

South Miami-Dade Exploratory Statistical Analysis Report

South Miami-Dade Water Issues Coordination Roundtable Discussion – March 1, 2011

**Kevin A. Rodberg, Science Supervisor
Resource Evaluation Division
Intergovernmental Programs Department**

Exploratory Statistical Analysis Report

Title of Report:

South Miami-Dade Statistical Data Analyses

Prepared for: SFWMD

Prepared by: Eric P. Smith

Purpose:

Develop a tool to provide insights into relationships among and between hydrologic and water quality data for the region

Exploratory Statistical Analysis Report (Cont.)

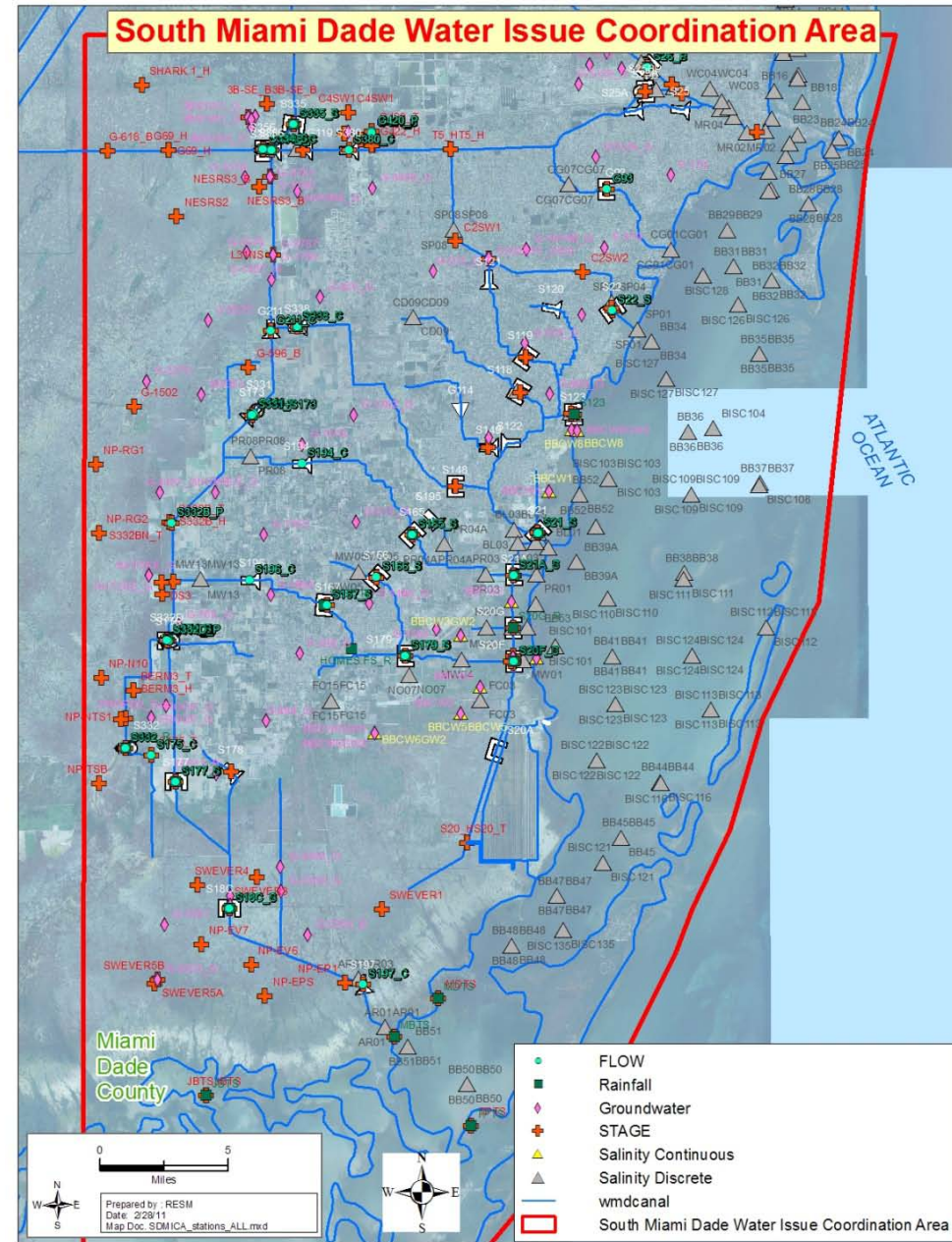
This report is the 3rd part of a three-phase study

1. Compilation of Data
2. Data Scrub
3. Exploratory Analysis

Scope of Study

Project Area is located:

- SE Miami-Dade County
- South of the C-4 canal
- East of the L31W levee
- Includes SFWMD water within Florida and Biscayne Bay



Exploratory Statistical Analysis Report (Cont.)

Source of Data for Statistical Analysis

- SFWMD, USGS, DERM and Biscayne National Park databases
- Data was 'scrubbed' under a separate contract
 - Eliminate anomalous data points and
 - Fill in data gaps with estimated data where appropriate

Period of Record – 01/01/2000 through 12-31-2009

Exploratory Statistical Analysis Report (Cont.)

Types of data analyzed:

Daily Timeseries

Groundwater Levels (168 stations)

Surface Water Stage (154 stations)

Canal/Structure Flow (60 stations)

Rainfall Gage Recorders (29 stations)

Discrete (random) and Continuous (15 minute interval)

Water Quality (182 Salinity stations)

Exploratory Statistical Analysis Report (Cont.)

The report is a tool to aid in scientific investigation of data relationships and trends in the area.

It provides insights into relationships among and between hydrologic and water quality data for the region.

The report should not be considered an “Answer Machine”

Exploratory Statistical Analysis Report (Cont.)

Report includes the computation of data correlations to aid in understanding the relationships in the data

A simple description or definition of Correlation Coefficients:

- A measure of the relationship between two variables
- Correlation coefficients are useful because they can indicate a predictive relationship in data sets

Exploratory Statistical Analysis Report (Cont.)

Types of Correlations calculated for this Report:

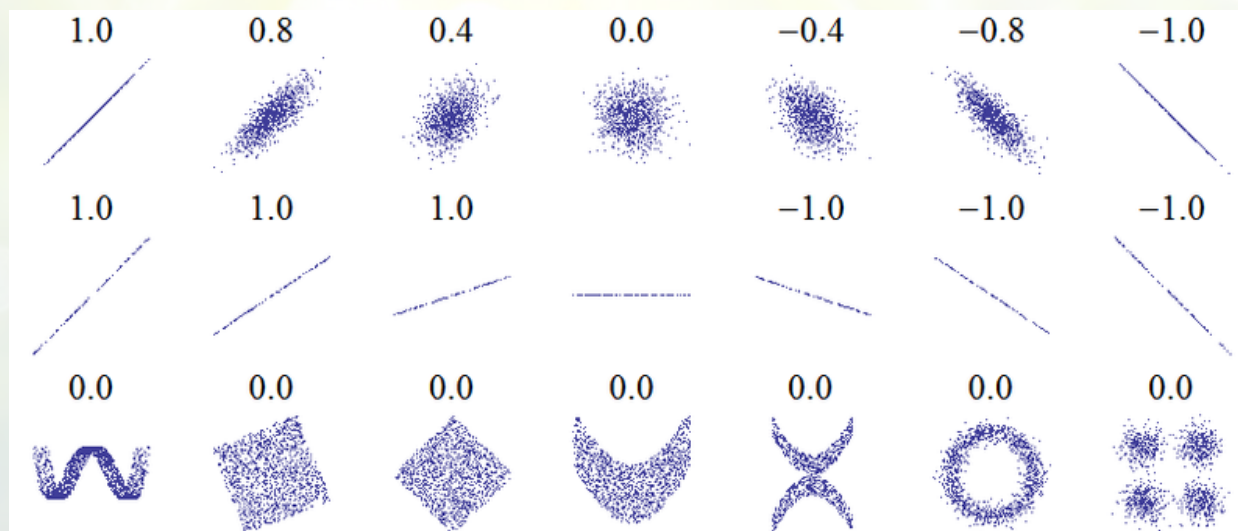
- **Pearson:**
Traditional Correlation
Good for Linear and Normal Distributions
- **Spearman:**
A Rank Correlation and doesn't require a normal distribution
Not as sensitive to odd observations or outliers
- **Kendall:**
Also a Rank or Order based Correlation
Good for monotonic relationships between series

Rank Correlations:

Ranks are computed and correlations between the ranks are calculated

Exploratory Statistical Analysis Report (Cont.)

Examples plots of Pearson's Correlation Coefficients

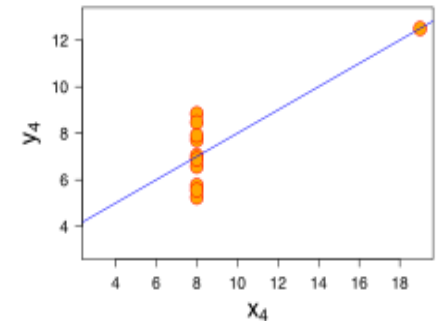
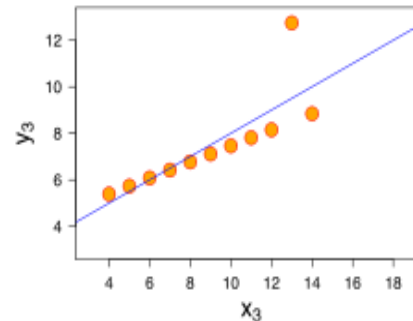
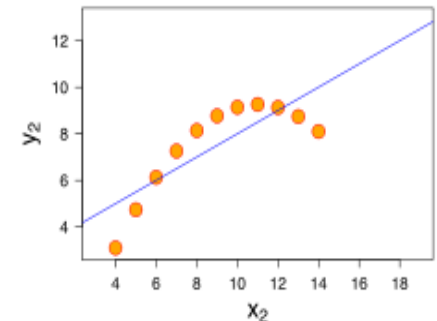
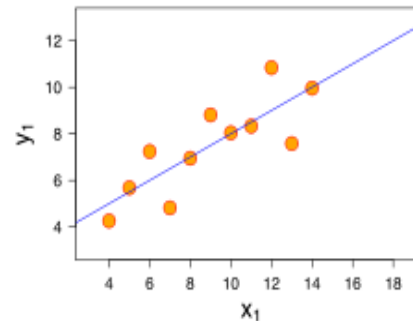


The correlation coefficients (Pearson's) reflect the noisiness and direction of a linear relationship. Not slope or non-linear variation.

Exploratory Statistical Analysis Report (Cont.)

Pearson's Correlation Coefficient and Linearity

1. Pearson's is good with a normal or linear distribution
2. Not distributed normally and not linear
3. Distribution is linear, but has a different regression line – one outlier changes the regression.
4. One outlier is enough to produce a high correlation



Each graph shows the same basic statistics: Mean (.75), Variance (4.12), Correlation (0.816), and Regression Line ($y = 3 + 0.5x$)

Exploratory Statistical Analysis Report (Cont.)

Analyses Included in the Report:

Correlation for stations with

(Stage, Groundwater, Rain, Flow, Salinity)

- Same data type
- Different types
- Full period of record for the study
- Wet season & dry season

Lag correlations

No lag, and 1 to 5 days. (i.e.: Lag 1 is the previous day)
between variables of the same type or different types

Neighboring station relationships

Correlations by station and type for each station with the six closest in distance

Exploratory Statistical Analysis Report (Cont.)

Principal Components Analysis(PCA)/Factor Analysis

Helps summarize data and relationships between variables.

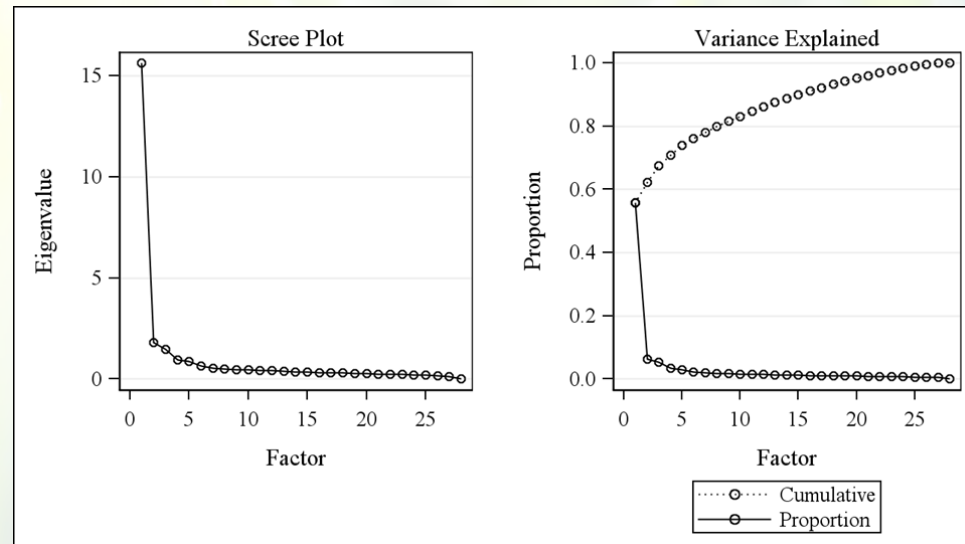
Scree plots :

Shows eigenvalues vs. eigenvectors or number of factors

Variance Explained plot:

Displays the proportion of total variance for each component and the cumulative proportion of the variance explained.

The PCA use the number of factors with eigenvalues above 1 .(Look for the elbow)



Exploratory Statistical Analysis Report (Cont.)

Principal Components Analysis(PCA)/Factor Analysis

Shows the sets of variables identified by the rotated factor pattern of rain data.

Variables with similar values grouped or flagged with a star

Rotated Factor Pattern				
	Factor1		Factor2	Factor3
S20G_Rain	77 *		27	23
S20F_Rain	77 *		28	22
S179_Rain	76 *		30	24
S21A_Rain	75 *		26	20
S21_Rain	72 *		35	24
S20_Rain	71 *		23	22
HOMESFS_rain	71 *		35	26
S165_Rain	71 *		40	21
S167_Rain	70 *		38	24
S18C_Rain	69 *		29	18
S177_Rain	67 *		38	17
MDTS_rain	67 *		18	15
tpts_rain	67 *		14	16
S332_Rain	63 *		38	19
S174_Rain	62 *		45	21
JBTS_rain	62 *		21	17
S123_rain	61 *		39	23
S334_rain	29		88 *	20
S335_rain	29		88 *	20
S336_Rain	36		83 *	21
S338_Rain	48 *		65 *	22
S331W_rain	53 *		62 *	23
S331_Rain	48 *		60 *	22
S26_rain	47		48 *	28
miamai_ap_rain	22		21	82 *
TAMI_AIR_rain	22		19	82 *
PERRINE_Rain	27		22	81 *
miami_2_Rain	25		15	77 *
Printed values are multiplied by 100 and rounded to the nearest integer. Values greater than 0.474122 are flagged by an '*'.				

- Bi-plots summarizes sites and times in one display
- Useful to identify sites with similar and extreme measurement patterns

[illegible]

Exploratory Statistical Analysis Report (Cont.)

Correlation Matrix for Rain Stations

rain_Spearman1.csv

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE
1	_TYPE	_NAME	HOMES	JBTS_r	MBTS_r	MDTS_r	PERRINE_R	S123_r	S165_Rain	S167_Rain	S174_Rain	S177_Rain	S179_Rain	S18C_Rain	S20F_Rain	S20G_R	S20_R	S21A_R	S21_R	S26_r	S331w_r	S331_R	S332_R	S334_ra	S335_ra	S336_R	S338_R	TAMI_A	miamai	miami_2	tps_rain
2	MEAN		0.1405	0.0843	0.0982337	0.1077018	0.161922296	0.1186329	0.135207	0.135413	0.13992	0.12385	0.144065	0.164078	0.120631	0.12724	0.1272	0.1184	0.1097	0.1551	0.1568	0.1249	0.1309	0.1667	0.1667	0.136	0.132	0.1429	0.173	0.1426	0.1187
3	STD		0.4932	0.3518	0.389512	0.4264105	0.483622783	0.4228182	0.451254	0.436966	0.42764	0.41217	0.510214	0.537026	0.391404	0.42663	0.4579	0.4116	0.3864	0.471	0.4967	0.4132	0.4059	0.5079	0.5079	0.41	0.4059	0.4458	0.5119	0.4443	0.4353
4	N		3433	3489	1703	3481	3449	3577	3634	3640	3563	3645	3488	3627	3646	3634	2818	3609	3623	3624	3626	3069	3627	3596	3596	3607	3599	3583	3640	3493	3342
5	CORR	HOMES	1	0.4952	0.5155851	0.5142516	0.472795454	0.5956339	0.698253	0.759515	0.64944	0.65154	0.776439	0.610345	0.66724	0.66852	0.6042	0.6382	0.6534	0.5532	0.6415	0.5796	0.6256	0.554	0.554	0.5901	0.6058	0.4344	0.4324	0.4127	0.4984
6	CORR	JBTS_r	0.4952	1	0.5948848	0.5402961	0.351353638	0.4419052	0.510973	0.494179	0.46559	0.50705	0.492068	0.531616	0.507468	0.52379	0.5212	0.4856	0.5036	0.4514	0.4719	0.4441	0.4987	0.4339	0.4339	0.4574	0.4678	0.3068	0.3221	0.341	0.5082
7	CORR	MBTS_r	0.5166	0.5949	1	0.6590065	0.348704523	0.4781809	0.486895	0.525464	0.46053	0.50358	0.534498	0.550152	0.525991	0.51972	0.5611	0.5309	0.5272	0.4527	0.4758	0.418	0.477	0.4452	0.4452	0.4749	0.4702	0.3186	0.3448	0.3438	0.637
8	CORR	MDTS_r	0.5143	0.5403	0.6590065	1	0.341202415	0.4697776	0.5036	0.521832	0.48725	0.50524	0.518806	0.529319	0.545038	0.52818	0.5388	0.4958	0.5111	0.4458	0.4678	0.4495	0.468	0.442	0.442	0.4602	0.4621	0.2987	0.3159	0.3424	0.6004
9	CORR	PERRINE_R	0.4728	0.3514	0.3487045	0.3412024	1	0.4251644	0.459387	0.468287	0.43562	0.40406	0.456237	0.405552	0.444393	0.4596	0.4158	0.4089	0.459	0.4203	0.4805	0.4681	0.4201	0.4265	0.4265	0.4301	0.4495	0.7407	0.6801	0.6109	0.333
10	CORR	S123_r	0.5956	0.4419	0.4781809	0.4697776	0.425164446	1	0.662384	0.612094	0.56341	0.54447	0.617076	0.514744	0.631672	0.63671	0.556	0.6223	0.6885	0.5736	0.6086	0.5598	0.5115	0.5473	0.5473	0.5746	0.5975	0.3878	0.4015	0.3833	0.4538
11	CORR	S165_R	0.6983	0.511	0.4868949	0.5036004	0.459386797	0.6623839	1	0.723992	0.63889	0.61408	0.722959	0.580738	0.688517	0.70833	0.5842	0.6985	0.7315	0.5732	0.6614	0.6328	0.5939	0.5728	0.5728	0.6116	0.6408	0.4012	0.402	0.3815	0.4853
12	CORR	S167_R	0.7595	0.4942	0.5254644	0.521892	0.462826842	0.6120938	0.723992	1	0.68252	0.64782	0.743127	0.608217	0.678717	0.66904	0.591	0.6341	0.6686	0.5498	0.6634	0.6271	0.6195	0.5712	0.5712	0.5977	0.6227	0.4328	0.4239	0.412	0.4988
13	CORR	S174_R	0.6494	0.4656	0.4605303	0.4872488	0.435620698	0.563411	0.638885	0.682519	1	0.69319	0.642427	0.592347	0.616061	0.60667	0.5645	0.5715	0.6071	0.5176	0.6601	0.6198	0.6814	0.5708	0.5708	0.6058	0.6429	0.4121	0.4045	0.3698	0.476
14	CORR	S177_R	0.6515	0.507	0.5035758	0.5052358	0.404061016	0.5444696	0.614081	0.647817	0.69319	1	0.650428	0.649727	0.598398	0.59935	0.5674	0.5701	0.5928	0.5234	0.6186	0.5783	0.7404	0.5395	0.5395	0.572	0.5907	0.3778	0.3706	0.3595	0.493
15	CORR	S179_R	0.7764	0.4921	0.5344975	0.5188059	0.456236857	0.6170764	0.722959	0.743127	0.62423	0.65043	1	0.603669	0.740855	0.73311	0.6242	0.6931	0.7019	0.5676	0.6501	0.5859	0.6086	0.5349	0.5349	0.579	0.6028	0.4258	0.4299	0.4193	0.5238
16	CORR	S18C_R	0.6103	0.5316	0.5501522	0.529319	0.405551633	0.5147439	0.580738	0.608217	0.59235	0.64973	0.603669	1	0.58745	0.58725	0.5998	0.5577	0.5675	0.5029	0.5842	0.5327	0.6267	0.4904	0.4904	0.5121	0.5511	0.3639	0.3595	0.3508	0.5056
17	CORR	S20F_R	0.6672	0.5075	0.5259911	0.5450381	0.444939188	0.6316716	0.688517	0.678717	0.61606	0.5984	0.740855	0.58745	1	0.80299	0.635	0.7083	0.7031	0.5512	0.6166	0.5505	0.5645	0.5341	0.5341	0.5661	0.5892	0.3882	0.4106	0.4049	0.5239
18	CORR	S20G_R	0.6685	0.5238	0.5191716	0.5281755	0.459602094	0.6367124	0.708332	0.669045	0.60667	0.59935	0.733112	0.587255	0.80299	1	0.6123	0.7364	0.7288	0.5624	0.6072	0.5543	0.5713	0.5302	0.5302	0.5616	0.587	0.3972	0.4194	0.4099	0.5185
19	CORR	S20_R	0.6042	0.5212	0.5610792	0.5989187	0.415811102	0.5559656	0.584189	0.591055	0.56449	0.56744	0.624211	0.599821	0.634999	0.61226	1	0.5646	0.6159	0.4934	0.5703	0.5486	0.5186	0.479	0.479	0.5081	0.5282	0.3909	0.3685	0.3795	0.5639
20	CORR	S21A_R	0.6382	0.4856	0.530962	0.4958469	0.408934158	0.62222819	0.698492	0.634077	0.57151	0.57005	0.693117	0.55773	0.708344	0.73643	0.5646	1	0.7538	0.5244	0.5909	0.5357	0.5504	0.4972	0.4972	0.5408	0.5681	0.3766	0.3844	0.3774	0.5148
21	CORR	S21_R	0.6534	0.5036	0.5271751	0.5111433	0.458984481	0.6885454	0.731488	0.686636	0.6071	0.59278	0.701942	0.56751	0.703095	0.72881	0.6159	0.7538	1	0.5818	0.6352	0.5945	0.562	0.5544	0.5544	0.6015	0.6286	0.4153	0.4146	0.4158	0.5119
22	CORR	S26_r	0.5532	0.4514	0.4526818	0.4457968	0.420312403	0.5736029	0.573151	0.549833	0.51759	0.52341	0.56761	0.502893	0.551163	0.56238	0.4934	0.5244	0.5818	1	0.5722	0.5368	0.4991	0.5838	0.5838	0.6025	0.5679	0.3958	0.4505	0.4084	0.4277
23	CORR	S331w_r	0.6415	0.4719	0.4757696	0.4676039	0.480519408	0.608621	0.661371	0.663398	0.68006	0.6186	0.650146	0.584245	0.616953	0.60721	0.5703	0.5909	0.6352	0.5722	1	0.842	0.605	0.651	0.651	0.8324	0.7582	0.4277	0.4284	0.393	0.4617
24	CORR	S331_R	0.5796	0.4441	0.4719612	0.4449454	0.4661244	0.5598084	0.632837	0.627057	0.61892	0.57835	0.588594	0.532748	0.550475	0.55434	0.5486	0.5357	0.5945	0.5368	0.842	1	0.5612	0.6023	0.6023	0.635	0.707	0.4046	0.4028	0.3636	0.4138
25	CORR	S332_R	0.6256	0.4987	0.4763923	0.4680012	0.420120249	0.5114922	0.593863	0.619549	0.6814	0.74037	0.608557	0.626745	0.564524	0.57133	0.5186	0.5504	0.562	0.4991	0.605	0.5612	1	0.5231	0.5231	0.5539	0.5744	0.3783	0.3749	0.3588	0.4541
26	CORR	S334_ra	0.554	0.4339	0.4452035	0.4419738	0.426462491	0.5472944	0.57285	0.571153	0.57079	0.53952	0.534898	0.490382	0.534085	0.5302	0.479	0.4972	0.5544	0.5838	0.651	0.6023	0.5231	1	1	0.8857	0.667	0.392	0.41	0.3854	0.4065
27	CORR	S335_ra	0.554	0.4339	0.4452035	0.4419738	0.426462491	0.5472944	0.57285	0.571153	0.57079	0.53952	0.534898	0.490382	0.534085	0.5302	0.479	0.4972	0.5544	0.5838	0.651	0.6023	0.5231	1	1	0.8857	0.667	0.392	0.41	0.3854	0.4065
28	CORR	S336_R	0.5901	0.4574	0.44747532	0.4602189	0.430122734	0.5745633	0.611555	0.597683	0.60581	0.57198	0.578991	0.512053	0.566149	0.56158	0.5081	0.5408	0.6015	0.6025	0.6324	0.635	0.5539	0.8857	0.8857	1	0.7158	0.4117	0.4229	0.3927	0.451
29	CORR	S338_R	0.6058	0.4678	0.470239	0.4621309	0.449520524	0.5975113	0.64078	0.622736	0.64289	0.5907	0.60282	0.551144	0.589209	0.58703	0.5282	0.5681	0.6286	0.5679	0.7582	0.707	0.5744	0.667	0.667	0.7158	0.41	0.4172	0.4206	0.3802	0.4405
30	CORR	TAMI_A	0.4344	0.3068	0.3185655	0.2987128	0.274069109	0.3878131	0.401229	0.432759	0.41211	0.37783	0.425791	0.363944	0.388228	0.3972	0.3909	0.3766	0.4153	0.3958	0.4277	0.4046	0.3783	0.392	0.392	0.4117	0.4172	1	0.665	0.5765	0.314
31	CORR	miamai	0.4324	0.3221	0.3448314	0.3158806	0.68014704	0.4015208	0.40448	0.423899	0.40448	0.37062	0.429918	0.359505	0.410642	0.39139	0.3685	0.3844	0.4146	0.4505	0.4284	0.4028	0.3749	0.41	0.41	0.4229	0.4206	0.685	1	0.663	0.3041
32	CORR	miami_2	0.4127	0.341	0.3438269	0.3424447	0.610943855	0.3833046	0.381524	0.411993	0.36982	0.35947	0.419268	0.35085	0.404864	0.40993	0.3795	0.3774	0.4158	0.4064	0.393	0.3636	0.3588	0.3854	0.3854	0.3927	0.3802	0.5765	0.663	1	0.3256
33	CORR	tps_rain	0.4984	0.5082	0.6370066	0.6003568	0.332954189	0.4537865	0.485298	0.499804	0.47603	0.49304	0.523761	0.505578	0.523907	0.51852	0.5639	0.5148	0.5119	0.4277	0.4617	0.4138	0.4541	0.4065	0.4065	0.451	0.4405	0.314	0.3041	0.3256	1

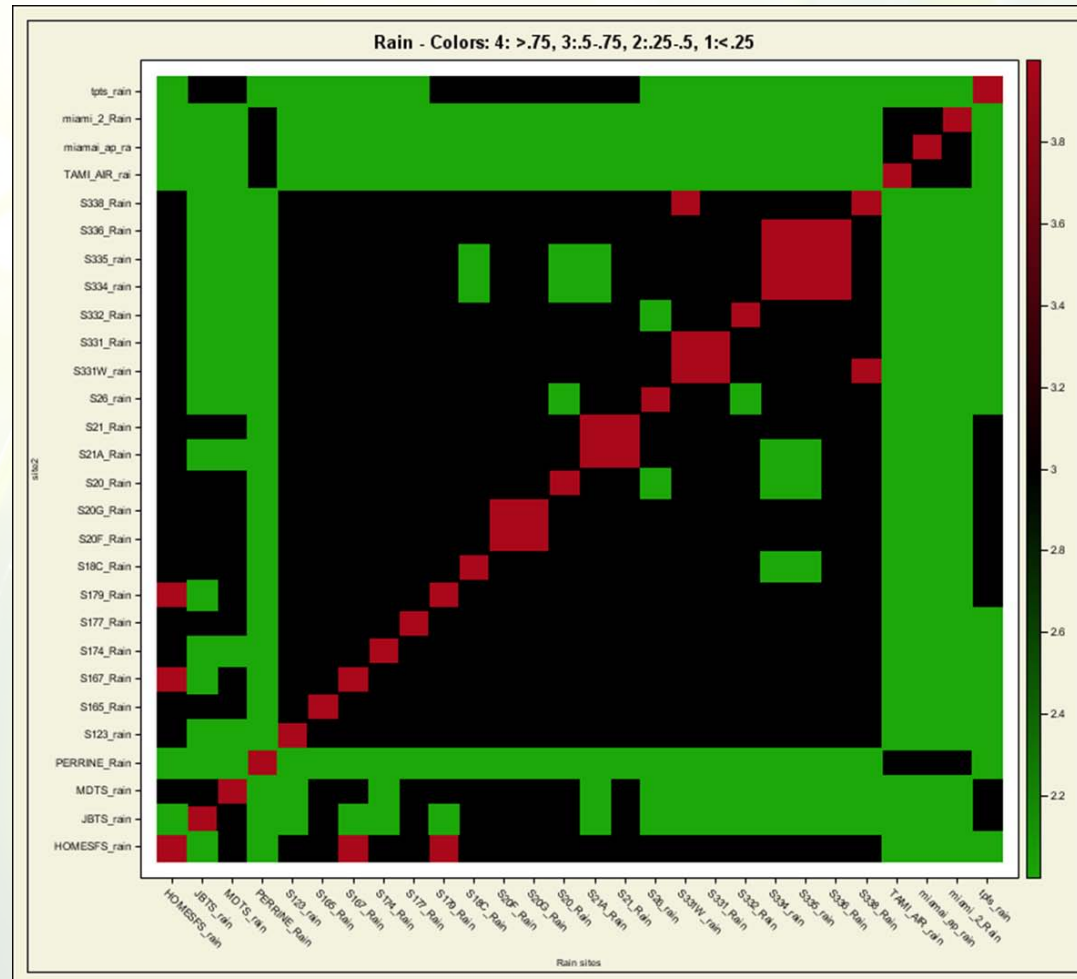
Exploratory Statistical Analysis Report (Cont.)

Heat Maps

Help summarize relationships of data by representing the size correlations in groups of colors on and off the matrix diagonal

Values of high correlation on the diagonal are comparing the same variables

High values off the diagonal are variables to look at



Exploratory Statistical Analysis Report (Cont.)

Regression Models (stepwise regression procedure)

DEPVAR: dependent variable

RMSE: square root of the mean square root

Intercept: gives the value of the intercept in the model

IN: Independent variables

P: number of parameters

EDF: Degrees of freedom

RSQ: R-squared of the model

RegsN_3500A.csv

	A	B	C	D	E	F	G	H	T	U	Y	CZ	DA	DB	DC
1	DEPVAR	RMSE_	Intercept	MDTS_S	S175_H	S175_T	S177_H	S177_T	SWEVER1	ANGEL_W	S178_C_Flow	IN_	_P_	_EDF_	_RSQ_
2	MBTS_Stage_J	0.03734	2.174185										5	6	3636 0.992055
3	s179_T_Stage_D	0.03476	0.027827										5	6	3636 0.986365
4	S175_H_Stage_J	0.047687	-0.03538		-1	0.06373							5	6	3636 0.99827
5	S175_T_Stage_J	0.096894	-0.07807		0.20697	-1		0.5313					5	6	3636 0.97294
6	S177_H_Stage_J	0.152644	0.014628		0.25106		-1						5	6	3636 0.91181
7	S177_T_Stage_J	0.064273	0.017561		-0.0331			-1					5	6	3636 0.979658
19	SWEVER1_Stage_J	0.103752	-0.08663	0.1204	-0.0707				-1				5	6	3636 0.912485
20	ANGEL_WELL_km	0.248584	0.023645		0.31699					-1			5	6	3636 0.935607
26	S178_C_Flow_SK	11.56409	7.990156								-1		5	6	3636 0.869254
27	S178_S_Flow_SK	7.181994	-1.78172								0.572281882		5	6	3636 0.87408
28	S20G_S_Flow_SK	32.52369	-43.158	22.129									5	6	3636 0.604927
29	S25_C_Flow_SKRE	15.27393	15.18167										5	6	3636 0.494211
30	_3B_SE_B_Stage_C	0.183141	-0.71708										5	6	3636 0.964557
31	G334_H_Stage_D	0.238433	3.340191										5	6	3636 0.761829
32	G334_T_Stage_D	0.04544	0.005767										5	6	3636 0.99415
33	G93 H Stage D	0.184288	-0.01013							0.101852			5	6	3636 0.755543

Exploratory Statistical Analysis Report (Cont.)

Example analysis: S20F flow and salinity

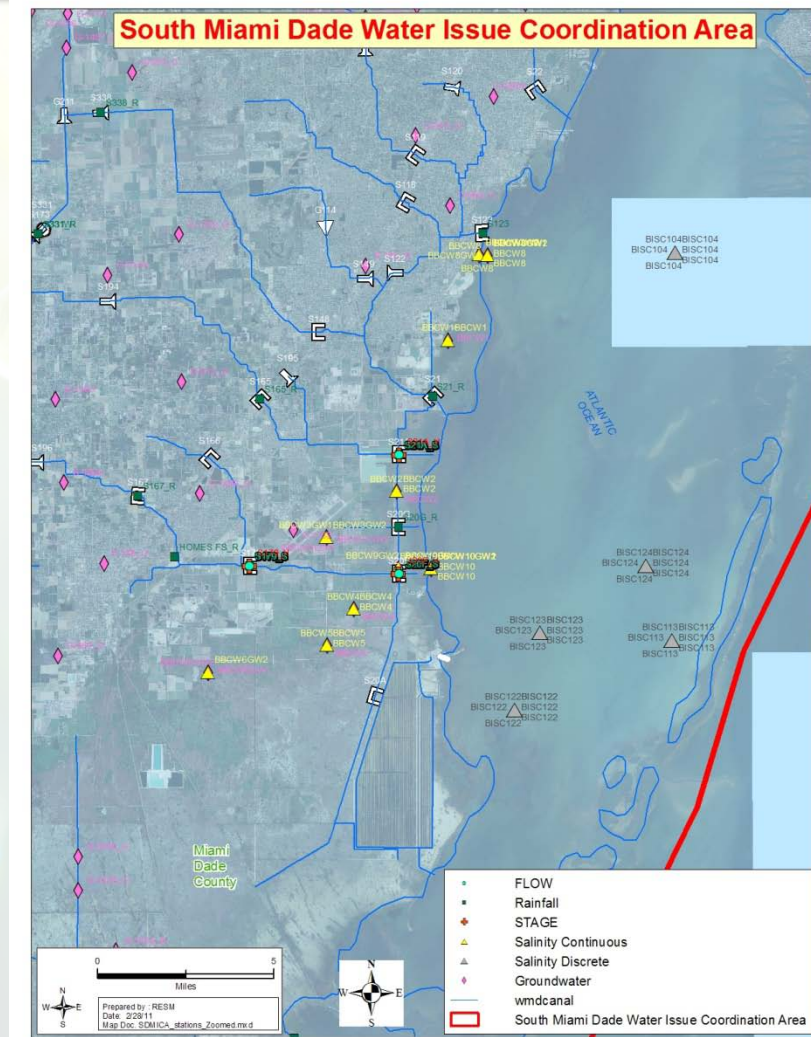
- Illustrates how the report can be used to explore data relationships
- Important factors to consider when interpreting the data relationships
- The example shows some of the complications the analysis of a simple question may need to provide a very general answer.

“Does flow out of S20F cause changes in salinity, stage and water levels nearby?”

Exploratory Statistical Analysis Report (Cont.)

Map of some of the stations for the Example Analysis:

Flow: S20F_S, S21A_S, S179_S
 Well: BBCW6 (BBCW6GW1, BBCW6GW2), G3356
 Stage: S20F_H, S21A_H, S179_H
 Salinity: BISC 124, BISC 123, BISC 122, BISC113, BISC16BOT, BISC24BOT, BISC34BOT, BISC14BOT, BISC32BOT, BISC22BOT, BISC129, BISC104



Exploratory Statistical Analysis Report (Cont.)

Answer:

- The analysis suggests there is a relationship between flow at S20F and salinity levels at nearby stations
- The relationship is strongest at the stations closer to S20F
- However the strength of the relationship is difficult to assess as there may be other factors involved
- This is suggested by the fact that salinity at stations farther from S20F are also correlated with flow at S20F, although the correlation is weaker



Questions?

Links to Information and Data:

[SFWMD Web site: Link to Miami-Dade Service Center Page](#)

[Map-based Interface to South Miami-Dade Water Issues Documents and Data](#)