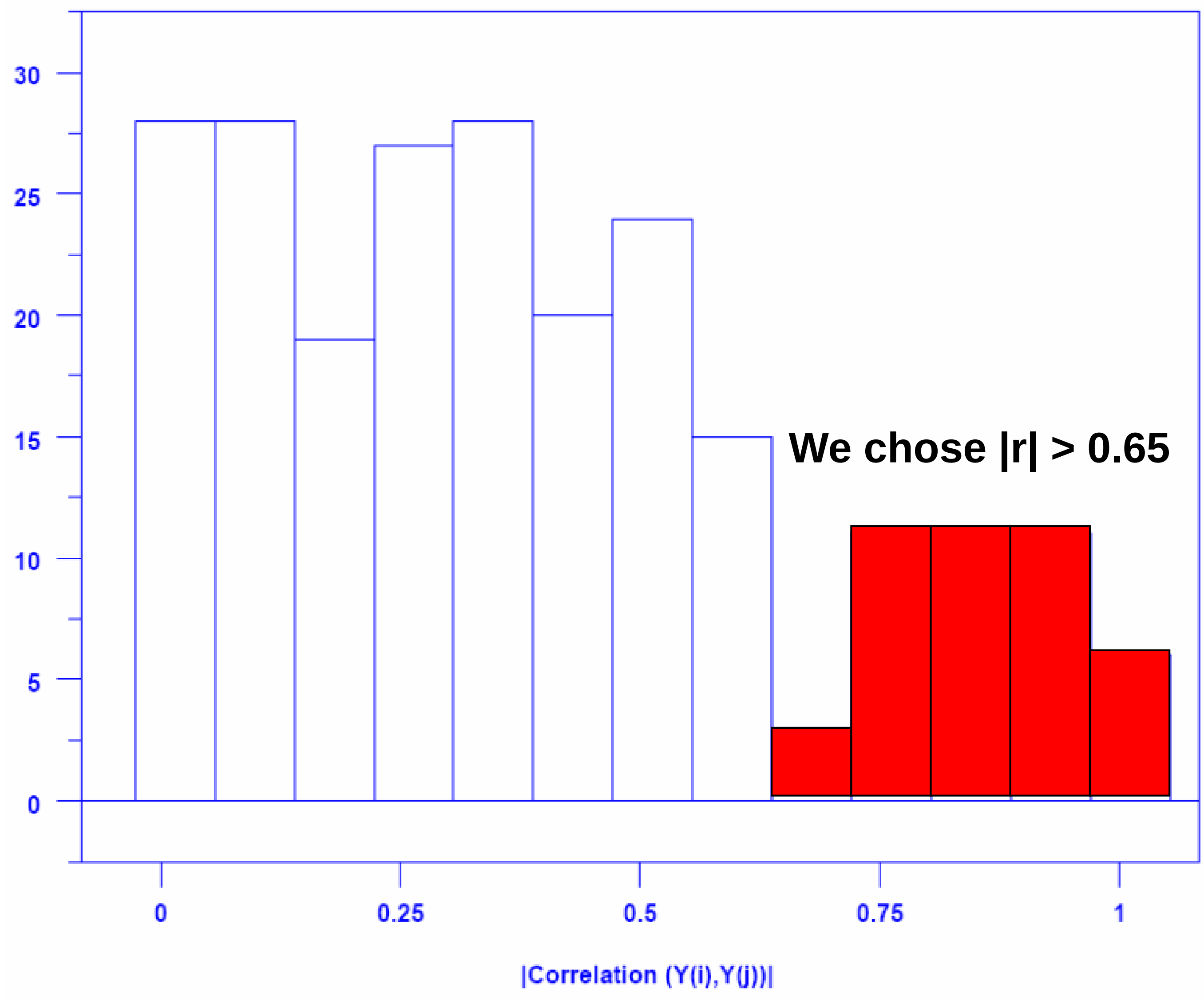
(a) Pair-wise Correlation Matrix



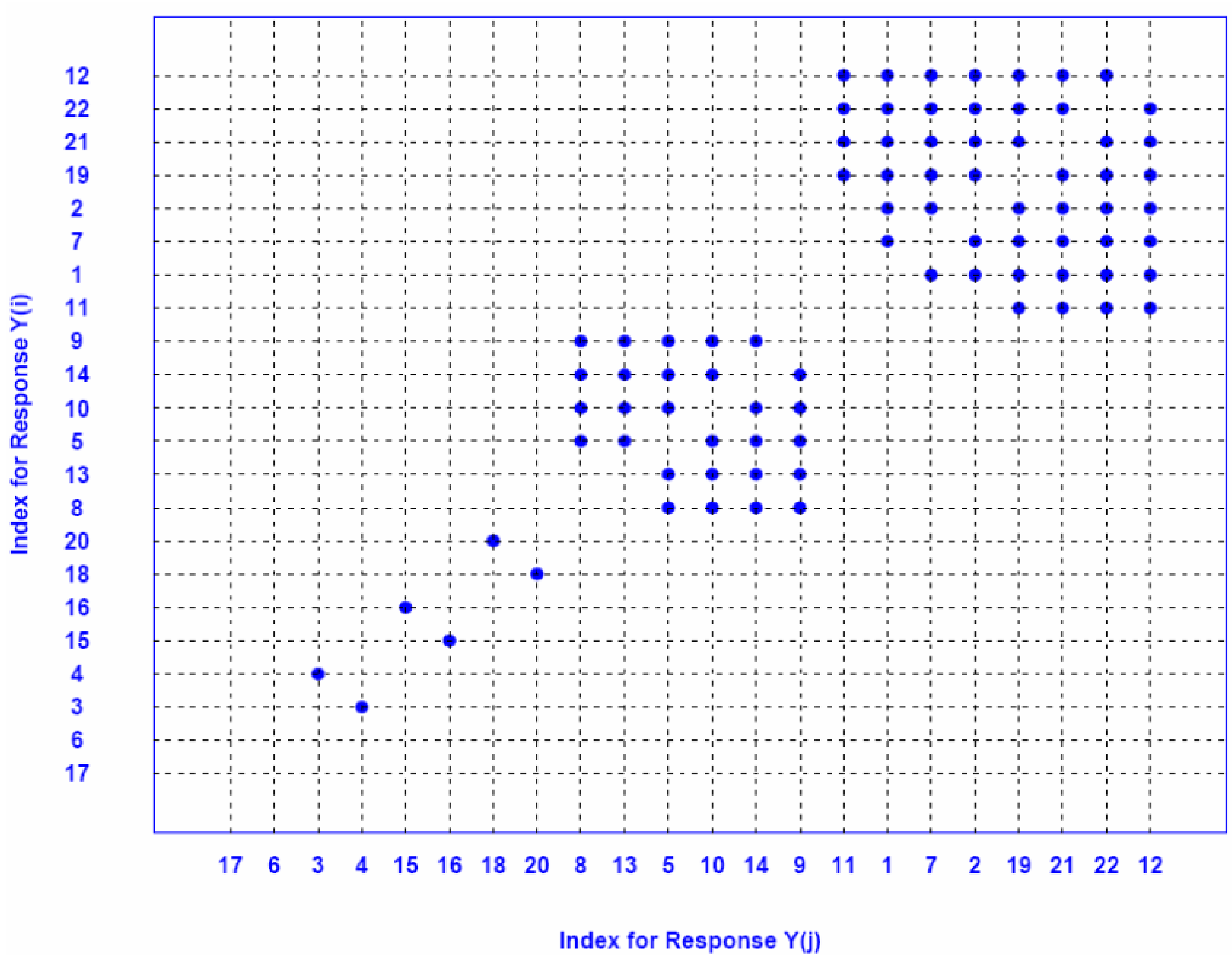
(b) Histogram: bins where |r| > 0.65 highlighted in red

Frequency Distribution of Absolute Value of Correlation Coefficients for All Response Pairs

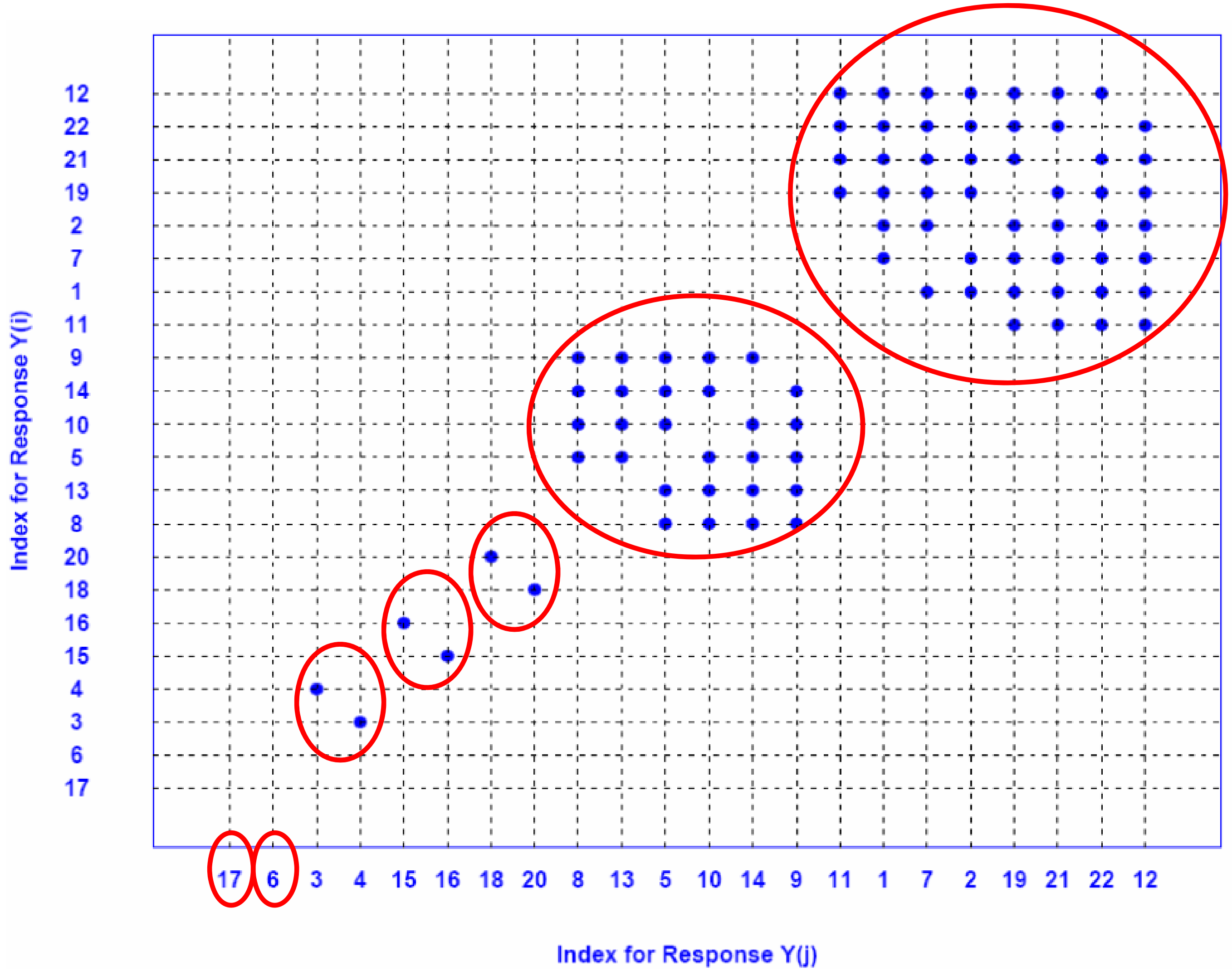
Select a threshold for |r| such that correlations above that threshold will be further considered



Response Index-Index Plot where $|r_{i,j}| > 0.65$ Clustered into Mutual Correlations



Response Index-Index Plot where $|r_{i,j}| > 0.65$ Clustered into Mutual Correlations



Plot suggests *MesoNet* exhibits 7 distinct behaviors

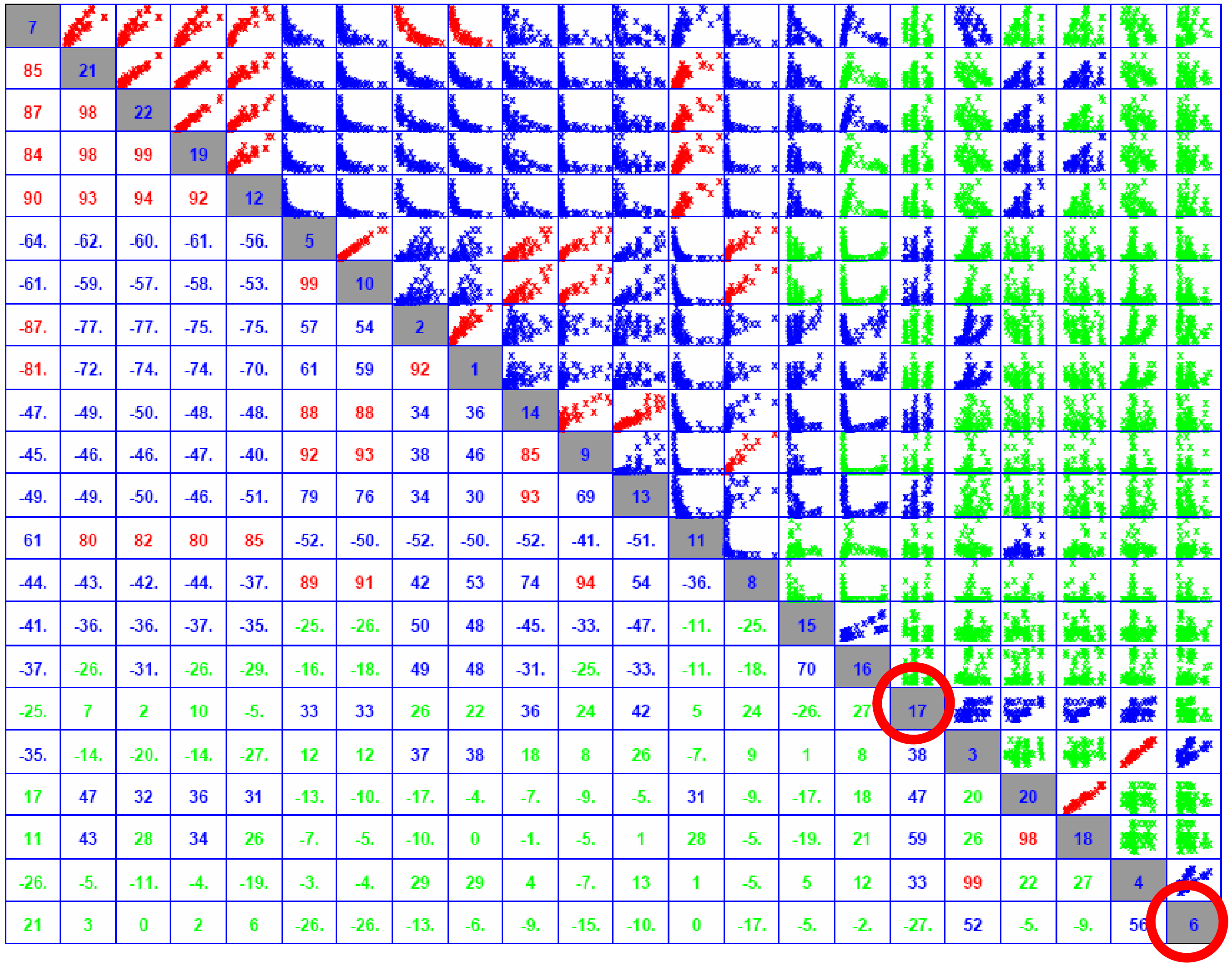
22 Responses: 16 Macro + 6 Throughput

Response	Definition
y1	Active Flows – flows attempting to transfer data
y2	Proportion of potential flows that were active: Active Flows/All Sources
y3	Data packets entering the network per measurement interval
y4	Data packets leaving the network per measurement interval
y5	Loss Rate: y4/(y3+y4)
y6	Flows Completed per measurement interval
y7	Flow-Completion Rate: y6/(y6+y1)
y8	Connection Failures per measurement interval
y9	Connection-Failure Rate: y8/(y8+y1)
y10	Retransmission Rate (ratio)
y11	Congestion Window per Flow (packets)
y12	Window Increases per Flow per measurement interval
y13	Negative Acknowledgments per Flow per measurement interval
y14	Timeouts per Flow per measurement interval
y15	Smoothed Round-Trip Time (ms)
y16	Relative queuing delay: y15/(x1x41)

y17	Average Throughput for Active DD Flows
y18	Average Throughput for Active DF Flows
y19	Average Throughput for Active DN Flows
y20	Average Throughput for Active FF Flows
y21	Average Throughput for Active FN Flows
y22	Average Throughput for Active NN Flows

(k=11,n=64,m=22)

Matrix of Pair-wise Scatter Plots & Correlation Coefficients (Ordered)



Red 80 ≥ |r|x100 ≤ 100 **Blue** 30 ≥ |r|x100 < 80 **Green** |r|x100 < 30

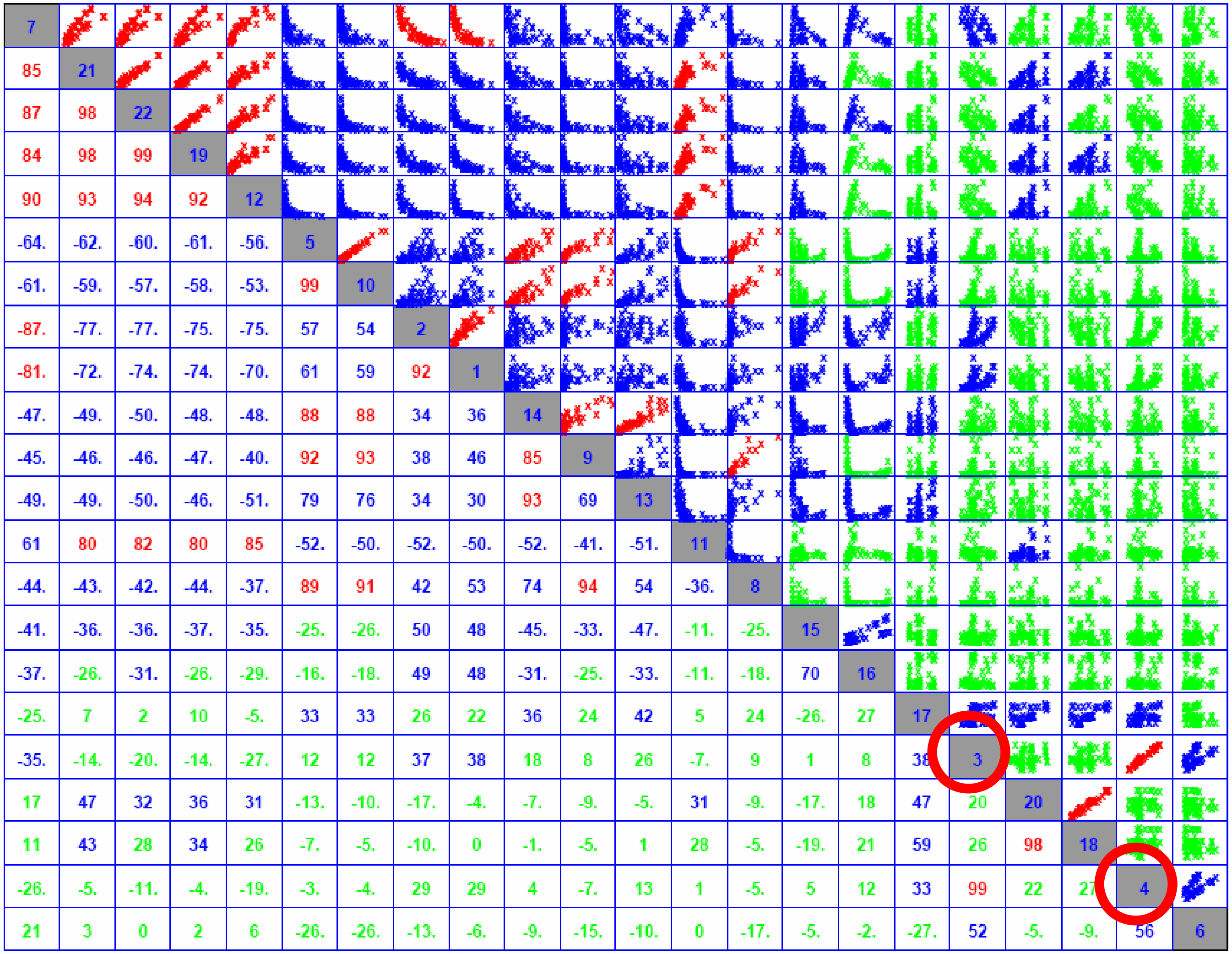
22 Responses: 16 Macro + 6 Throughput

Response	Definition
y1	Active Flows – flows attempting to transfer data
y2	Proportion of potential flows that were active: Active Flows/All Sources
y3	Data packets entering the network per measurement interval
y4	Data packets leaving the network per measurement interval
y5	Loss Rate: y4/(y3+y4)
y6	Flows Completed per measurement interval
y7	Flow-Completion Rate: y6/(y6+y1)
y8	Connection Failures per measurement interval
y9	Connection-Failure Rate: y8/(y8+y1)
y10	Retransmission Rate (ratio)
y11	Congestion Window per Flow (packets)
y12	Window Increases per Flow per measurement interval
y13	Negative Acknowledgments per Flow per measurement interval
y14	Timeouts per Flow per measurement interval
y15	Smoothed Round-Trip Time (ms)
y16	Relative queuing delay: y15/(x1x41)

y17	Average Throughput for Active DD Flows
y18	Average Throughput for Active DF Flows
y19	Average Throughput for Active DN Flows
y20	Average Throughput for Active FF Flows
y21	Average Throughput for Active FN Flows
y22	Average Throughput for Active NN Flows

(k=11,n=64,m=22)

Matrix of Pair-wise Scatter Plots & Correlation Coefficients (Ordered)



Red 80 ≥ |r|x100 ≤ 100 **Blue** 30 ≥ |r|x100 < 80 **Green** |r|x100 < 30

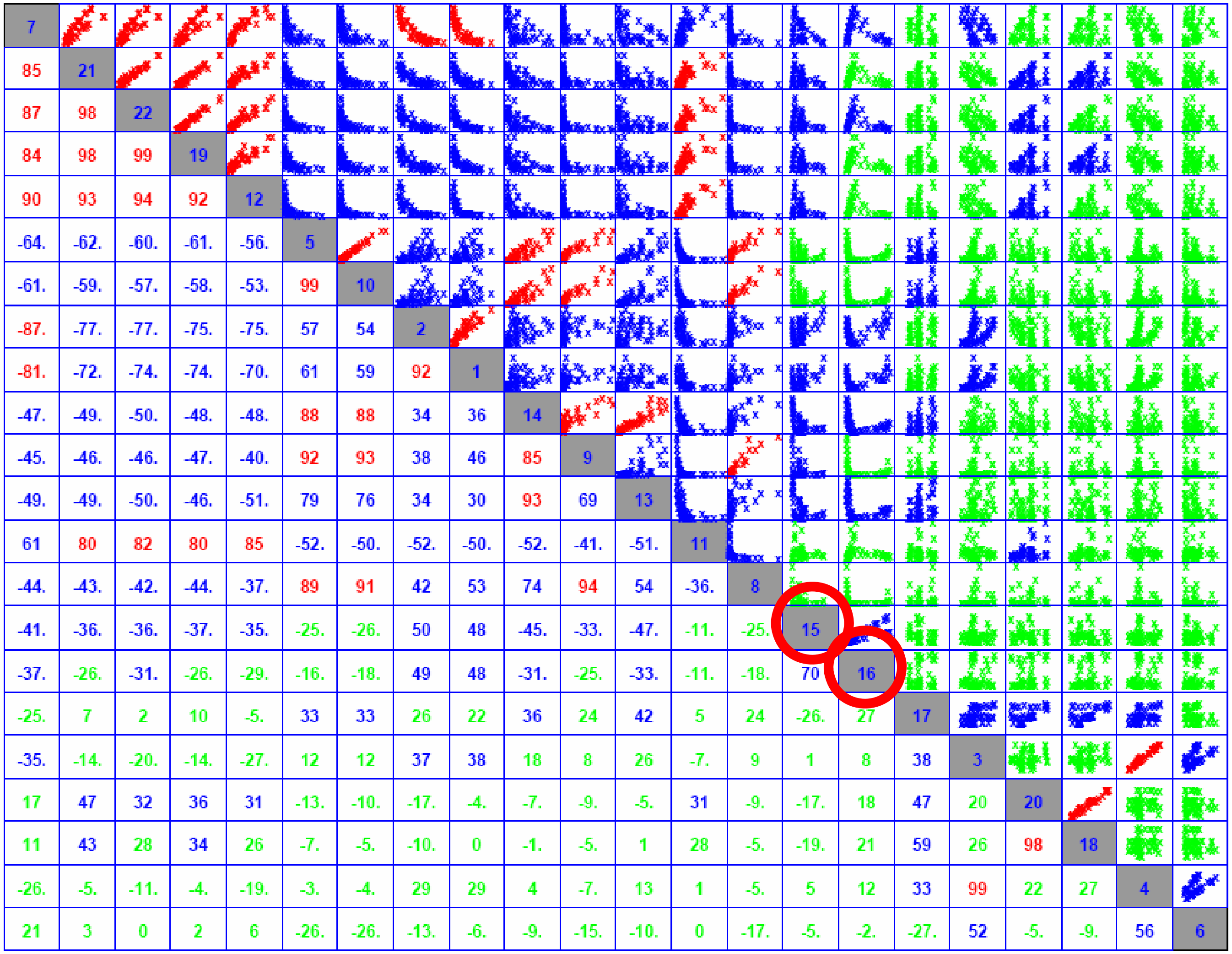
22 Responses: 16 Macro + 6 Throughput

Response	Definition
y1	Active Flows – flows attempting to transfer data
y2	Proportion of potential flows that were active: Active Flows/All Sources
y3	Data packets entering the network per measurement interval
y4	Data packets leaving the network per measurement interval
y5	Loss Rate: y4/(y3+y4)
y6	Flows Completed per measurement interval
y7	Flow-Completion Rate: y6/(y6+y1)
y8	Connection Failures per measurement interval
y9	Connection-Failure Rate: y8/(y8+y1)
y10	Retransmission Rate (ratio)
y11	Congestion Window per Flow (packets)
y12	Window Increases per Flow per measurement interval
y13	Negative Acknowledgments per Flow per measurement interval
y14	Timeouts per Flow per measurement interval
y15	Smoothed Round-Trip Time (ms)
y16	Relative queuing delay: y15/(x1x41)

y17	Average Throughput for Active DD Flows
y18	Average Throughput for Active DF Flows
y19	Average Throughput for Active DN Flows
y20	Average Throughput for Active FF Flows
y21	Average Throughput for Active FN Flows
y22	Average Throughput for Active NN Flows

(k=11,n=64,m=22)

Matrix of Pair-wise Scatter Plots & Correlation Coefficients (Ordered)



Red 80 ≥ |r|x100 ≤ 100 **Blue** 30 ≥ |r|x100 < 80 **Green** |r|x100 < 30

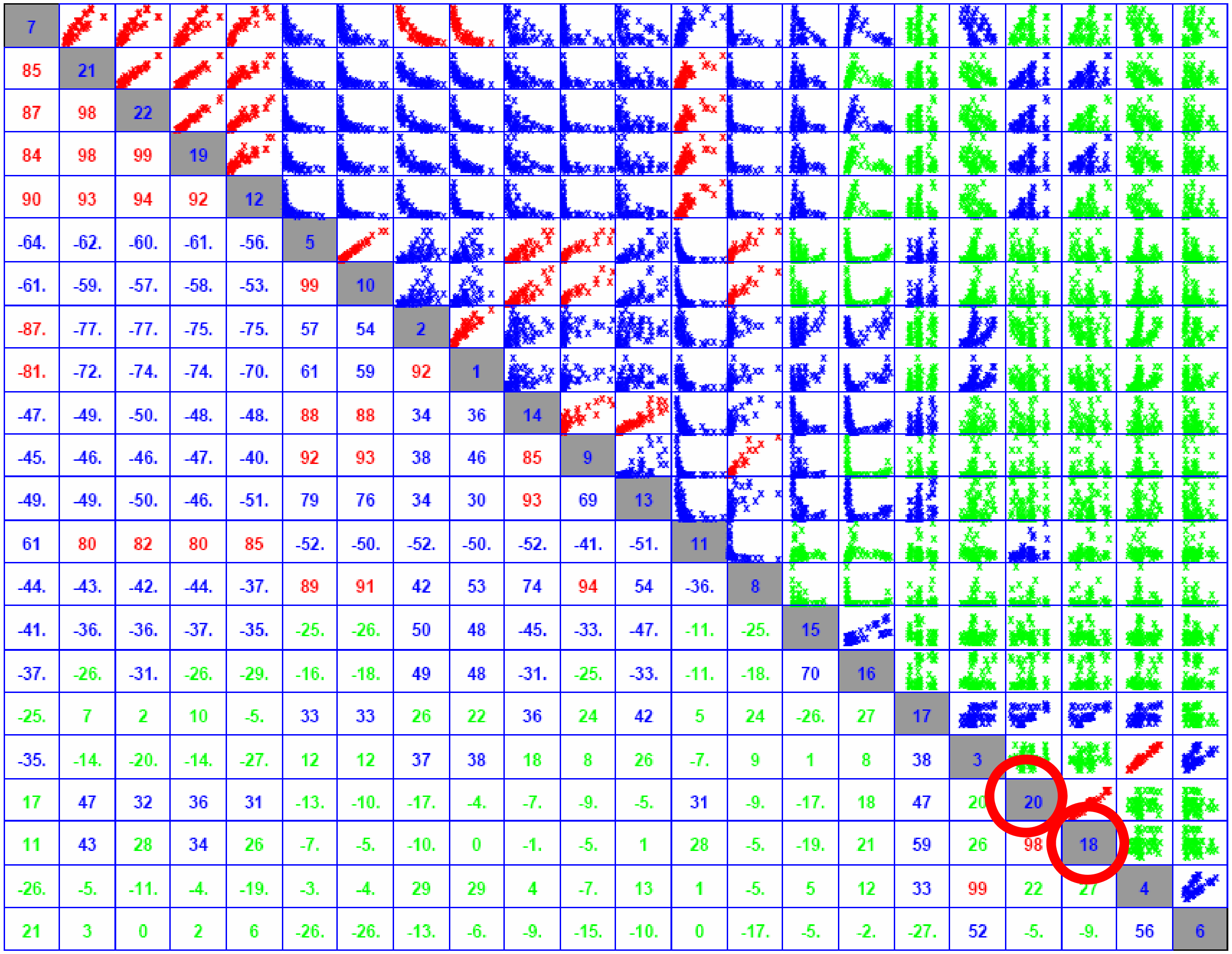
22 Responses: 16 Macro + 6 Throughput

Response	Definition
y1	Active Flows – flows attempting to transfer data
y2	Proportion of potential flows that were active: Active Flows/All Sources
y3	Data packets entering the network per measurement interval
y4	Data packets leaving the network per measurement interval
y5	Loss Rate: y4/(y3+y4)
y6	Flows Completed per measurement interval
y7	Flow-Completion Rate: y6/(y6+y1)
y8	Connection Failures per measurement interval
y9	Connection-Failure Rate: y8/(y8+y1)
y10	Retransmission Rate (ratio)
y11	Congestion Window per Flow (packets)
y12	Window Increases per Flow per measurement interval
y13	Negative Acknowledgments per Flow per measurement interval
y14	Timeouts per Flow per measurement interval
y15	Smoothed Round-Trip Time (ms)
y16	Relative queuing delay: y15/(x1x41)

y17	Average Throughput for Active DD Flows
y18	Average Throughput for Active DF Flows
y19	Average Throughput for Active DN Flows
y20	Average Throughput for Active FF Flows
y21	Average Throughput for Active FN Flows
y22	Average Throughput for Active NN Flows

(k=11,n=64,m=22)

Matrix of Pair-wise Scatter Plots & Correlation Coefficients (Ordered)



Red $80 \geq |r|\text{x}100 \leq 100$ **Blue** $30 \geq |r|\text{x}100 < 80$ **Green** $|r|\text{x}100 < 30$

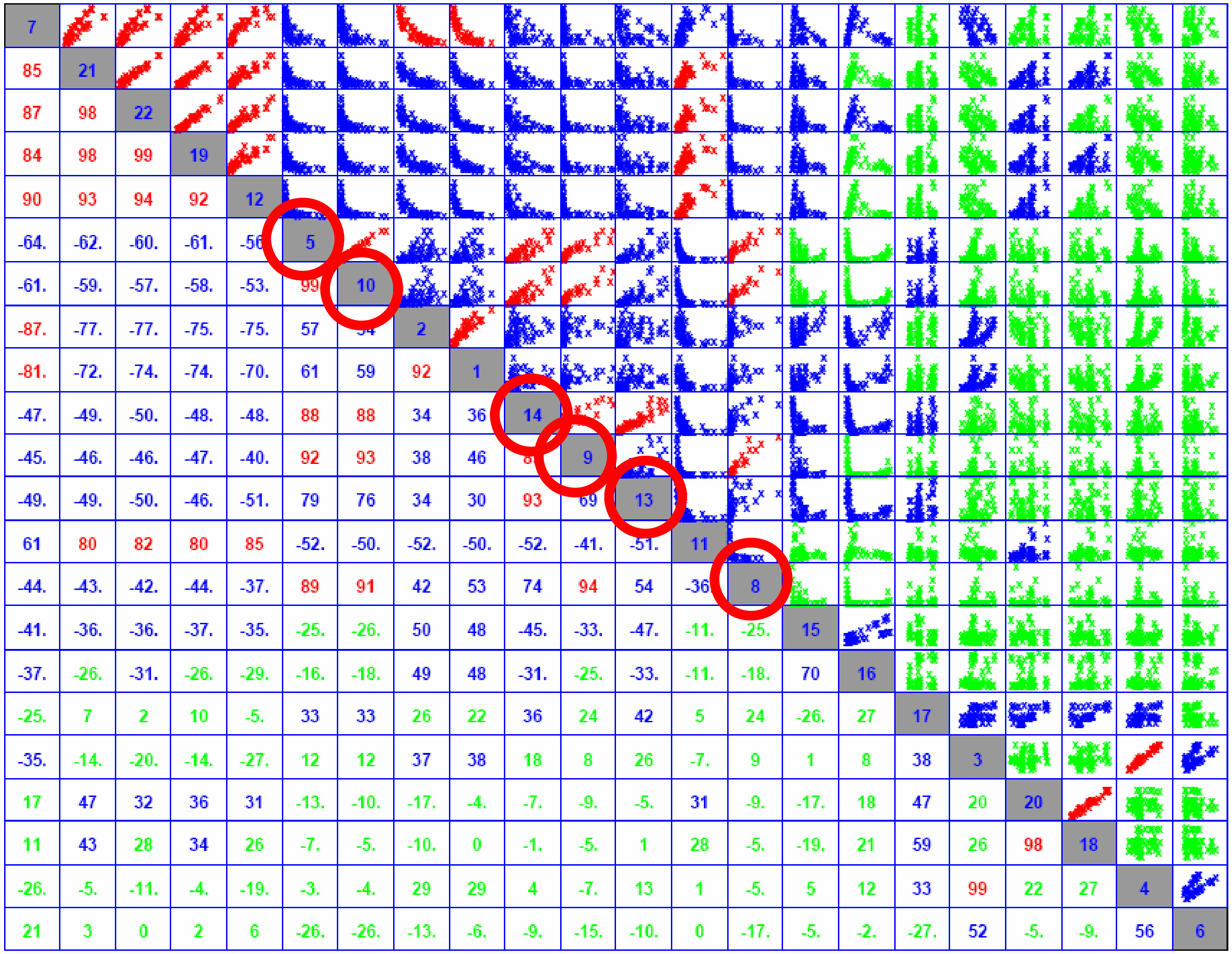
22 Responses: 16 Macro + 6 Throughput

Response	Definition
y1	Active Flows – flows attempting to transfer data
y2	Proportion of potential flows that were active: Active Flows/All Sources
y3	Data packets entering the network per measurement interval
y4	Data packets leaving the network per measurement interval
y5	Loss Rate: y4/(y3+y4)
y6	Flows Completed per measurement interval
y7	Flow-Completion Rate: y6/(y6+y1)
y8	Connection Failures per measurement interval
y9	Connection-Failure Rate: y8/(y8+y1)
y10	Retransmission Rate (ratio)
y11	Congestion Window per Flow (packets)
y12	Window Increases per Flow per measurement interval
y13	Negative Acknowledgments per Flow per measurement interval
y14	Timeouts per Flow per measurement interval
y15	Smoothed Round-Trip Time (ms)
y16	Relative queuing delay: y15/(x1x41)

y17	Average Throughput for Active DD Flows
y18	Average Throughput for Active DF Flows
y19	Average Throughput for Active DN Flows
y20	Average Throughput for Active FF Flows
y21	Average Throughput for Active FN Flows
y22	Average Throughput for Active NN Flows

(k=11,n=64,m=22)

Matrix of Pair-wise Scatter Plots & Correlation Coefficients (Ordered)



Red 80 ≥ |r|x100 ≤ 100 **Blue** 30 ≥ |r|x100 < 80 **Green** |r|x100 < 30

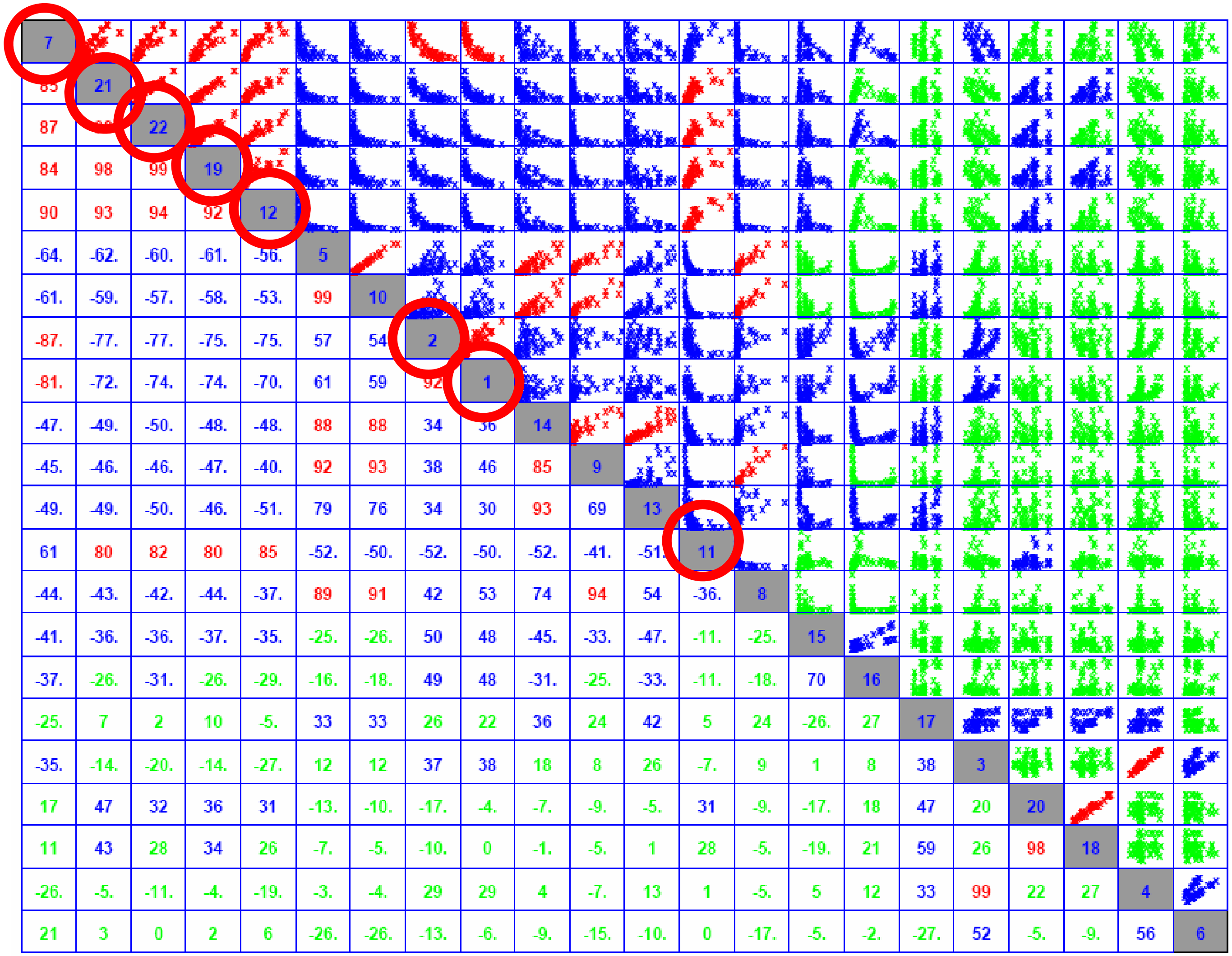
22 Responses: 16 Macro + 6 Throughput

Response	Definition
y1	Active Flows – flows attempting to transfer data
y2	Proportion of potential flows that were active: Active Flows/All Sources
y3	Data packets entering the network per measurement interval
y4	Data packets leaving the network per measurement interval
y5	Loss Rate: y4/(y3+y4)
y6	Flows Completed per measurement interval
y7	Flow-Completion Rate: y6/(y6+y1)
y8	Connection Failures per measurement interval
y9	Connection-Failure Rate: y8/(y8+y1)
y10	Retransmission Rate (ratio)
y11	Congestion Window per Flow (packets)
y12	Window Increases per Flow per measurement interval
y13	Negative Acknowledgments per Flow per measurement interval
y14	Timeouts per Flow per measurement interval
y15	Smoothed Round-Trip Time (ms)
y16	Relative queuing delay: y15/(x1x41)

y17	Average Throughput for Active DD Flows
y18	Average Throughput for Active DF Flows
y19	Average Throughput for Active DN Flows
y20	Average Throughput for Active FF Flows
y21	Average Throughput for Active FN Flows
y22	Average Throughput for Active NN Flows

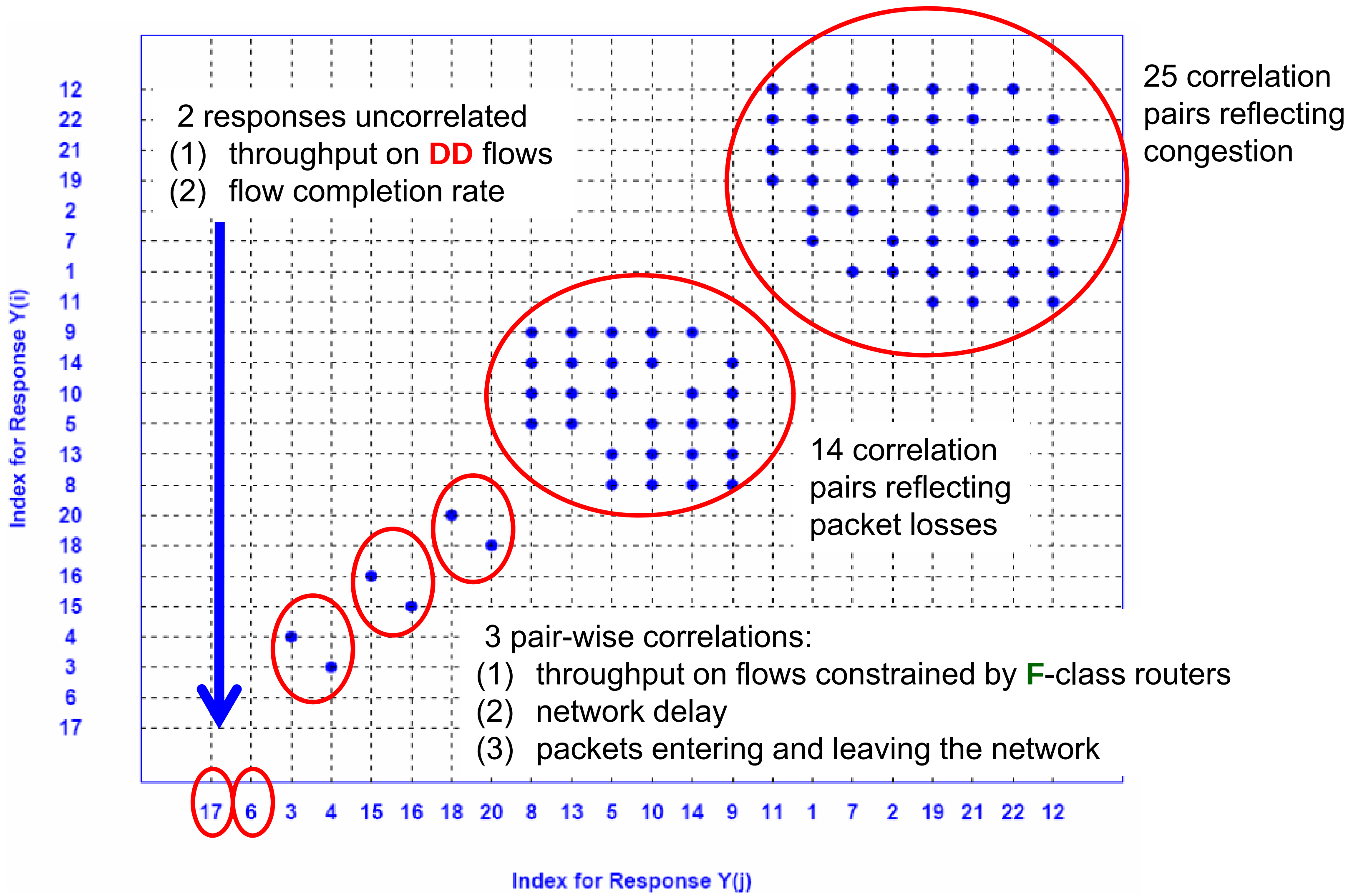
(k=11,n=64,m=22)

Matrix of Pair-wise Scatter Plots & Correlation Coefficients (Ordered)



Red 80 ≥ |r|x100 ≤ 100 **Blue** 30 ≥ |r|x100 < 80 **Green** |r|x100 < 30

Summary: Response Index-Index Plot where $|r_{i,j}| > 0.65$ Clustered into Mutual Correlations



Plot suggests *MesoNet* exhibits 7 distinct behaviors

Summary of Correlation Results

Correlation Analysis

Dimension	Responses
Congestion	y1, y2, y7, y11, y12, y19, y21, y22
Losses	y5, y8, y9, y10, y13, y14
Delay	y15, y16
F-class TP	y18, y20
D-class TP	y17
Packet TP	y3, y4
Flow TP	y6

Summary of Correlation Results

Correlation Analysis

Dimension	Responses
Congestion	y1, y2, y7, y11, y12, y19, y21, y22
Losses	y5, y8, y9, y10, y13, y14
Delay	y15, y16
F-class TP	y18, y20
D-class TP	y17
Packet TP	y3, y4
Flow TP	y6

22 Responses: 16 Macro + 6 Throughput

Response	Definition
y1	Active Flows – flows attempting to transfer data
y2	Proportion of potential flows that were active: Active Flows/All Sources
y3	Data packets entering the network per measurement interval
y4	Data packets leaving the network per measurement interval
y5	Loss Rate: y4/(y3+y4)
y6	Flows Completed per measurement interval
y7	Flow-Completion Rate: y6/(y6+y1)
y8	Connection Failures per measurement interval
y9	Connection-Failure Rate: y8/(y8+y1)
y10	Retransmission Rate (ratio)
y11	Congestion Window per Flow (packets)
y12	Window Increases per Flow per measurement interval
y13	Negative Acknowledgments per Flow per measurement interval
y14	Timeouts per Flow per measurement interval
y15	Smoothed Round-Trip Time (ms)
y16	Relative queuing delay: y15/(x1x41)

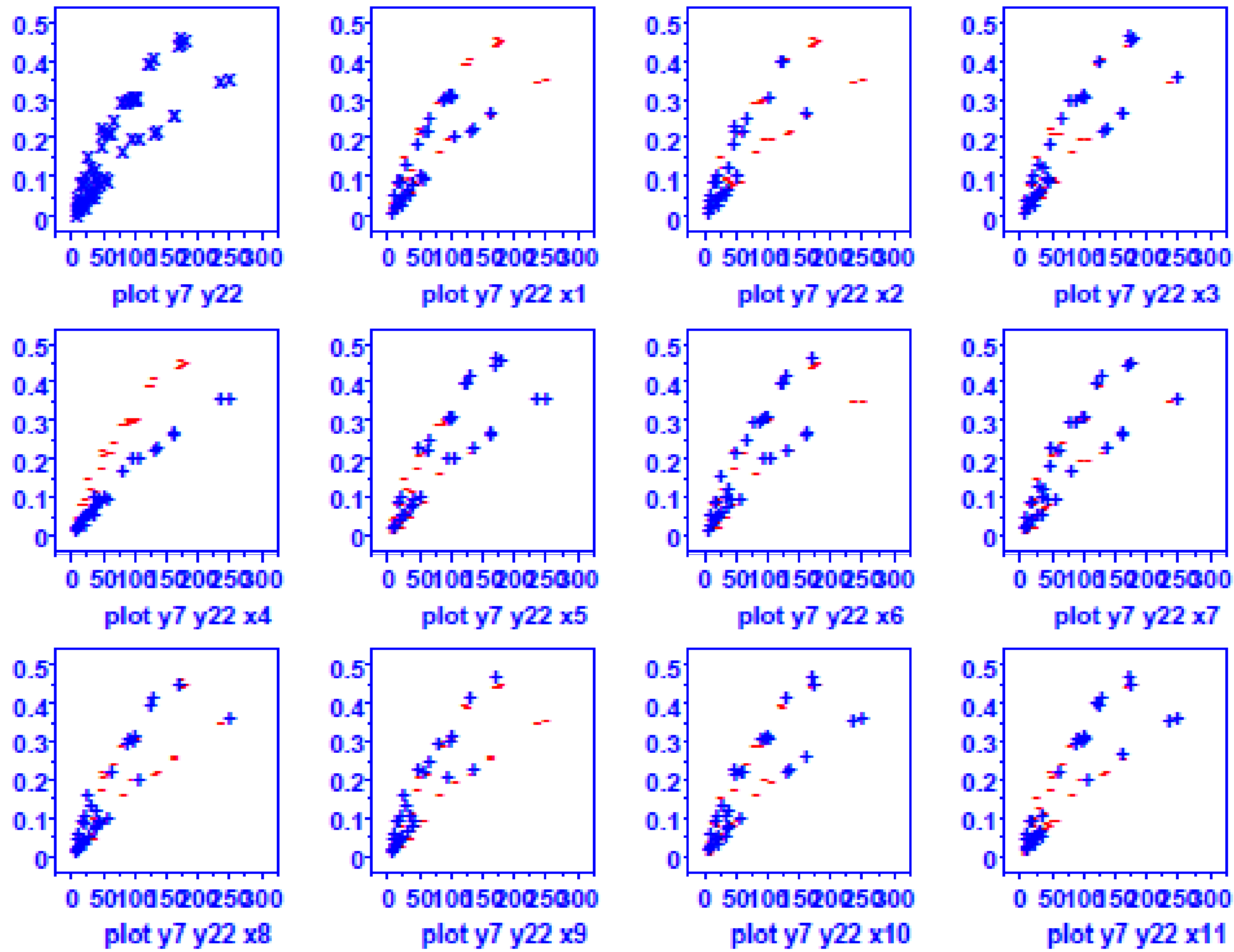
y17	Average Throughput for Active DD Flows
y18	Average Throughput for Active DF Flows
y19	Average Throughput for Active DN Flows
y20	Average Throughput for Active FF Flows
y21	Average Throughput for Active FN Flows
y22	Average Throughput for Active NN Flows

(k=11,n=64, m=22➔7)

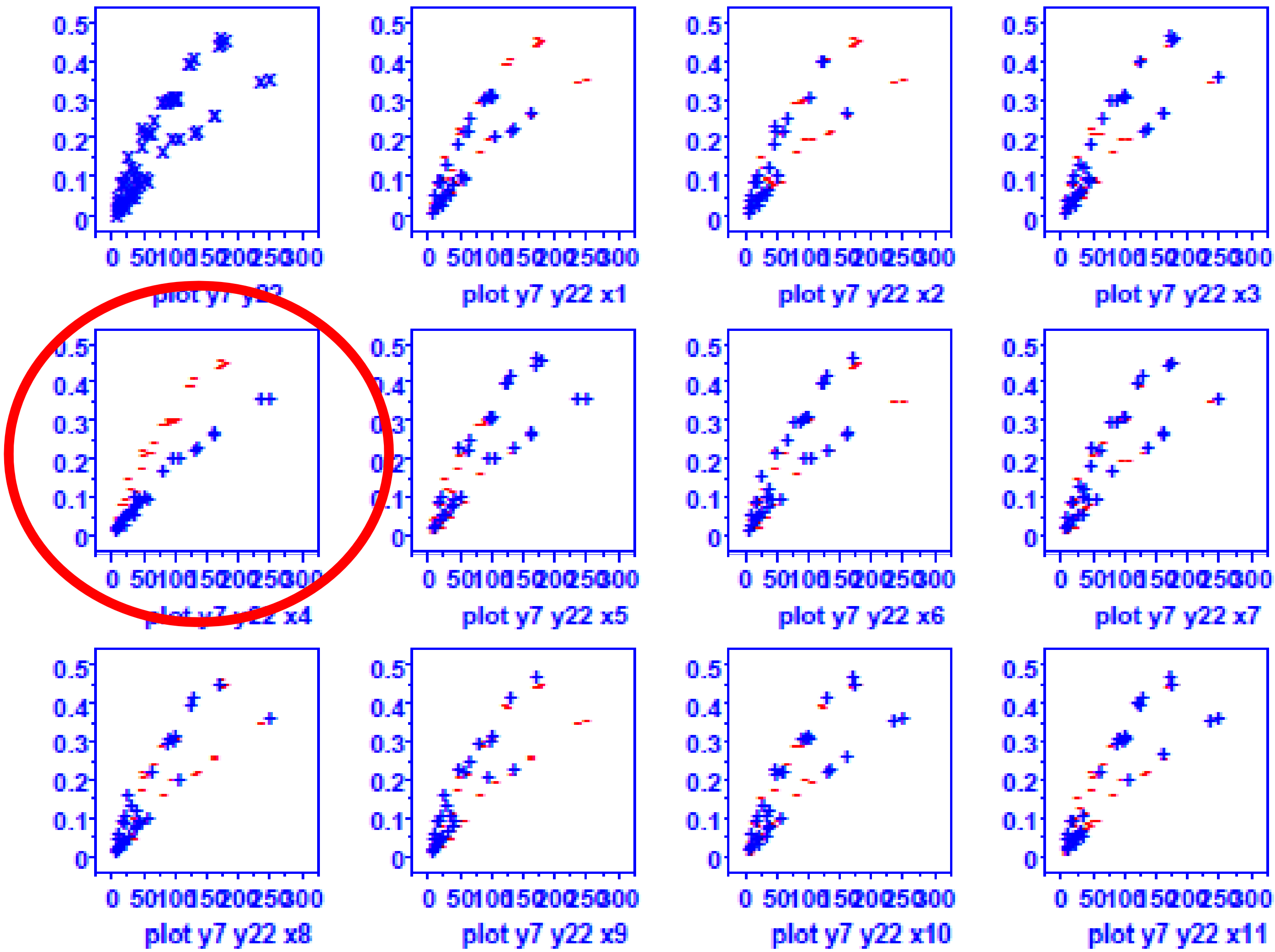
Correlation Analysis & Clustering Suggests ***MesoNet*** Behavior Reflected in Only 7 Responses

Response	Definition
y4	**Average number of packet output per measurement interval** ***(network throughput in packets/sec)***
y6	**Average number of flows completed per measurement interval** ***(network throughput in flows/sec)***
y10	**Average retransmission rate** ***(packet loss)***
y15	**Average smoothed round-trip time** ***(network delay)***
y17	**Average instantaneous throughput for DD flows** ***(throughput in packets/sec for the most advantaged users)***
y20	**Average instantaneous throughput for FF flows** ***(throughput in packets/sec for 2nd most advantaged users)***
y22	**Average instantaneous throughput for NN flows** ***(network congestion)***

Q. Why is the Scatter Plot of Y7 vs Y22 Bifurcated?

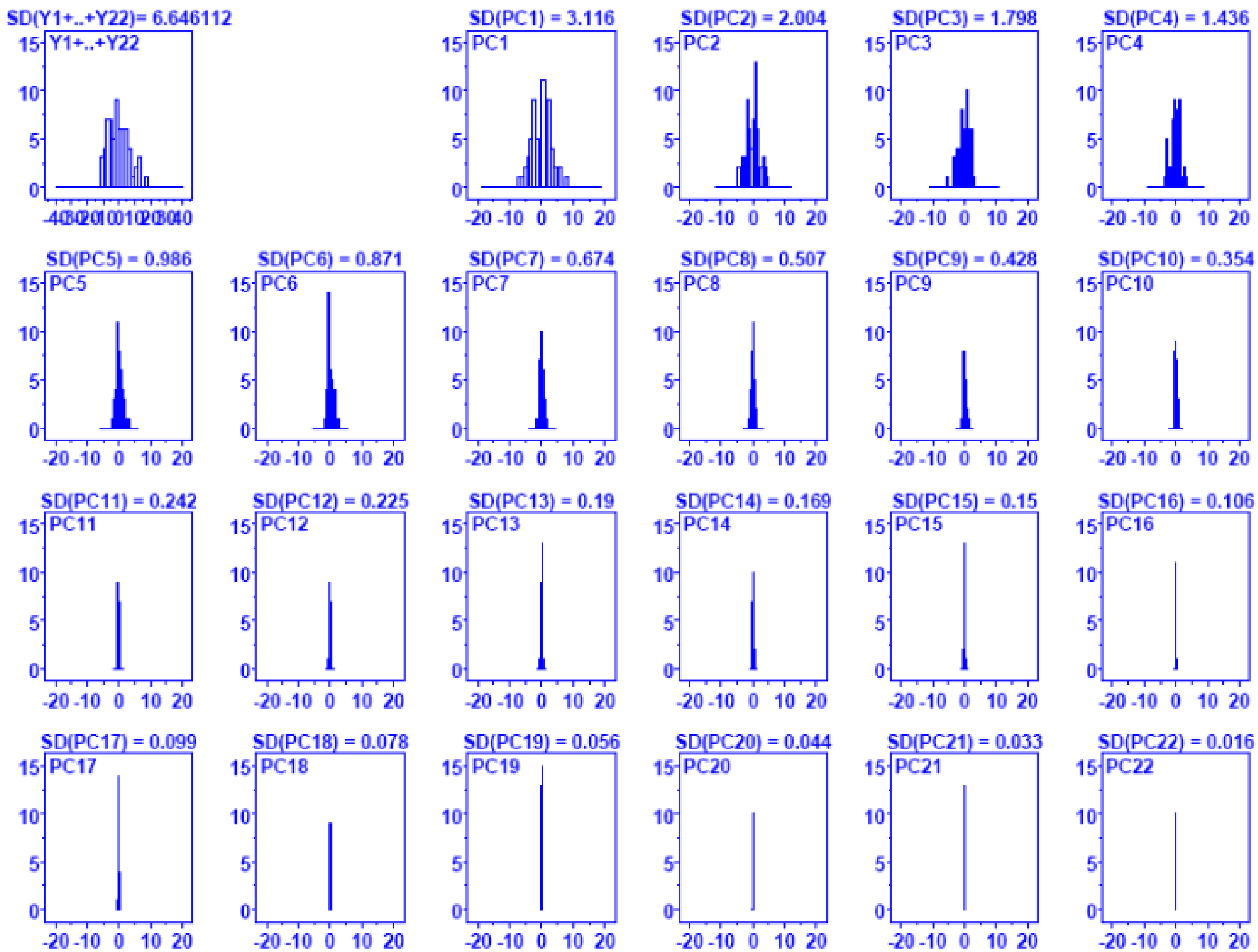


Q. Why is the Scatter Plot of Y7 vs Y22 Bifurcated?



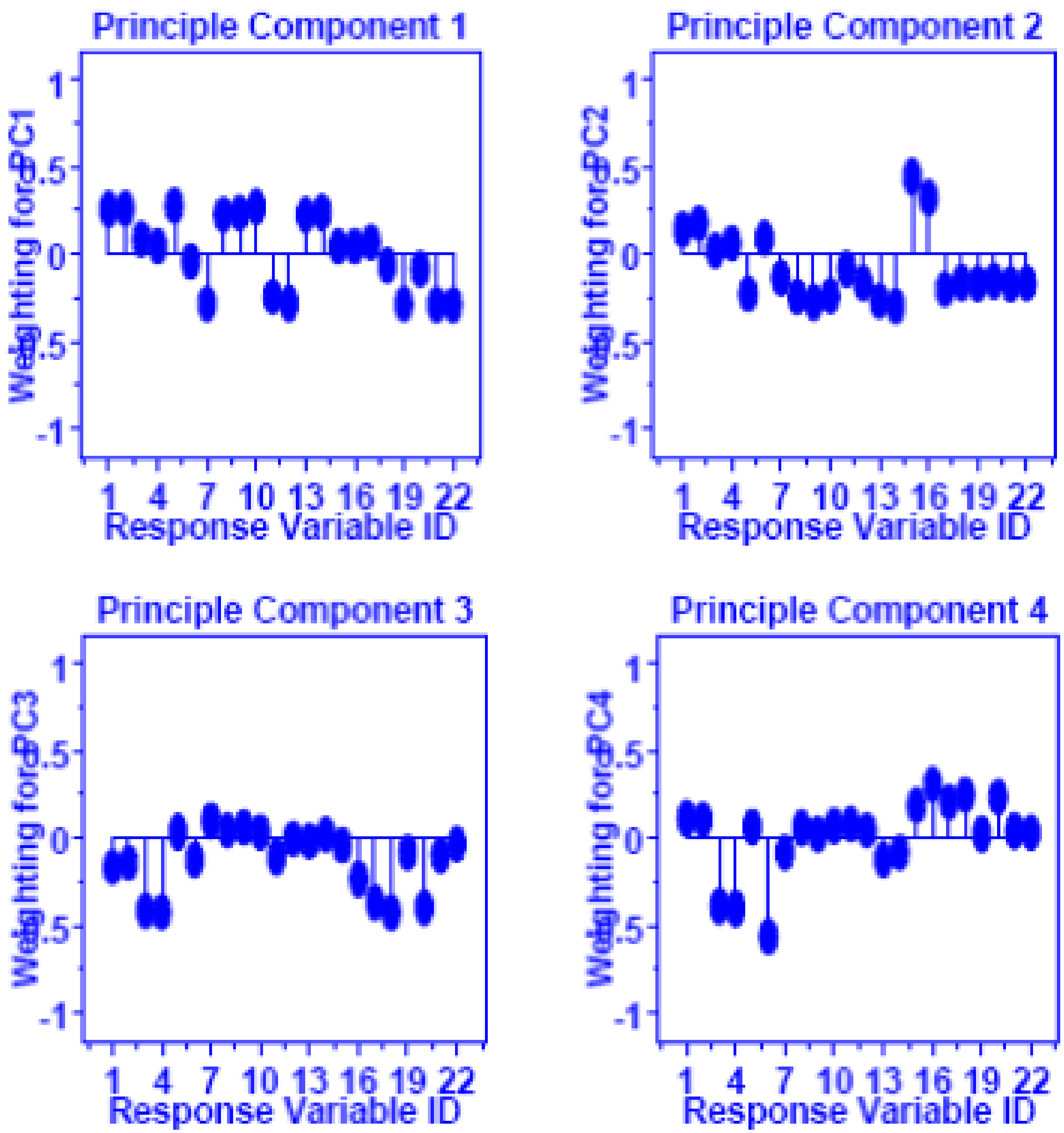
Method 2: Principal Components Analysis

Principal Components Analysis of 22 *MesoNet* Responses

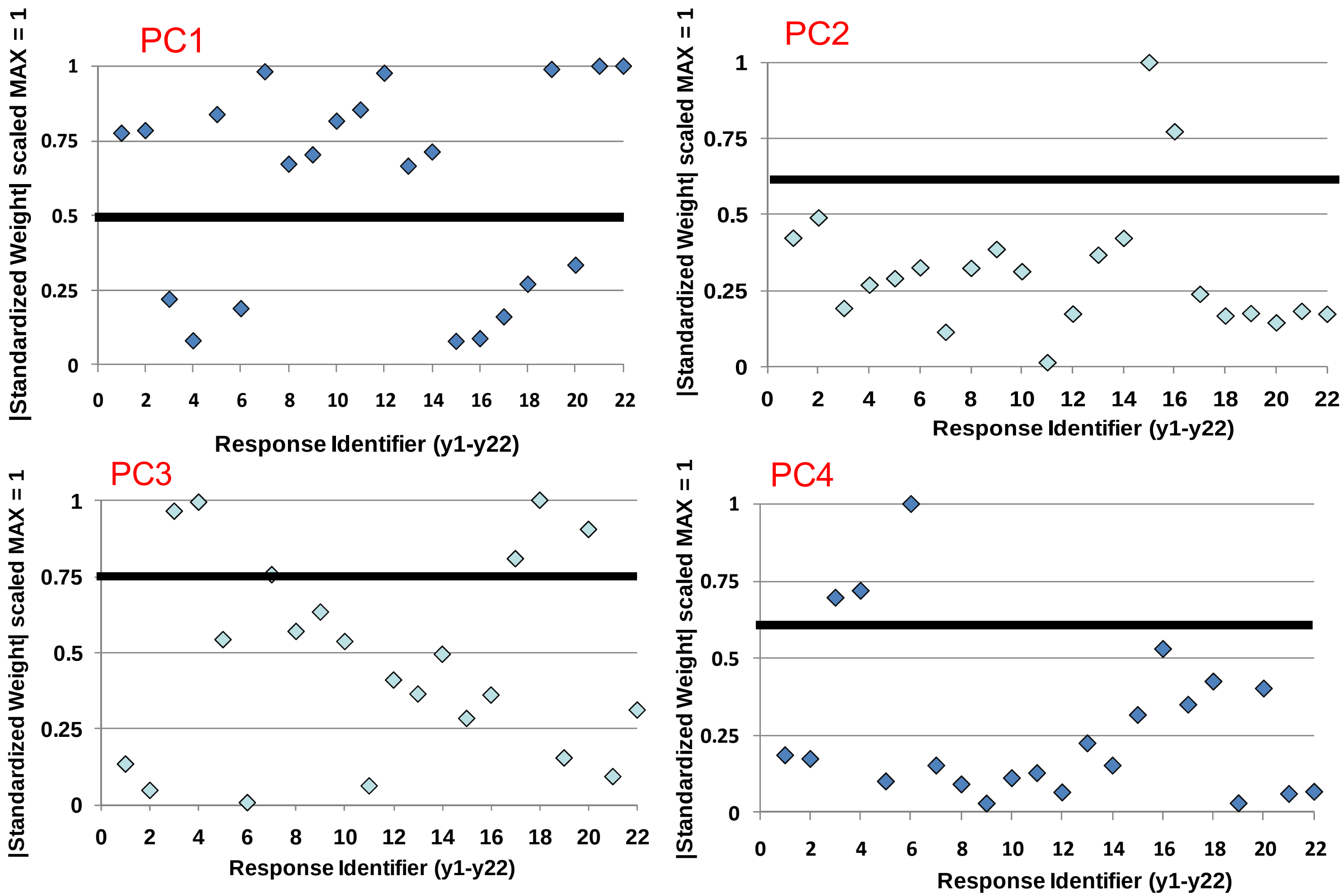


Most response variance appears to be accounted for by the first 4 components

Weight Vectors for the first 4 Components



|Weight| Vectors for the first 4 Components



Significant Responses in PC1 *(congestion)*

Response	Definition
y1	Average number of active flows
y2	Proportion of possible flows that are active
y5	Loss rate
y7	Flow-completion rate
y8	Connection failures
y9	Connection-failure rate
y10	Retransmission rate
y11	Average congestion window
y12	Window-increase rate
y13	Negative-acknowledgment rate
y14	Timeout rate
y19	Average instantaneous throughput for DN flows
y21	Average instantaneous throughput for FN flows
y22	Average instantaneous throughput for NN flows

Significant Responses in PC2 *(delay)*

Response	Definition
y15	Smoothed round-trip time
y16	Relative queuing delay

Significant Responses in PC3 *(throughput for advantaged users)*

Response	Definition
y3	Packets input
y4	Packets output
y17	Average instantaneous throughput for DD flows
y18	Average instantaneous throughput for DF flows
y20	Average instantaneous throughput for FF flows

Significant Responses in PC4 *(network throughput in flows/second)*

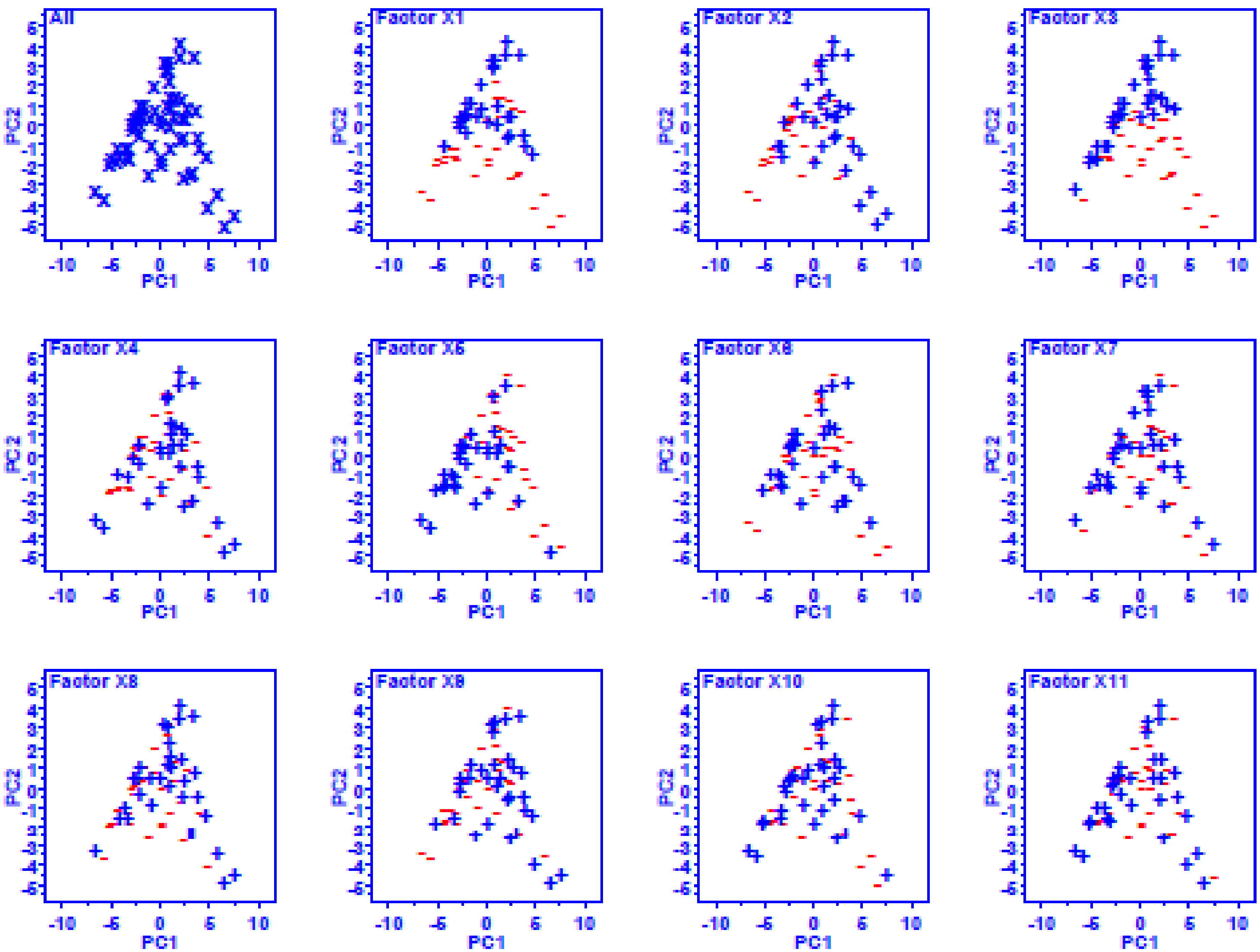
Response	Definition
y3	Packets input
y4	Packets output
y6	Flows completed per measurement interval

Summary of PCA Results

PCA Analysis

Dimension	Responses
PC1: Congestion	y1, y2, y5, y7, y8, y9, y10, y11, y12, y13, y14, y19, y21, y22
PC2: Delay	y15, y16
PC3: D-class & F-class TP	y3, y4, y17, 18, y20
PC4: Flow TP	y3, y4, y6

Abilene Network PCA Analysis (Standardized Responses) (Kevin Mills) Exp. 3 (2to(11-5))
Q. Dominant Factors (Out of the 11)?
PC1 & PC2 Character Plot



Comparing Correlation & PCA Results

Correlation Analysis		PCA	
Dimension	**Responses**	**Dimension**	**Responses**
Congestion	**y1, y2, y7, y11, y12, y19, y21, y22**	**PC1: Congestion**	**y1, y2, y5, y7, y8, y9, y10, y11, y12, y13, y14, y19, y21, y22**
Losses	**y5, y8, y9, y10, y13, y14**		
Delay	**y15, y16**	**PC2: Delay**	**y15, y16**
F-class TP	**y18, y20**	**PC3: D-class & F-class TP**	**y3, y4, y17, 18, y20**
D-class TP	**y17**		
Packet TP	**y3, y4**		
Flow TP	**y6**	**PC4: Flow TP**	**y3, y4, y6**

The results show good alignment:

PCA1 merges congestion + losses;

PCA2 & Correlation identical for delay;

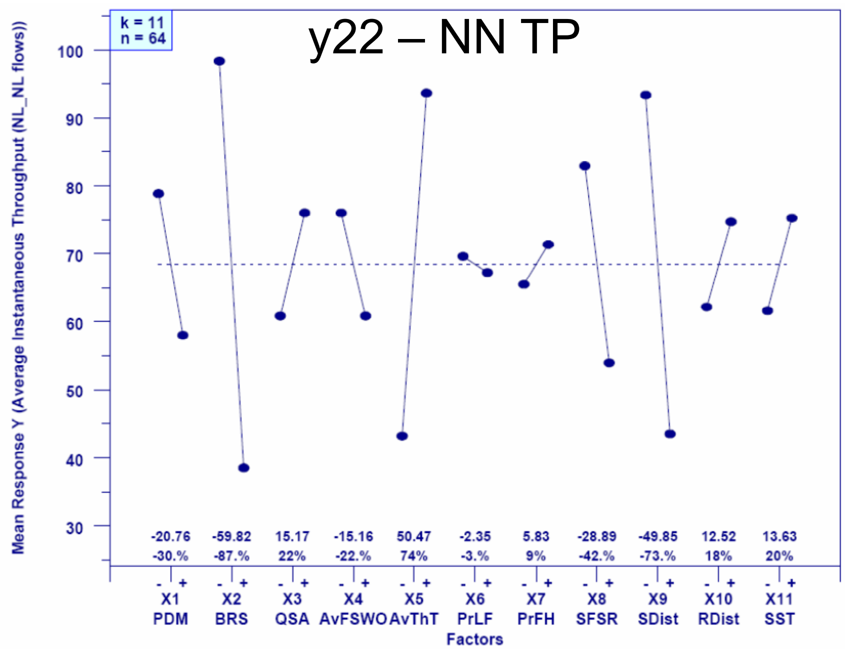
PCA3 merges D-class & F-class Throughput;

PCA4 splits Packet TP acrosstwo dimensions (D- & F-class TP and Flow TP)

Identifying Significant Response Dimensions for *MesoNet*: 4 or 7 or something between?

Note X2 is miscoded so I reverse +/- for X2

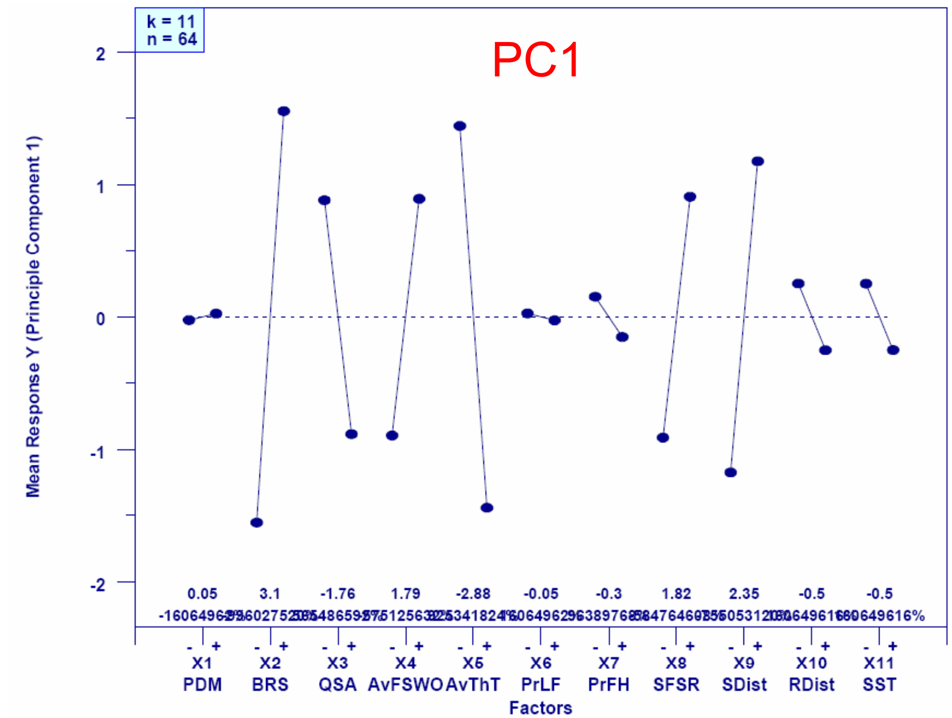
PC1

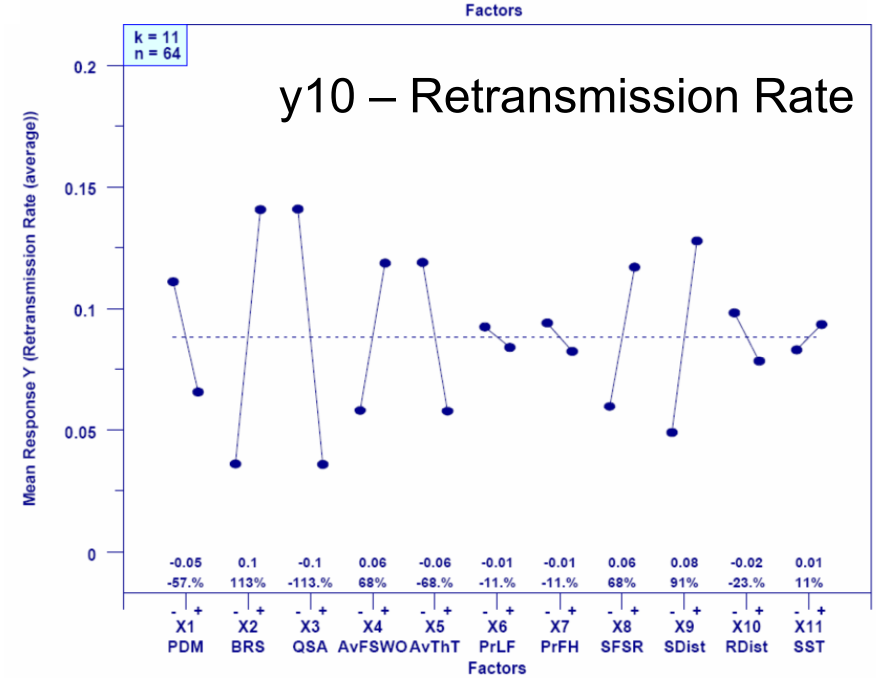


HIGHER CONGESTION IS LOWER TP: -X2, -X5, +X9, +X8, +X1

Note that PC interpretation is possible only by resorting to cross-mapping with response variables

PC+ IS: -X2, -X5, +X9, +X8, +X4, -X3

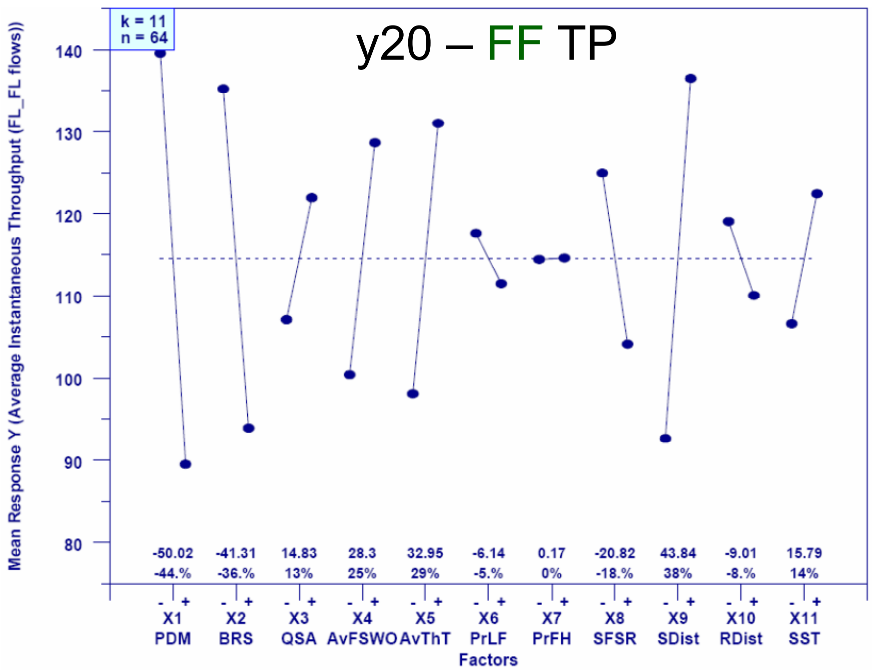




I THINK LOSS & CONGESTION SHOULD BE SEPARATE – SIMILAR CAUSES BUT SUBTLE DIFFERENCES

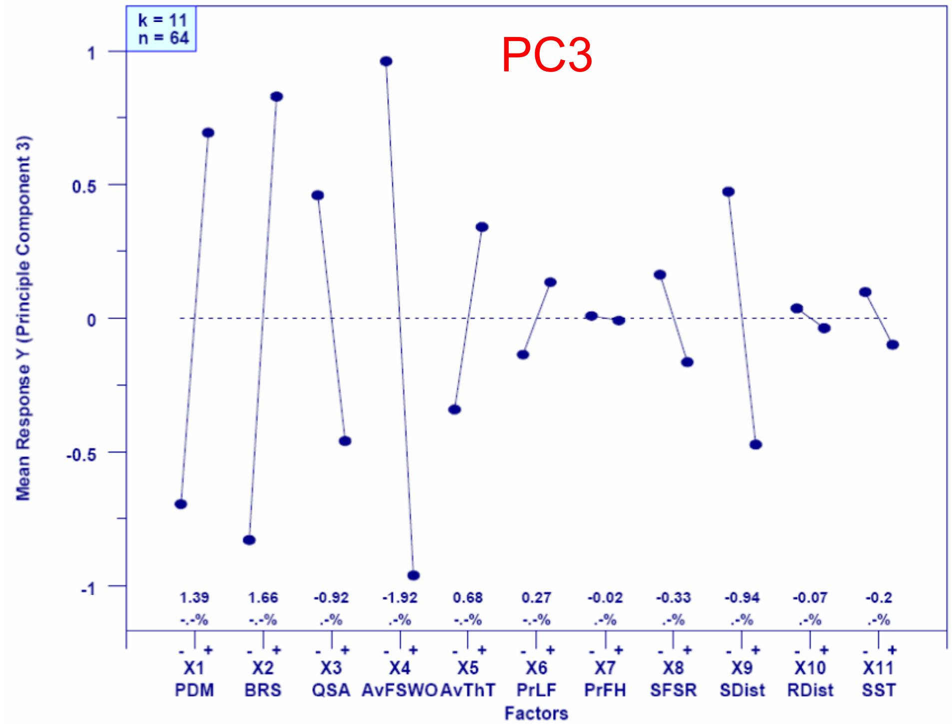
HIGHER IS -X2, -X3, +X9, +X4, -X5, +X8, -X1

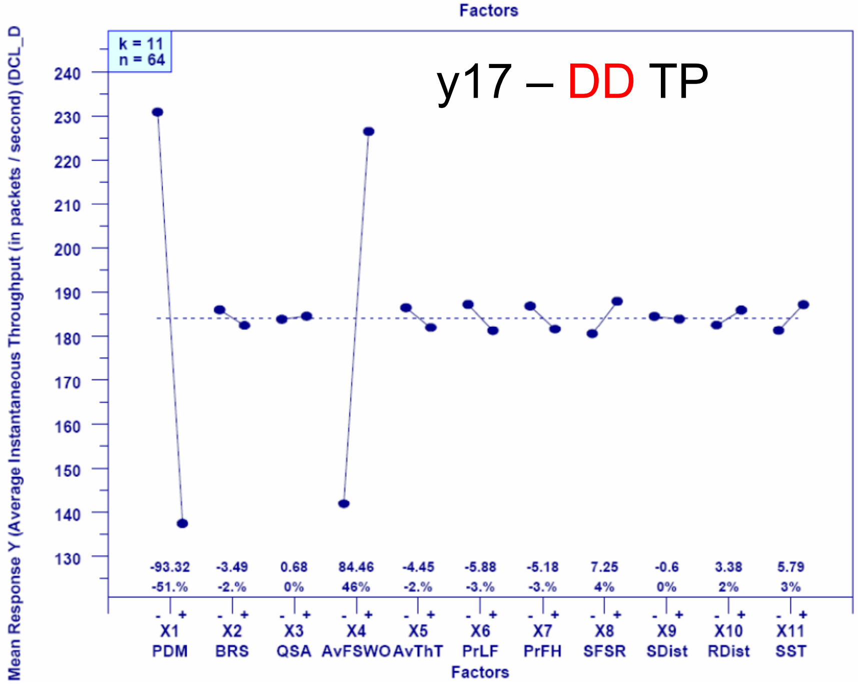
PC3



HIGHER TP: -X1, +X9, +X2, +X5, +X4

PC- IS: +X4, +X2, -X1, +X9, +X3

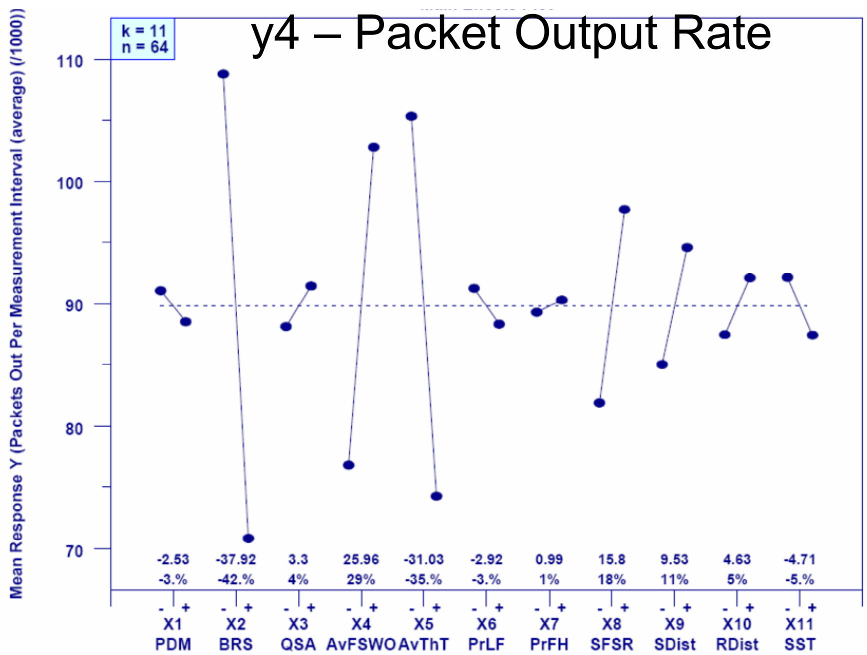




I THINK D-class & F-class THROUGHPUT SHOULD BE SEPARATE – ONLY TWO INPUT FACTORS INFLUENCE D-class THROUGHPUT

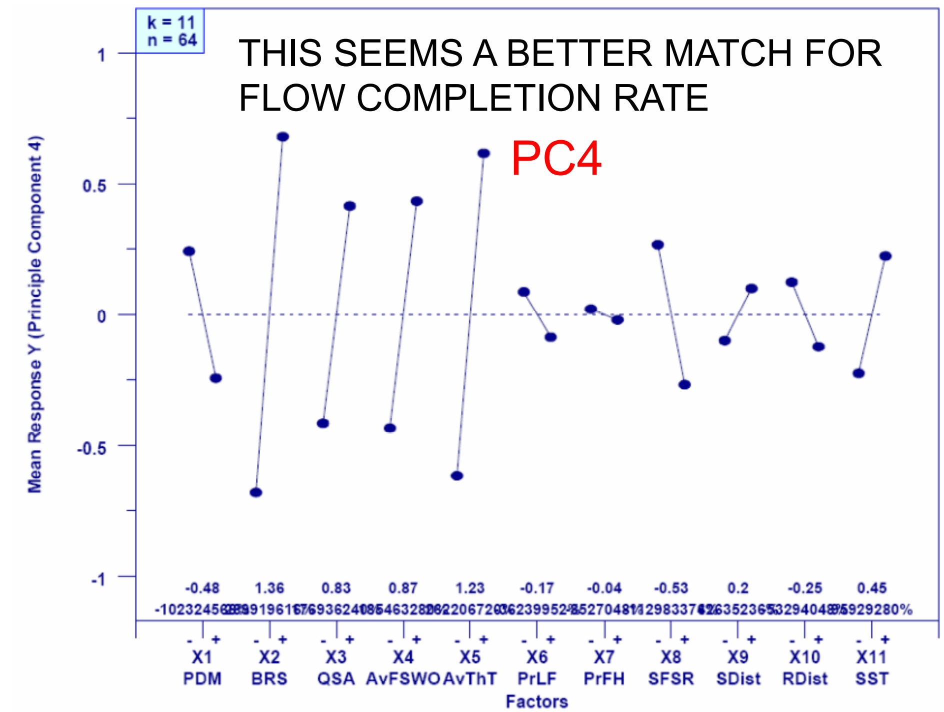
HIGHER TP IS –X1, +X4

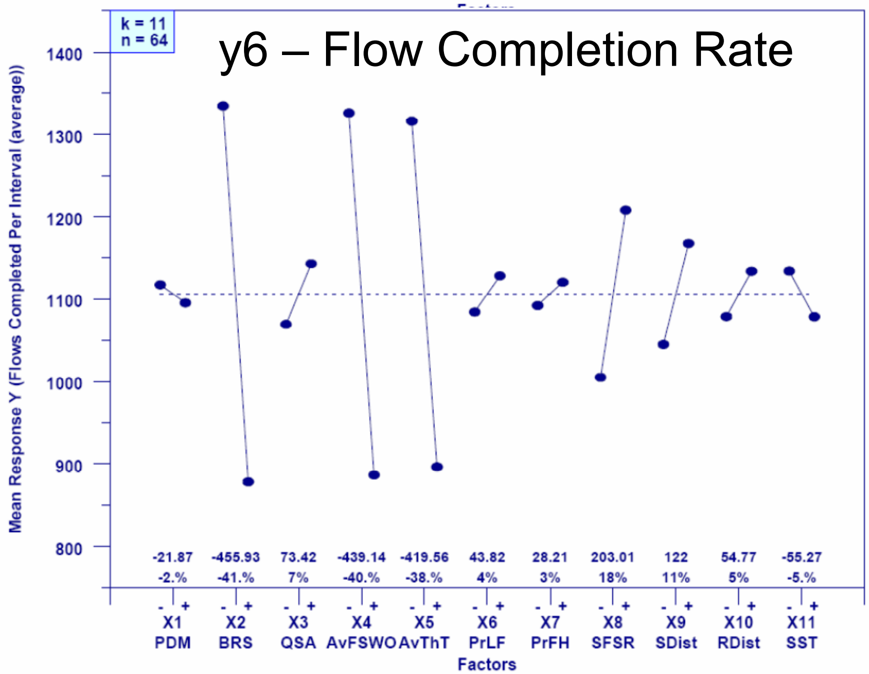
PC4



HIGHER PO: +X2, -X5, +X4

PC- IS: +X2, -X5, -X4, -X3





I THINK FLOW COMPLETE RATE & PACKET THROUGHPUT RATE SHOULD BE KEPT SEPARATE BECAUSE FLOW COMPLETE IS HIGHER WITH SMALL FILE SIZE & PACKET OUTPUT IS HIGHER WITH LARGE FILE SIZE

HIGHER FC IS +X2, -X4, -X5

Note: The Domain Analyst Sides With the Correlation Analysis Results

Dimension	Definition
1	Congestion
2	Loss
3	Delay
4	Throughput for the most advantaged users
5	Throughput for the somewhat advantaged users
6	Network-wide Packet Throughput
7	Network-wide Flow Throughput

Pros & Cons of the 2 Dimension Reduction Techniques

Pros/Cons of Correlation Analysis & Clustering

Pros

- Provided effective dimension reduction (22 → 7) through correlations that could be vetted by a domain expert
- Examining response correlations helped to validate *MesoNet*
- Uncovered nuanced differences between flow and packet throughput rates in a network

Cons

- A second 2^{11-5} OFF experiment with different level settings revealed some (valid) differences in correlations – thus separate correlation analyses must be conducted for different level settings

Pros/Cons of Principal Components Analysis

Pros

- Provided greater dimension reduction (22 → 4) than correlation analysis & clustering

Cons

- There is no specific domain interpretation of even the top 2 or 3 principal components – in the case shown here we were able to arrive at a reasonable interpretation; in other cases, we were not
- Principal components take on + and – values, which present domain analysts with difficulty assigning meaning – we had to infer meaning of components by comparing them with meaning derived from analyzing individual responses
- Principal components proved coarser than corresponding groupings generated by clustering mutual correlations
- A second 2^{11-5} OFF experiment with different level settings revealed some differences in principal components – such differences are difficult to understand without assistance from other analyses

Summary: Correlation Analysis or PCA?

- If limited to one technique, correlation analysis provides results easier for a domain analyst to comprehend
- Principal components take on + and – values, which present domain analysts with difficulty assigning meaning – we had to infer meaning by comparing main effects plots of principal components with main effects plots from responses chosen from groupings established by correlation analysis
- Principal components proved coarser than corresponding groupings generated by clustering mutual correlations
- PCA provides a reasonable complement to correlation analysis by giving a separate view of the data, which should be consistent with correlation results, thus helping to validate a model

MesoNet Conclusions

- We investigated correlation and PC analyses as two techniques to reduce the dimension of responses from *MesoNet*, a network simulator
- We demonstrated that both techniques can significantly reduce the dimension of response data
- We also showed that both techniques could be used to validate a model, but that PCA is more suited as a complement to correlation analysis
- We found that PCA results are difficult for a domain analyst to interpret without comparison to analyses of individual responses
- We also found that results from correlation and PC analyses with one set of parameter values cannot necessarily be extrapolated to a different set of values

Stat Conclusions

1. Stat Framework/Approach & Methodology: Demo beginning-to-"end"

2. Critical importance of domain expert

3. Dimension Reduction dependency on DEX & Sensitivity Analysis

4. Internet Modeling Conclusions & Insight

Methodology Applications

1. MesoNet Analysis #1 (k=11,n = 64,m=22 ➔ 7)
   Sensitivity & Dimension-Reduction Analysis <today's talk>

2. MesoNet Analysis #2 (k=20,n=256,m=22)
   Sensitivity Analysis

3. MesoNet TCP Congestion/Control Alg. Comparison (k=6,n=32)(5)

4. Cloud Computing Analysis (k=11,n=64,m=42 => 8) (Koala)
   Sensitivity & Dimension-Reduction Analysis

5. Cloud Computing VM Placement Alg. Comparison (k=6,n=32)
   (Koala)

Graphical Methods

1. Main Effects Plots
2. Interaction Effects Matrix
3. Ordered Data Plots
4. Pairwide Scatter Plot Matrix (Unordered)
5. Pairwise Scatter Plot Matrix (Ordered)
6. Stacked Main Effects Plot
7. Multiplot of (1-Way) ANOVA CDF Values
8. Index-Index Cluster Plot
9. Character Plots
10. PCA Weights Plot

Presentations

J. Filliben, "Sensitivity Analysis Methodology for a Complex System Computational Model", 39th Symposium on the Interface: Computing Science and Statistics, Philadelphia, PA, May 26, 2007.

K. Mills and J. Filliben, "An Efficient Sensitivity Analysis Method for Mesoscopic Network Models", Complex Systems Study Group, NIST, February 2, 2010.

K. Mills and J. Filliben, "Comparing Two Dimension-Reduction Methods for Network Simulation Models", Winter Simulation Conference (WSC 2010), Baltimore, Maryland, Dec. 6, 2010.

K. Mills and J. Filliben, "Using Sensitivity Analysis to Identify Significant Parameters in a Network Simulation", Winter Simulation Conference (WSC 2010), Baltimore, Maryland, Dec. 6, 2010.

K. Mills, J. Filliben, D.-Y. Cho and E. Schwartz, "Predicting Macroscopic Dynamics in Large Distributed Systems", LSN Seminar on Complex Networks and Information Systems, Gaithersburg, Maryland, June 30, 2011.

K. Mills, J. Filliben and C. Dabrowski, "An Efficient Sensitivity Analysis Method for Large Cloud Simulations", IEEE Cloud 2011, Washington, D.C., July 8, 2011.

K. Mills, J. Filliben, D.-Y. Cho and E. Schwartz, "Predicting Macroscopic Dynamics in Large Distributed Systems", American Society of Mechanical Engineers
2011 Conference on Pressure Vessels & Piping, Baltimore, MD, July 21, 2011.

References

K. Mills, “Measurement Science for Complex Information Systems”, NIST/ITL Web Page for the Complex Systems Project:
http://www.nist.gov/itl/antd/emergent_behavior.cfm

K. Mills, J. Filliben, D. Cho, E. Schwartz and D. Genin, "Study of Proposed Internet Congestion Control Mechanisms“, NIST Special Publication 500-282, May 2010, 534 pages. http://www.nist.gov/itl/antd/Congestion_Control_Study.cfm

K. Mills, J. Filliben and C. Dabrowski, "An Efficient Sensitivity Analysis Method for Large Cloud Simulations“, Proceedings of the 4th International Cloud Computing Conference, IEEE, Washington, D.C., July 5-9, 2011.

K. Mills, J. Filliben, D-Y. Cho and E. Schwartz, "Predicting Macroscopic Dynamics in Large Distributed Systems“, Proceedings of ASME 2011 Conference on Pressure Vessels & Piping, Baltimore, MD, July 17-22, 2011.

K. Mills and J. Filliben, "Comparison of Two Dimension-Reduction Methods for Network Simulation Models“, Journal of Research of the National Institute of Standards and Technology, 116-5, September-October 2011, in press.

K. Mills, J. Filliben and C. Dabrowski, "Comparing VM-Placement Algorithms for On-Demand Clouds“, (submitted to IEEE CloudCom 2011, under review.

Web References

Complex Systems Project
http://www.nist.gov/itl/antd/emergent_behavior.cfm

NIST SP 500-282 (534 pages)
http://www.nist.gov/itl/antd/Congestion_Control_Study.cfm

NIST/SEMATECH Engineering Statistics Handbook
http://www.itl.nist.gov/div898/handbook/

Dataplot
http://www.itl.nist.gov/div898/software/dataplot/

This Talk
http://stat.nist.gov/~filliben/fillibenmillsnistsedtalk092211.pdf
http://www.nist.gov/itl/antd/upload/millsjjfsedtalk092211.pdf