Spatial Analytics w/ SAS

Exploring Parking Violations and Spatial Modeling in the District of Columbia

Agenda

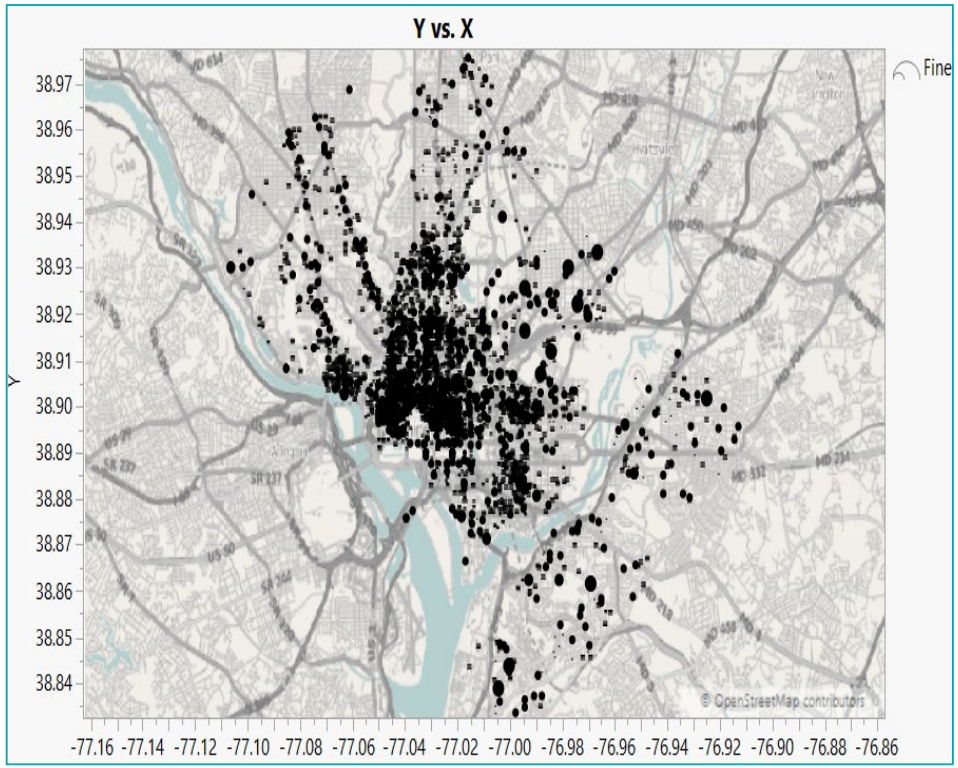
- Exploratory Data Analysis
- Empirical and Directional Variograms
- Kriging
- Regression
- Generate & Test Multiple Models
- Summary and Conclusion

Agenda

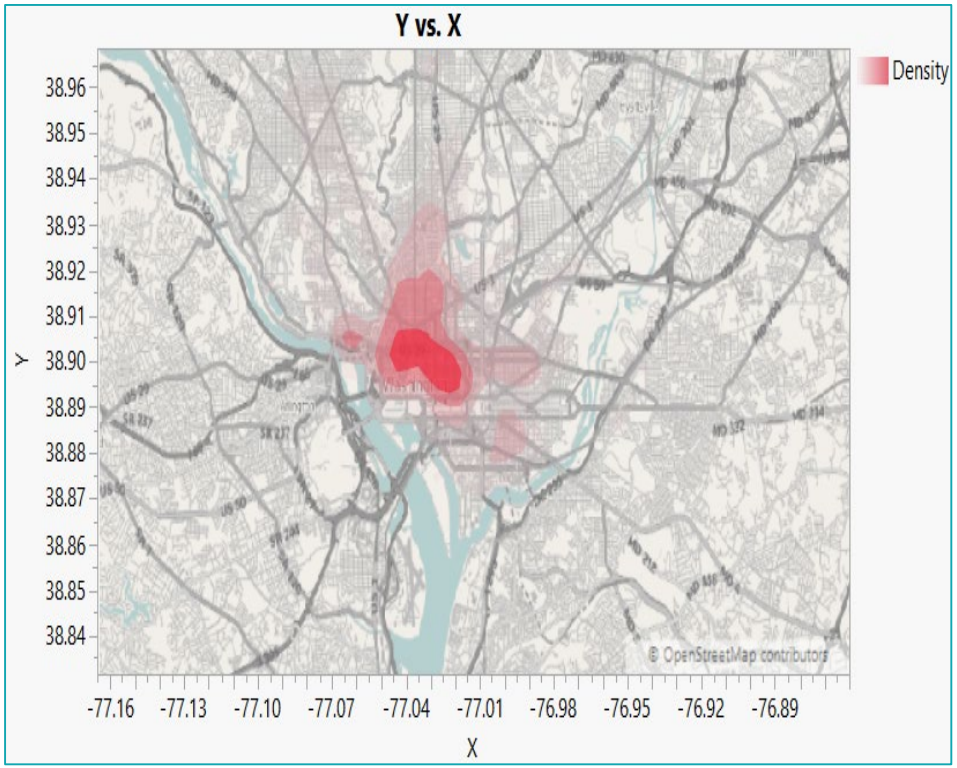
- **Exploratory Data Analysis**
- Empirical and Directional Variograms
- Kriging
- Regression
- Generate & Test Multiple Models

Exploratory Data Analysis

SAS JMP Spatial Map of Fine Amount

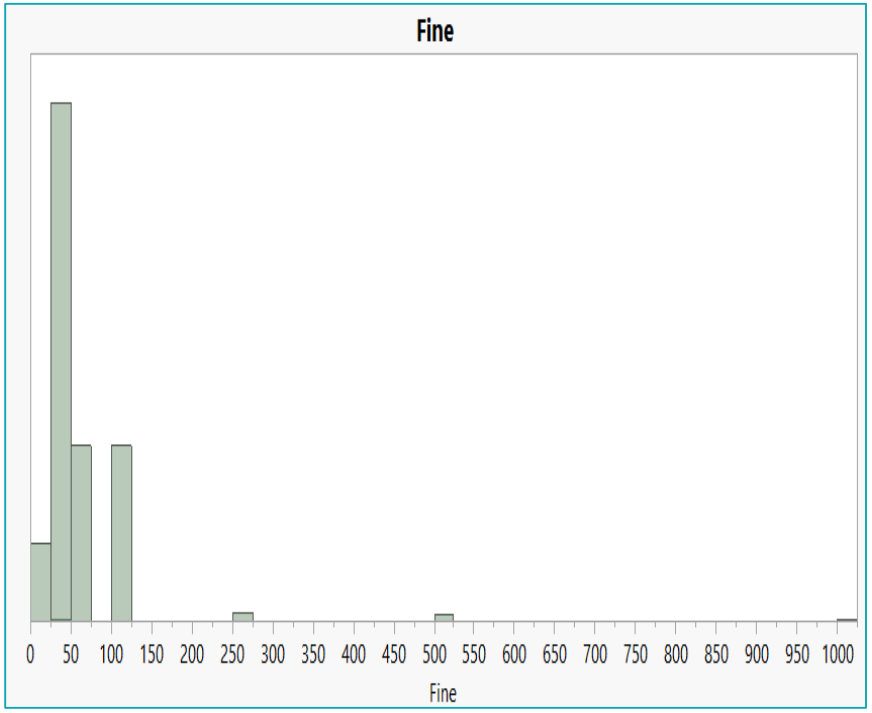


SAS JMP Contour Plot of Fine Amount

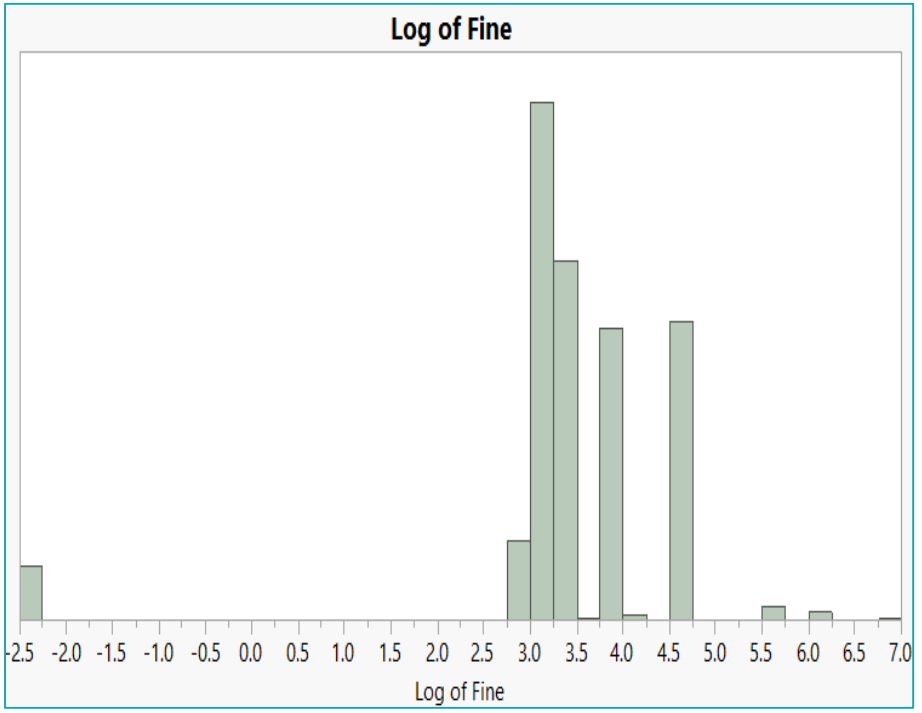


Exploratory Data Analysis

Histogram of Fine Amount



Histogram of Log Fine Amount



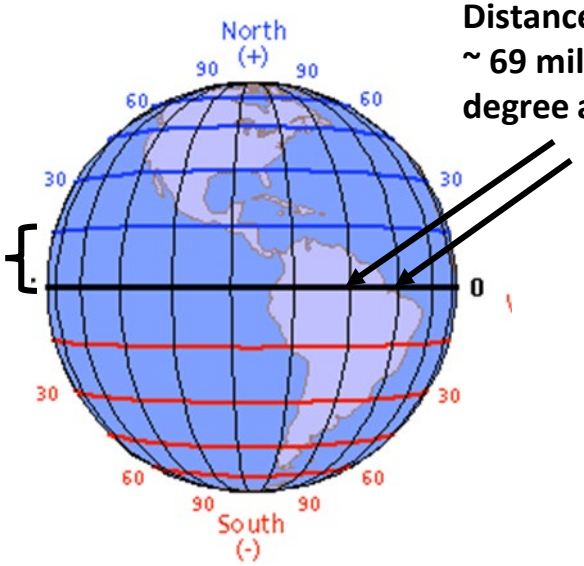
Agenda

- Exploratory Data Analysis
- **Empirical and Directional Variograms**
- Kriging
- Regression
- Generate & Test Multiple Models

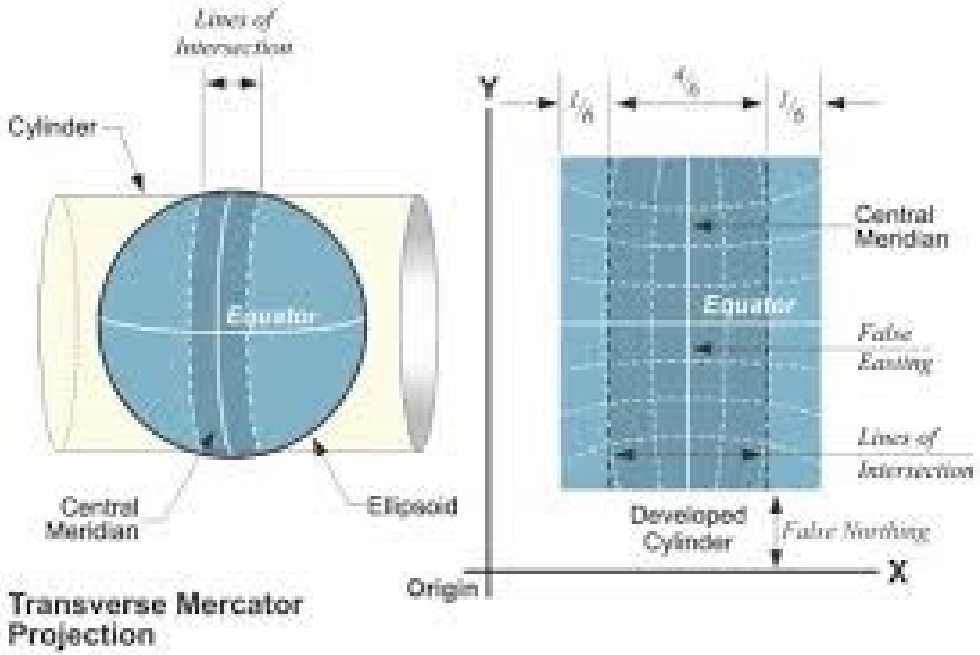
Coordinate Transformation

Except at equator, distance of a degree is not the same in EW direction as NS direction.

Difference in longitude. Distance varies. ~ 69 miles per degree at equator.



~ 69 miles / degree latitude



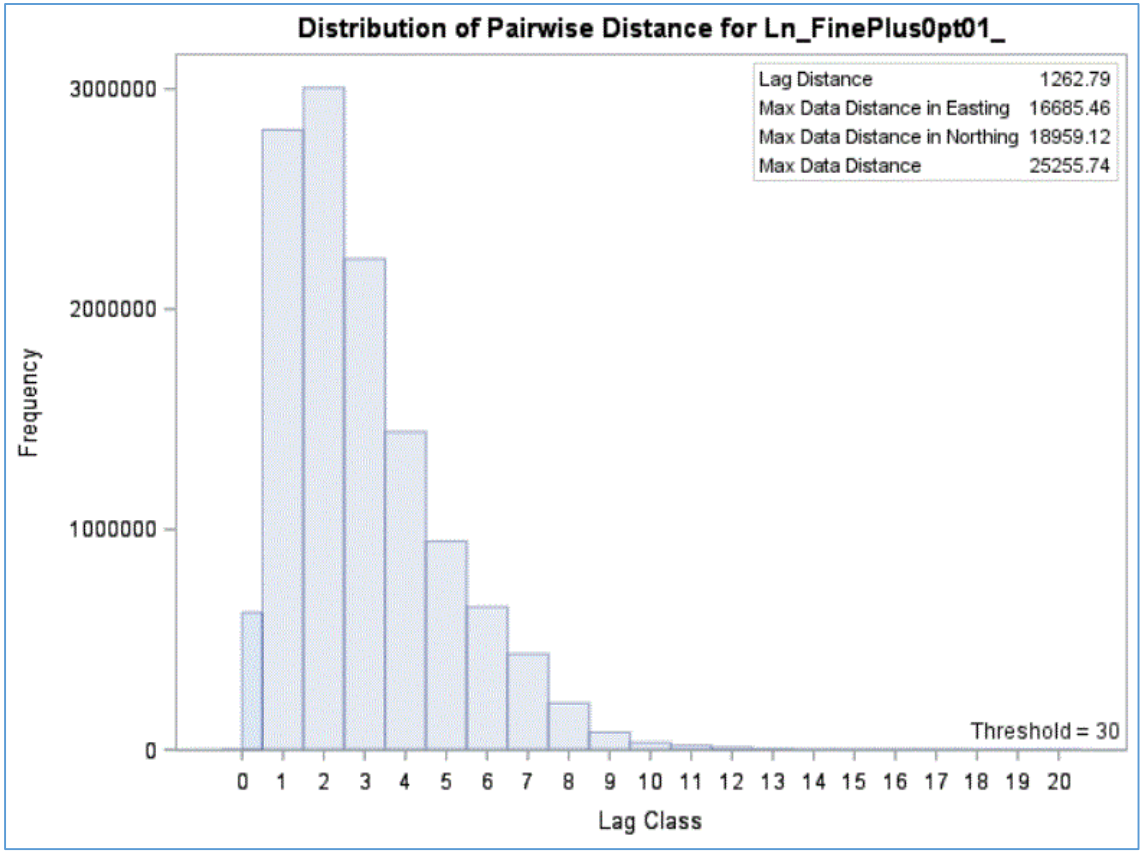
Project onto cylinder and unwrap. Still imperfect, but distortion minimal over a UTM zone.

UTM Zone 18, band S

Empirical Variograms

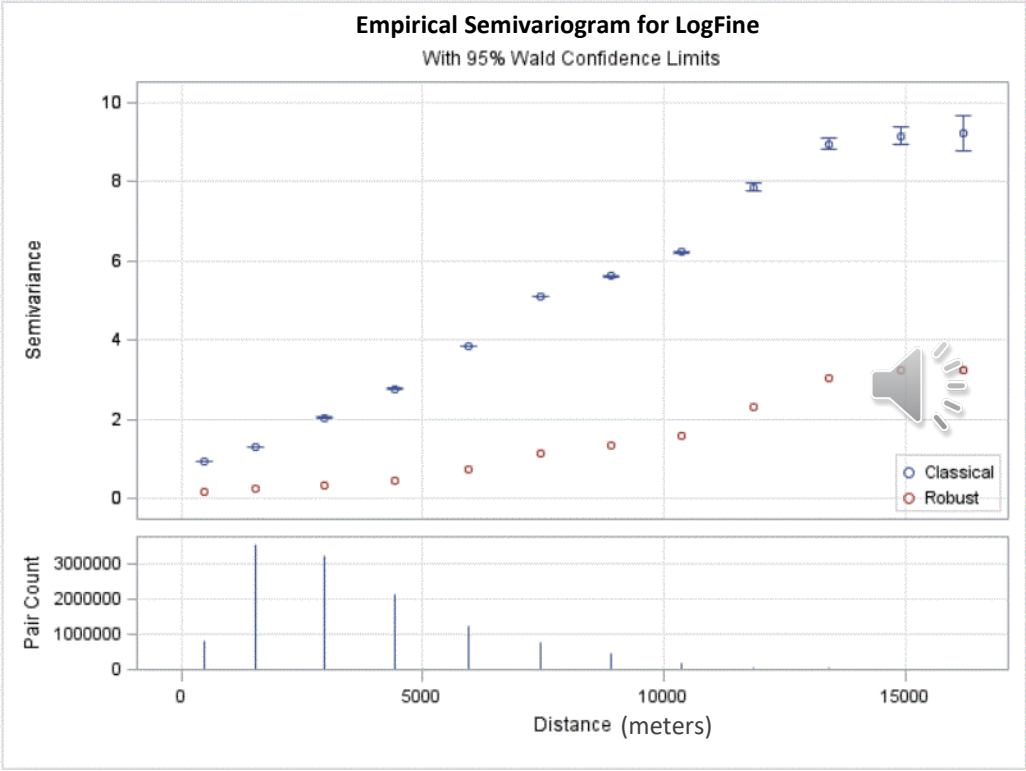
Variogram Histogram

Pairwise Distance Intervals				
Lag Class	Bounds		Number of Pairs	Percentage of Pairs
0	0.00	631.39	622995	4.98%
1	631.39	1894.18	2.81E6	22.51%
2	1894.18	3156.97	3.01E6	24.05%
3	3156.97	4419.76	2.23E6	17.83%
4	4419.76	5682.54	1.44E6	11.54%
5	5682.54	6945.33	946152	7.57%
6	6945.33	8208.12	648009	5.19%
7	8208.12	9470.90	434740	3.48%
8	9470.90	10733.69	211726	1.69%
9	10733.69	11996.48	78398	0.63%
10	11996.48	13259.27	32299	0.26%
11	13259.27	14522.05	19099	0.15%
12	14522.05	15784.84	9989	0.08%
13	15784.84	17047.63	3050	0.02%
14	17047.63	18310.41	257	0.00%
15	18310.41	19573.20	38	0.00%
16	19573.20	20835.99	0	0.00%
17	20835.99	22098.78	0	0.00%



Empirical Variograms

Variogram with Binned Averages and Confidence Limits



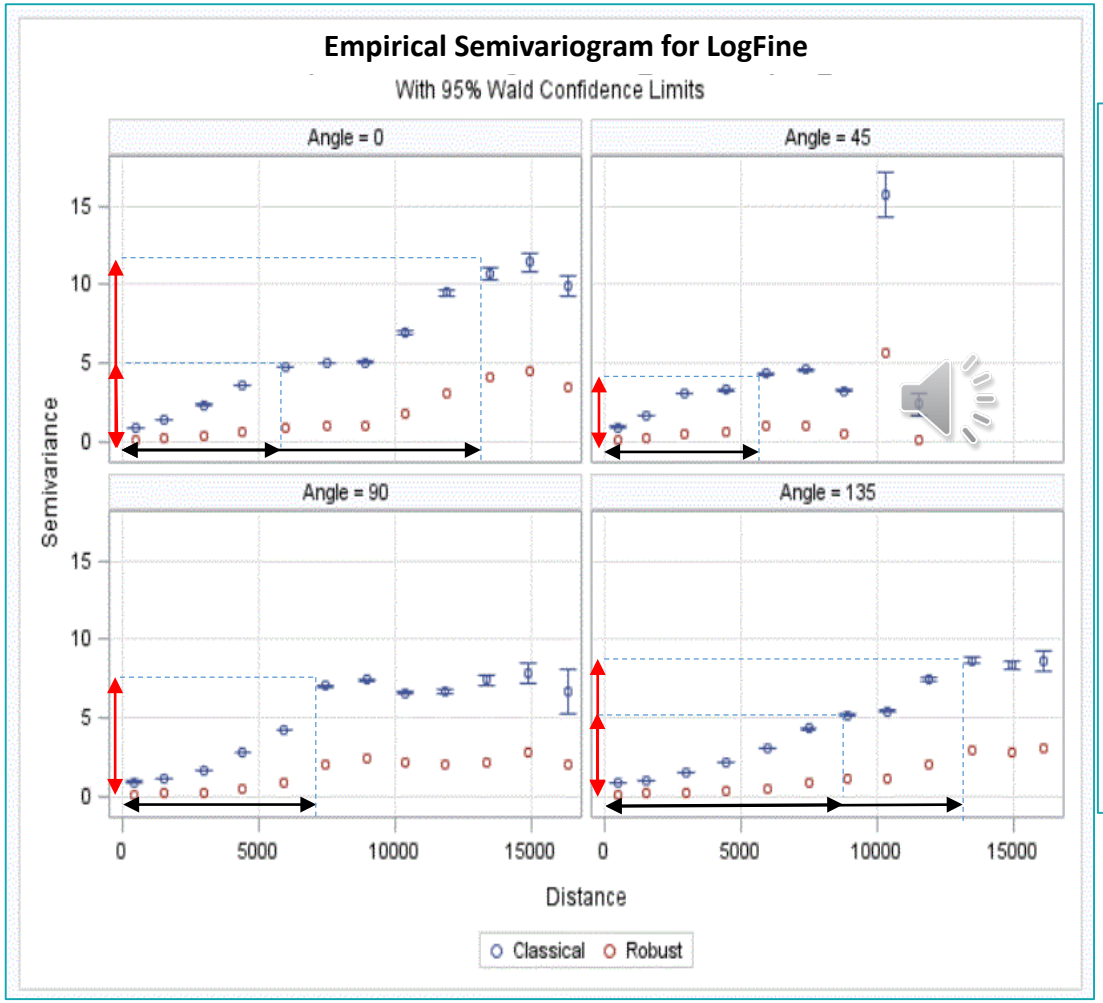
The VARIOGRAM Procedure
Dependent Variable: LogFine

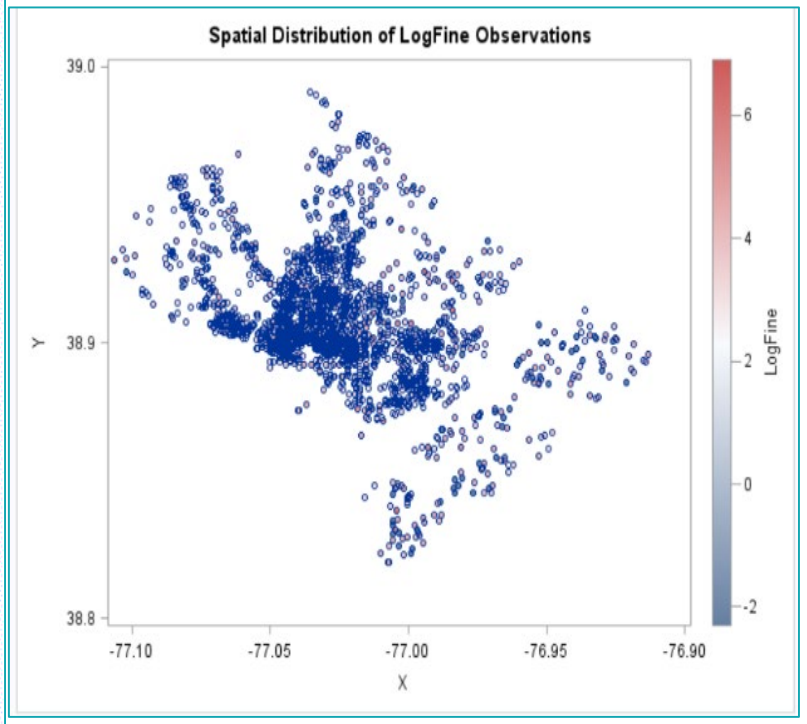
Empirical Semivariogram

Lag Class	Pair Count	Average Distance	Semivariance				
			Robust	Classical	Standard Error	95% Confidence Limits	
0	830015	470	0.182	0.949	0.00147	0.946	0.952
1	3.53E6	1544	0.226	1.311	0.00099	1.309	1.313
2	3.22E6	2956	0.334	2.043	0.00161	2.040	2.046
3	2.13E6	4433	0.465	2.773	0.00269	2.768	2.779
4	1.23E6	5950	0.715	3.853	0.00491	3.843	3.862
5	790539	7459	1.139	5.100	0.00811	5.084	5.115
6	475332	8891	1.340	5.620	0.012	5.597	5.642
7	193784	10362	1.597	6.215	0.020	6.176	6.254
8	57065	11866	2.307	7.864	0.047	7.773	7.955
9	27291	13425	3.045	8.958	0.077	8.807	9.108
10	12752	14907	3.254	9.164	0.115	8.939	9.389
11	3318	16213	3.245	9.247	0.227	8.802	9.692

Directional Variograms

Directional Variograms





Directional Variograms

Directional Variograms

Empirical Semivariogram at Angle=0

Lag Class	Pair Count	Average Distance	Robust	Classical
0	203903	470	0.183	0.953
1	816393	1552	0.240	1.467
2	723777	2956	0.395	2.391
3	403041	4397	0.632	3.615
4	270315	5989	0.950	4.744
5	210524	7491	1.087	5.032
6	139536	8877	1.099	5.081
7	52944	10345	1.791	6.952
8	16265	11882	3.057	9.497
9	6338	13475	4.086	10.747
10	3382	14934	4.465	11.432
11	1782	16313	3.457	9.914

Empirical Semivariogram at Angle=45

Lag Class	Pair Count	Average Distance	Robust	Classical
0	191750	470	0.188	0.970
1	706428	1525	0.272	1.734
2	512850	2911	0.502	3.098
3	258167	4417	0.658	3.302
4	127035	5913	0.984	4.345
5	51216	7362	1.049	4.604
6	15572	8749	0.531	3.277
7	966	10286	5.637	15.752
8	78	11507	0.129	2.410
9	0	.	.	.
10	0	.	.	.
11	0	.	.	.

Empirical Semivariogram at Angle=90

Lag Class	Pair Count	Average Distance	Robust	Classical
0	219583	456	0.177	0.974
1	1.01E6	1548	0.207	1.122
2	974308	2965	0.292	1.695
3	611770	4414	0.492	2.812
4	295930	5935	0.866	4.226
5	162867	7454	2.070	7.078
6	107294	8931	2.421	7.458
7	49689	10360	2.223	6.617
8	14893	11829	2.100	6.747
9	4609	13333	2.250	7.449
10	1259	14827	2.867	7.839
11	175	16297	2.045	6.743

Empirical Semivariogram at Angle=135

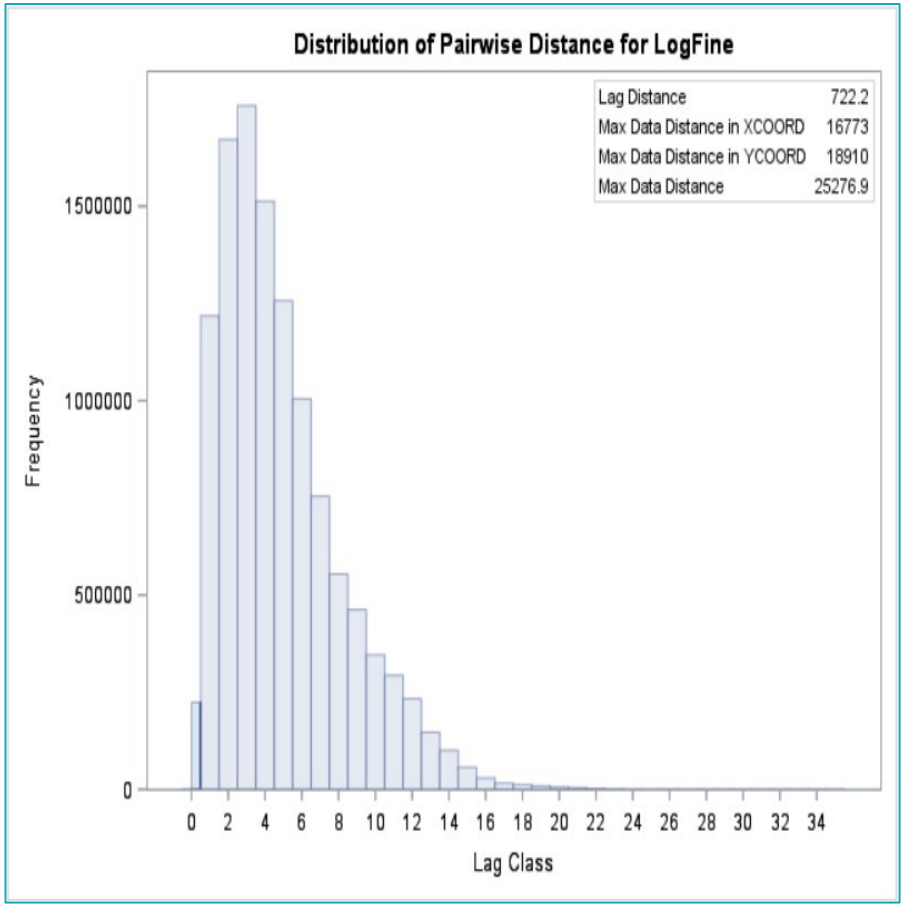
Lag Class	Pair Count	Average Distance	Robust	Classical
0	214779	485	0.181	0.902
1	992054	1548	0.206	1.075
2	1.01E6	2969	0.269	1.593
3	854663	4468	0.340	2.189
4	537231	5948	0.505	3.082
5	365932	7458	0.881	4.327
6	212930	8890	1.169	5.218
7	90185	10374	1.202	5.459
8	25829	11878	2.034	7.496
9	16344	13432	2.935	8.689
10	8111	14908	2.886	8.424
11	1361	16072	3.152	8.695

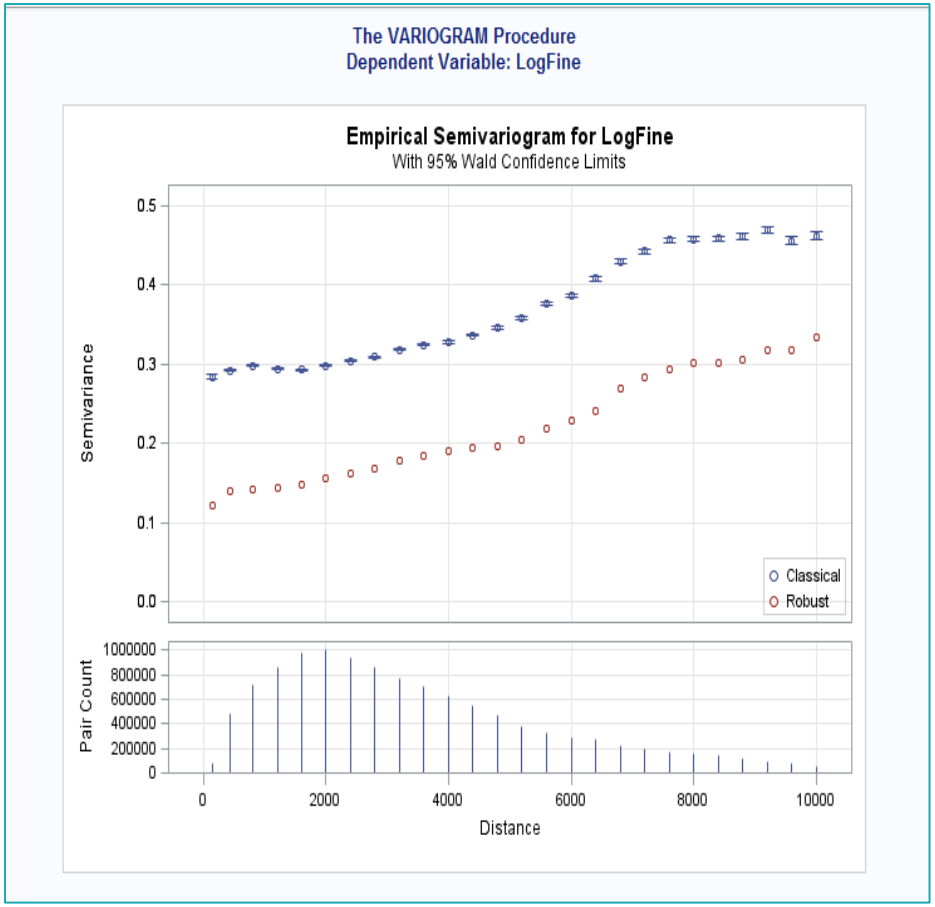
Angle	N-S (0)	NE-SW (45)	E-W (90)	SE-NW (135)
Nugget	1	1	1	1
Spatial Range	6,000 (15,000?)	6,000	7,000	8,000 (13,000?)
Sill	5 (11?)	4	7.5	5 (8.5?)

Agenda

- Exploratory Data Analysis
- Empirical and Directional Variograms
- **Kriging**
- Regression
- Generate & Test Multiple Models

Distribution of Pairwise Distance for Log Fine





Models & Parameter Estimates

Fit Summary				
Class	Model	SSE	AIC	Notes
1	Sph-Gau	0.0033723	-222.70649	Questionable fit
2	Sph-Exp	0.0068668	-204.21788	Questionable fit
3	Gau	0.0096620	-199.33884	
	Gau-Exp	0.0096623	-195.33805	
	Gau-Sph	0.0096624	-195.33796	
	Gau-Gau	0.0096624	-195.33788	
	Gau-Mat	0.0096624	-193.33796	
	Mat-Exp	0.0099340	-192.61699	
4	Sph	0.01446	-188.85878	

Parameter Estimates					
Parameter	Estimate	Approx Std Error	DF	t Value	Approx Pr > |t|
Nugget	0.2799	0.4307	21	0.65	0.5229
SphScale1	1E-6	0.7576	21	0.00	1.0000
SphRange1	4995.12	0	21	.	.
GauScale2	0.2263	0.7695	21	0.29	0.7716
GauRange2	7053.48	0	21	.	.

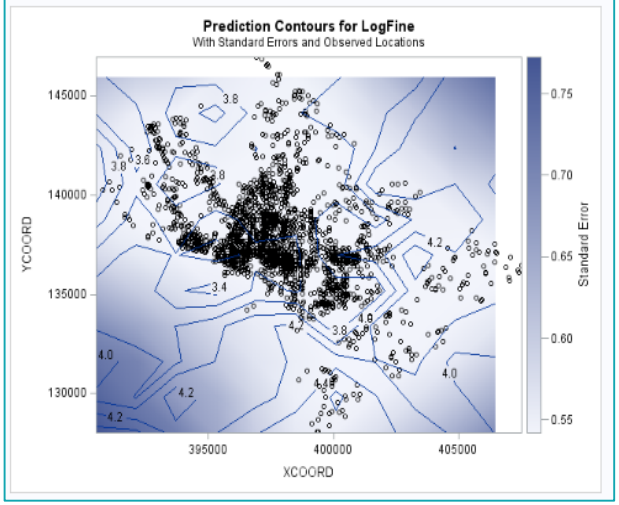
Kriging

Kriging Example 1

Kriging Information	
Prediction Grid Points	121
Type of Analysis	Local
Neighborhood Search Radius	60
Grid Points with Radius Incremented	121
Maximum Radius	9351.2363
Minimum Neighbors	20
Maximum Neighbors	All Within Radius

The KRIGE2D Procedure
Dependent Variable: LogFine
Prediction: Pred1, Model: Model1

Covariance Model Information	
Nested Structure 1 Type	Spherical
Nested Structure 1 Sill	1E-6
Nested Structure 1 Range	4995.1225
Nested Structure 2 Type	Gaussian
Nested Structure 2 Sill	0.2262609
Nested Structure 2 Range	7053.4814
Nested Structure 2 Effective Range	12216.988
Nugget Effect	0.2798662

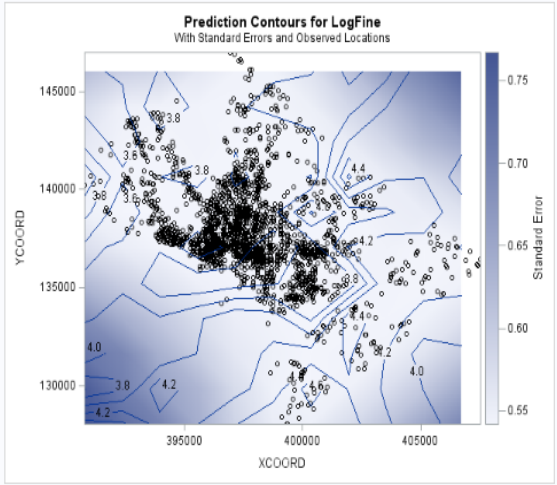


Kriging Example 2

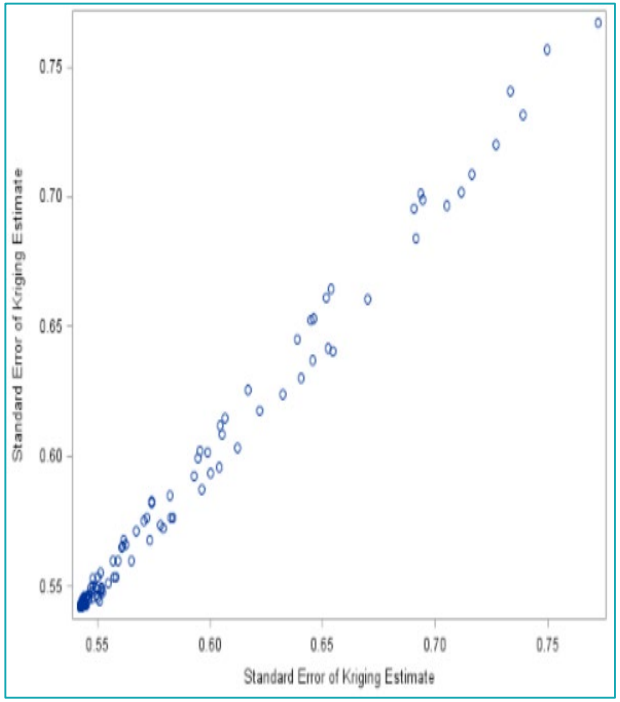
Kriging Information	
Prediction Grid Points	121
Type of Analysis	Local
Neighborhood Search Radius	60
Grid Points with Radius Incremented	121
Maximum Radius	9082.0012
Minimum Neighbors	20
Maximum Neighbors	All Within Radius

The KRIGE2D Procedure
Dependent Variable: LogFine
Prediction: Pred1, Model: Model1

Covariance Model Information	
Nested Structure 1 Type	Spherical
Nested Structure 1 Sill	1E-6
Nested Structure 1 Range	4995.1225
Nested Structure 2 Type	Gaussian
Nested Structure 2 Sill	0.2262609
Nested Structure 2 Range	7053.4814
Nested Structure 2 Effective Range	12216.988
Nugget Effect	0.2798662



Standard Error Scatterplot



Agenda

- Exploratory Data Analysis
- Empirical and Directional Variograms
- Kriging
- **Regression**
- Generate & Test Multiple Models

Regression

Fit Regression

The REG Procedure
Model: MODEL1
Dependent Variable: LogFine

Number of Observations Read	5000
Number of Observations Used	4834
Number of Observations with Missing Values	166

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	59.27850	14.81963	45.95	<.0001
Error	4829	1557.33520	0.32250		
Corrected Total	4833	1616.61370			

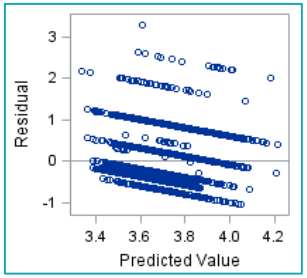
Root MSE	0.56789	R-Square	0.0367
Dependent Mean	3.67368	Adj R-Sq	0.0359
Coeff Var	15.45828		

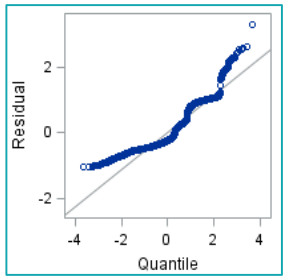
Parameter Estimates

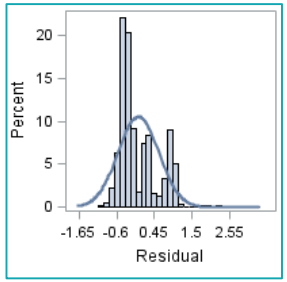
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	218.99064	26.59334	8.23	<.0001
X	1	2.97091	0.35865	8.28	<.0001
Y	1	0.33338	0.46275	0.72	0.4713
ISSUE_TIME	1	-0.00013988	0.00001582	-8.84	<.0001
WEEK_OF_YEAR	1	0.01472	0.00615	2.39	0.0168

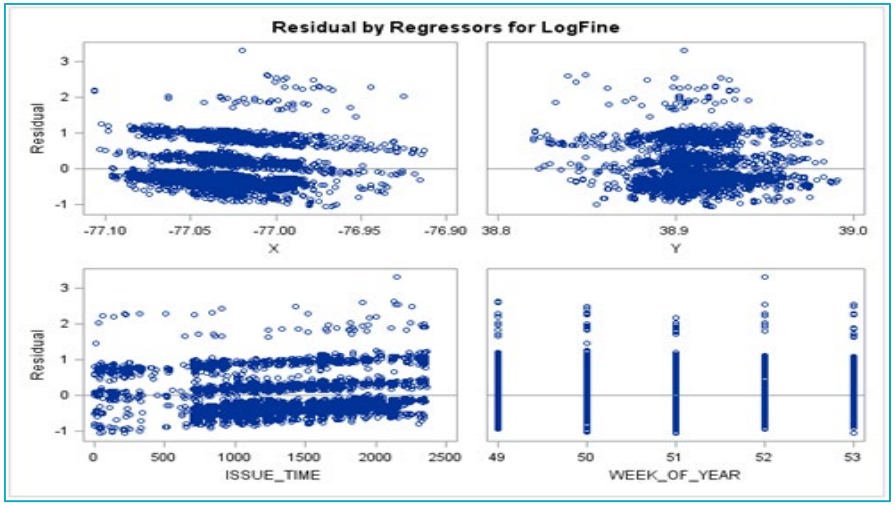
Calculate Residuals

```
proc reg data=&data;
model  LogFine= x y issue_time Week_of_year;
output out=regressionoutput
r=resid;
run;
```









Regression

Interpretation

Non-Binary

- X increases by 1-unit it is expected that the log (fine) increases 72.5%
- Y increases by 1-unit it is expected that the log (fine) decreases 32.9%
- Issue time increases by 1 minute it is expected that the log (fine) decreases .01%
- Week of year increases by 1-week it is expected that the log (fine) increases 1.5%

Binary

- If it is a Holiday then log (fine) increase 7.9%
- If it is a Rush Hour then log (fine) increases 23.6%
- If it is a N/A Violation then log (fine) decreases -914.6%
- If it is a Parking Violation then log (fine decreases 9.4%

Fit Regression

The REG Procedure
Model: MODEL1
Dependent Variable: LogFine

Number of Observations Read	5000
Number of Observations Used	5000

Analysis of Variance

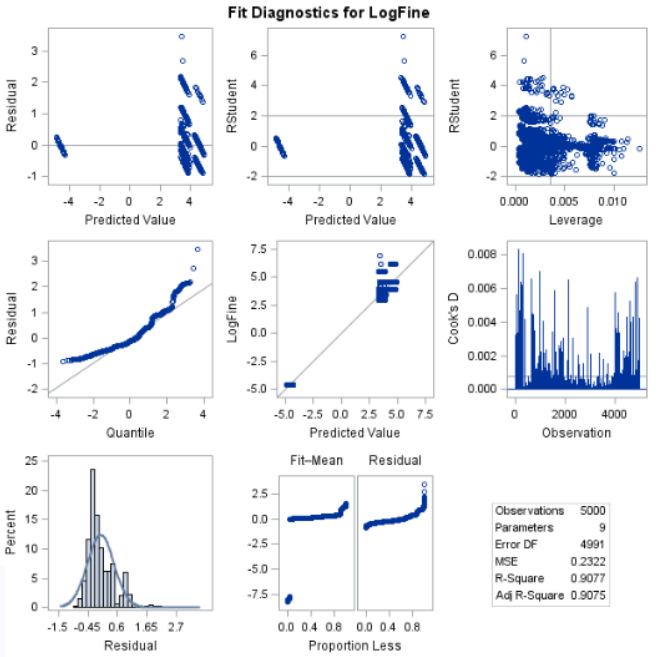
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	11395	1424.35017	6133.28	<.0001
Error	4991	1159.07479	0.23223		
Corrected Total	4999	12554			

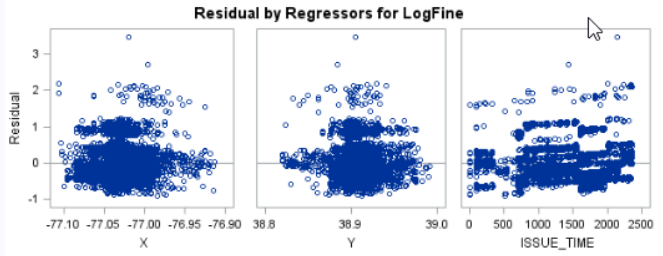
Root MSE	0.48191	R-Square	0.9077
Dependent Mean	3.40094	Adj R-Sq	0.9075
Coeff Var	14.16976		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	72.51956	22.08708	3.28	0.0010	0
X	1	0.72606	0.29932	2.43	0.0153	1.26067
Y	1	-0.32883	0.37505	-0.88	0.3807	1.18393
ISSUE_TIME	1	-0.00010425	0.00001306	-7.98	<.0001	1.10544
WEEK_OF_YEAR	1	0.01538	0.00537	2.86	0.0042	1.08965
HOLIDAY	1	0.07872	0.04349	1.81	0.0704	1.09290
RushHr	1	0.23605	0.01497	15.77	<.0001	1.06241
NAVio	1	-9.13765	0.04380	-208.61	<.0001	1.31821
ParkVio	1	-0.93642	0.02405	-38.94	<.0001	1.38945

Residuals



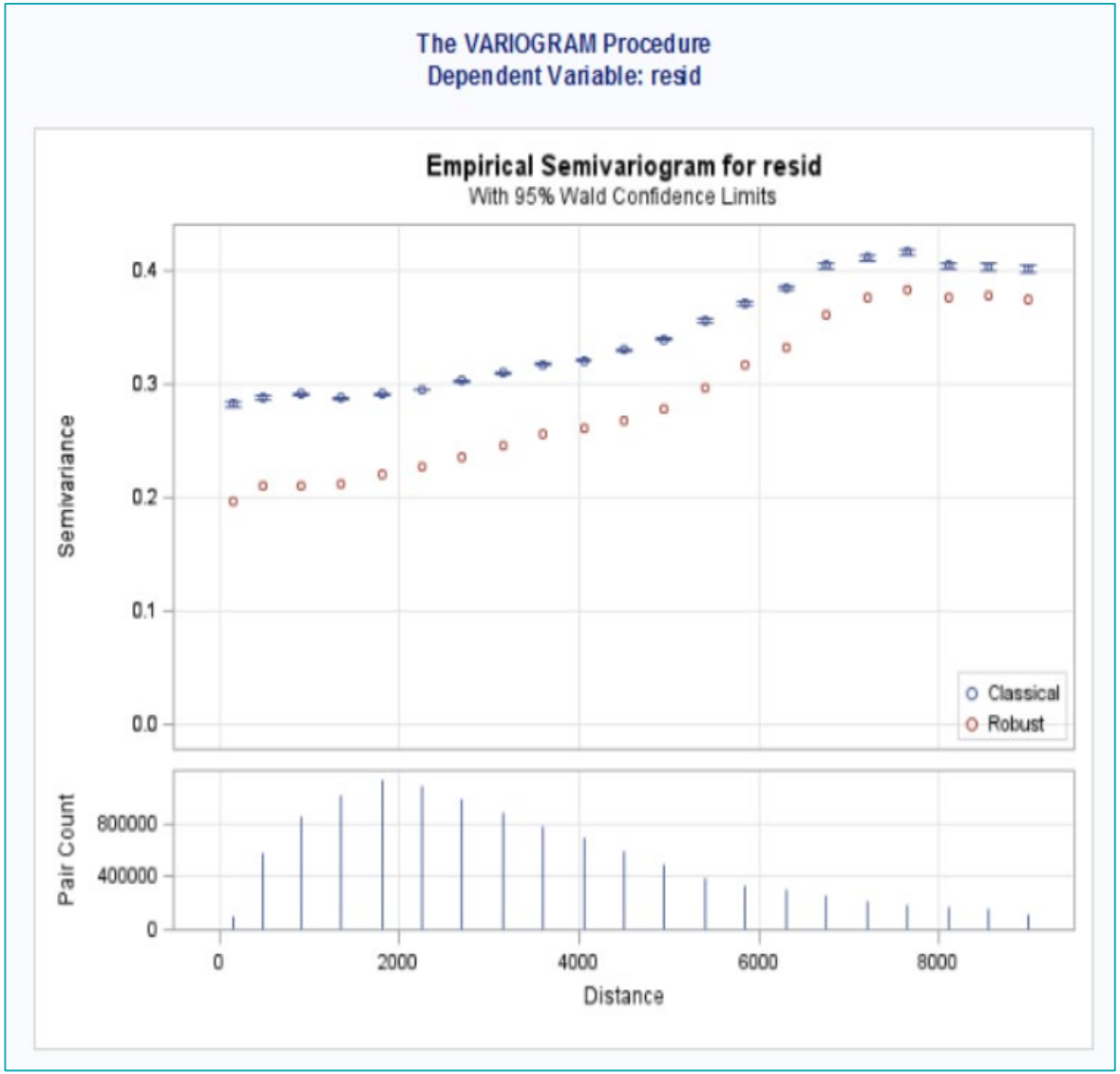


Agenda

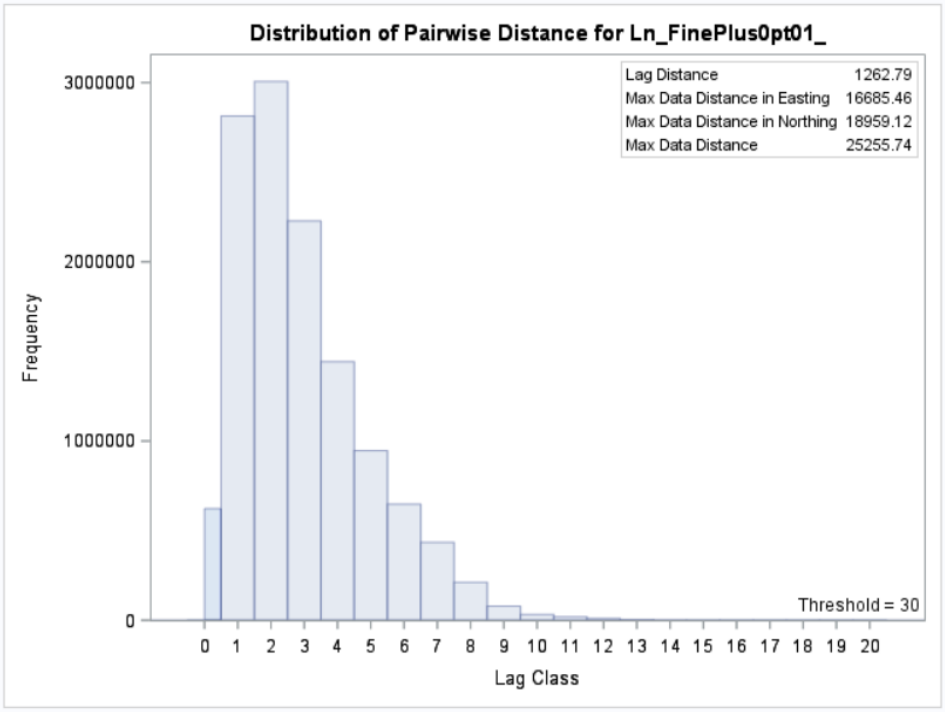
- Exploratory Data Analysis
- Empirical and Directional Variograms
- Kriging
- Regression
- **Generate & Test Multiple Models**

Modeling

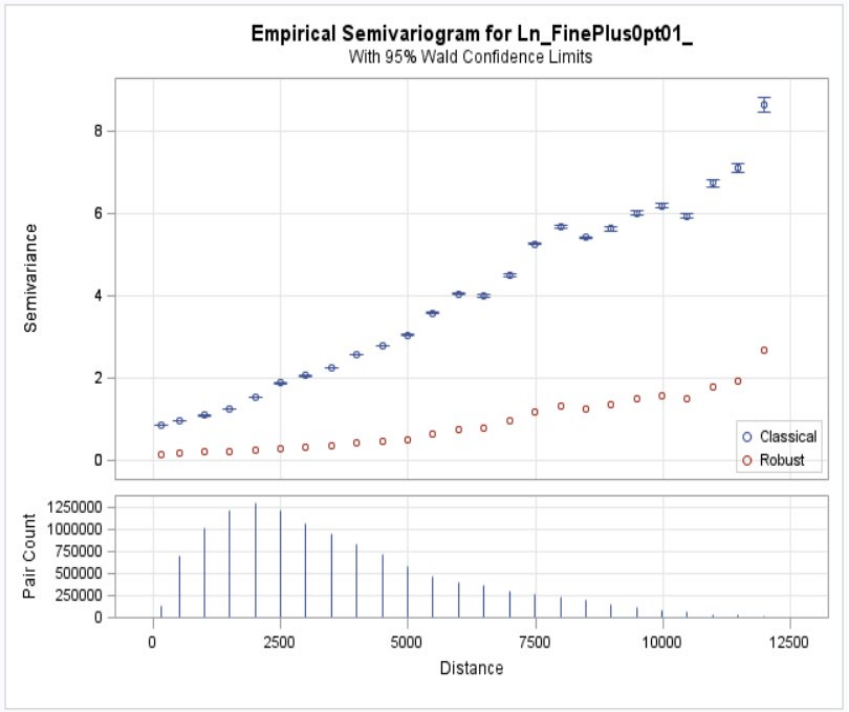
```
title j=left color=red "*** variogram lagd=450 maxlag=20 *** ";
proc variogram data=regressionoutput plot(only)=semivar;
    compute lagd=450 maxlag=20 cl robust;
    coordinates xc=XCOORD yc=YCOORD;
    var resid;
run;
```



Distribution of Pairwise Distances

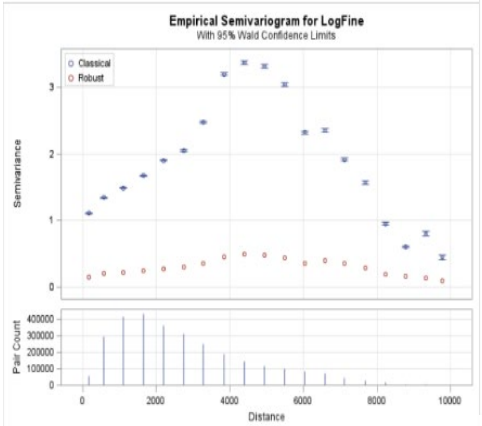


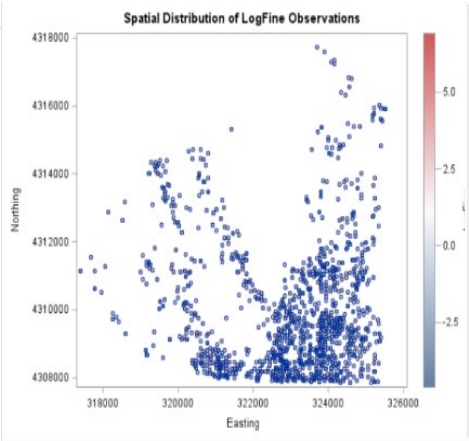
Semi-Variogram of the Full Dataset

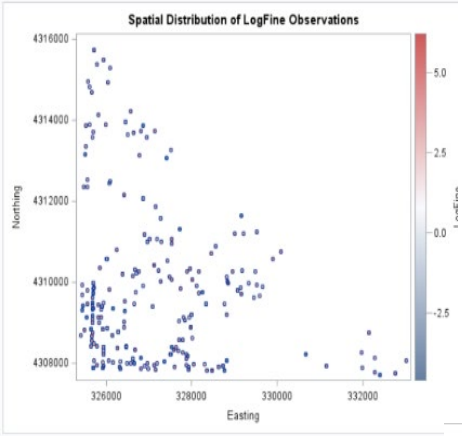


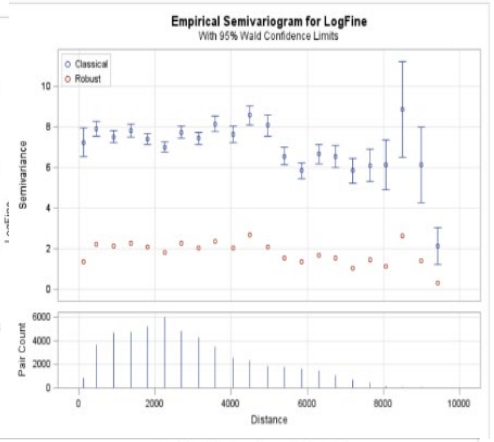
Modeling

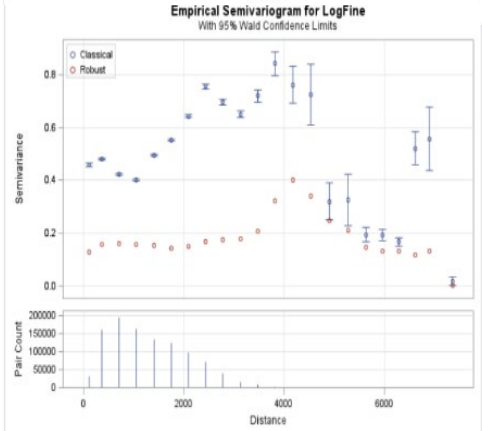
Semi-Variogram of Quadrant Data

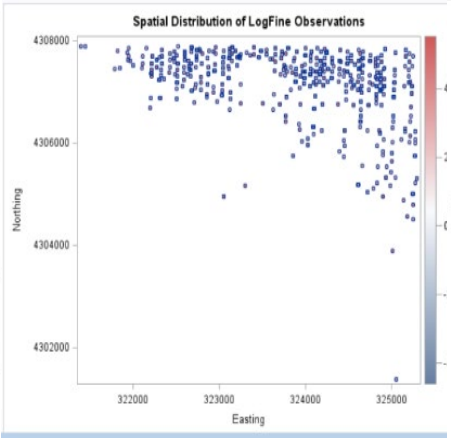


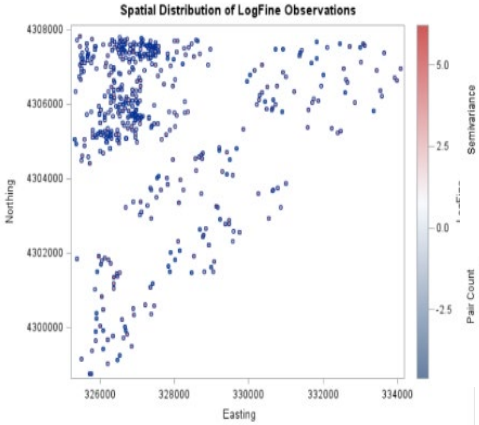


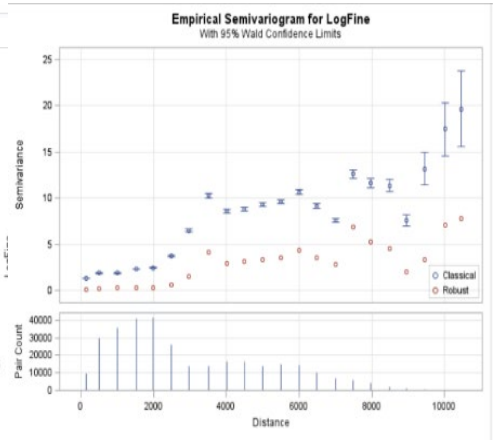












Modeling

Fit Regression on 80% Train Dataset

ML gaussian w/ Easting Northing

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Variance	Intercept	0.06296	0.01953	3.22	0.0006
SP(GAU)	Intercept	2.8369	2.5841	1.10	0.1361
Residual		0.1552	0.01922	8.07	<.0001

Fit Statistics	
-2 Log Likelihood	1037.6
AIC (Smaller is Better)	1053.6
AICC (Smaller is Better)	1053.7
BIC (Smaller is Better)	1091.0

Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
2	13.25	0.0013

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	4.4229	0.05553	0	79.65	.
RushHr	0.2242	0.03579	795	6.26	<.0001
NAVio	-9.0339	0.09499	795	-95.11	<.0001
ParkVio	-0.9190	0.05807	795	-15.83	<.0001
HOLIDAY	0.2261	0.09399	795	2.41	0.0164

ML exponential w/ Easting Northing

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Variance	Intercept	0.09173	0.02151	4.26	<.0001
SP(EXP)	Intercept	2.6414	2.3931	1.10	0.1349
Residual		0.1291	0.02033	6.35	<.0001

Fit Statistics	
-2 Log Likelihood	1040.4
AIC (Smaller is Better)	1056.4
AICC (Smaller is Better)	1056.6
BIC (Smaller is Better)	1093.8

Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
2	10.45	0.0054

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	4.4193	0.05554	0	79.57	.
RushHr	0.2224	0.03581	795	6.21	<.0001
NAVio	-9.0311	0.09514	795	-94.92	<.0001
ParkVio	-0.9138	0.05797	795	-15.76	<.0001
HOLIDAY	0.2249	0.09384	795	2.40	0.0168

AIC on 20% Test Dataset

AIC Scores	REML	ML
gau	1064.7	1053.6
mat	1070	1058.9
exp	1067.6	1056.4
sph	1076	1065.6

SAS Code: Proc Mix for Regression & Parameters

```
PROC MIXED covtest data=train method=ml
scoring=5 convf=1e-1;
model LogFine = RushHr NAVIO PARKVIO HOLIDAY/
solution;
repeated / subject=intercept local
type=sp(gau)(Easting Northing);
run;

*LOWEST AIC  1053.6  residual vars created*;
resid_MLgauEN = LogFine - (4.4229 +
0.2242*RushHr - 9.0339*NAVio - 0.919*ParkVio
            + 0.2261*HOLIDAY);
```

Modeling

The KRIGE2D Procedure
Dependent Variable: resid_MLgauEN
Prediction: Pred1, Model: Model1

Covariance Model Information	
Type	Gaussian
Sill	0.06296
Range	2.8369
Effective Range	4.9136549
Nugget Effect	0.1552

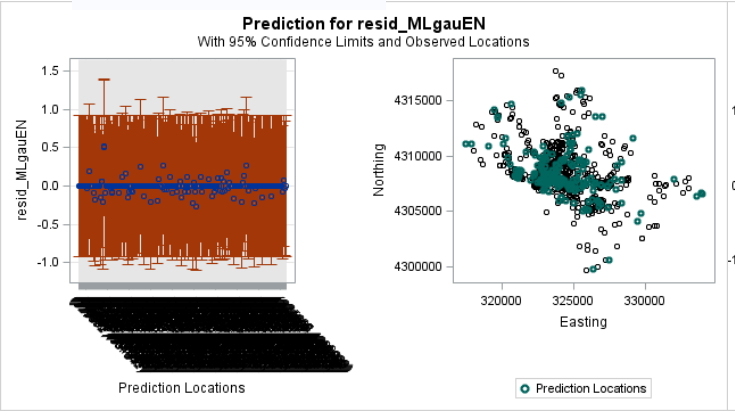
The KRIGE2D Procedure
Dependent Variable: resid_MLexpEN

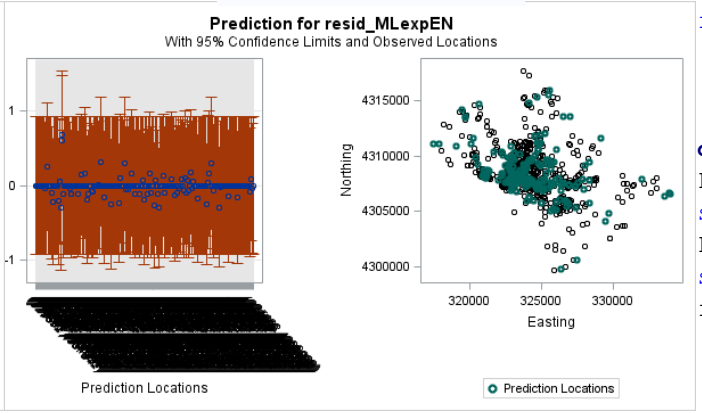
Number of Observations Read	800
Number of Observations Used	800

Kriging Information	
Prediction Grid Points	200
Type of Analysis	Global

The KRIGE2D Procedure
Dependent Variable: resid_MLexpEN
Prediction: Pred1, Model: Model1

Covariance Model Information	
Type	Exponential
Sill	0.09173
Range	2.6414
Effective Range	7.9242
Nugget Effect	0.1291





Krig Values using the Residuals

REML	*X Y Coordinates*	***UTM Shifted***	***ML***	*X Y Coordinates*	***UTM Shifted***
gau MSE	0.2511789	0.2462179	*gau MSE*	n/a	0.2461608
gau MAE	0.3716209	0.3623198	*gau MAE*	n/a	**0.3618963**
mat MSE	0.2541386	0.2487546	*mat MSE*	0.2511674	0.2487108
mat MAE	0.3910720	0.3646501	*mat MAE*	0.3695892	0.3646516
exp MSE	0.2529205	0.2479724	*exp MSE*	n/a	0.2478817
exp MAE	0.3863332	0.3620828	*exp MAE*	n/a	0.3620768
sph MSE	0.2518858	0.2501950	*sph MSE*	n/a	0.2503963
sph MAE	0.3807165	0.3655414	*sph MAE*	n/a	0.3664073

SAS Code: Proc Krige2d: Target, Coordinates, Parameters

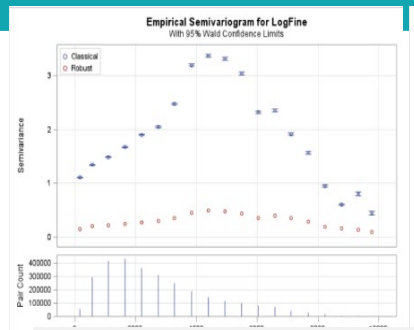
```
proc krige2d data=lik outest=
Proj3.krigresidresid_MLgauEN;
coord xc=Easting yc=Northing;
grid gdata=validate xc=Easting
yc=Northing;
pred var=resid_MLgauEN;
model scale =0.06296 Range =
2.8369 nugget = 0.1552
form=gau;
```

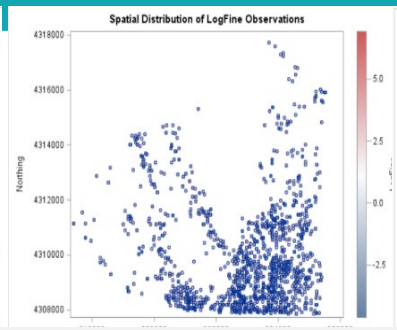
SAS Code: Validation Set used to Compare Actual and Predicted

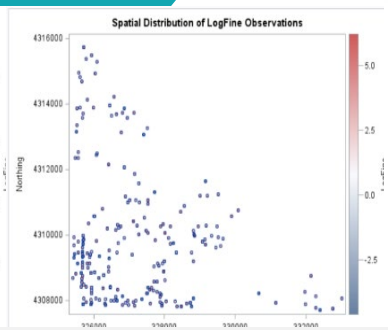
```
data
Proj3.krig_validate_MLgauEN;
set
Proj3.krigresidresid_MLgauEN;
set validate;
final = estimate + (4.4229
              +0.2242*RushHr
              -9.0339*NAVio
              -0.919*ParkVio
              +0.2261*HOLIDAY);
error = final - LogFine;
error2 = (final - LogFine)**2;
abs = abs(final - LogFine);
run;
```

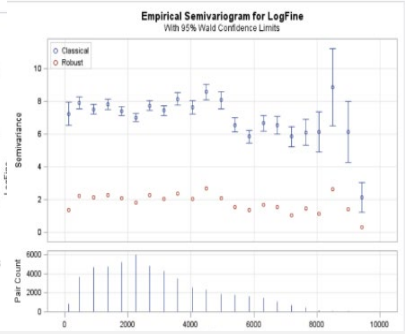
Val. Set: Performance Presented

```
title1 "Validation data kriging
results.";
title2 "Covariance modeled with
ML_gauEN";
proc means
data=Proj3.krig_validate_MLgauE
N;
var LogFine final error error2
abs;
run;
```

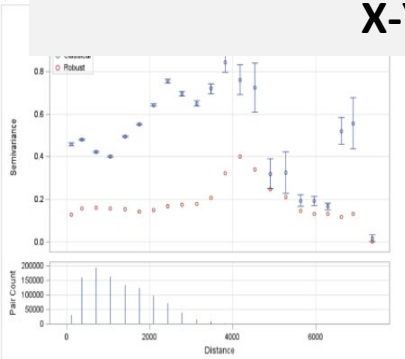


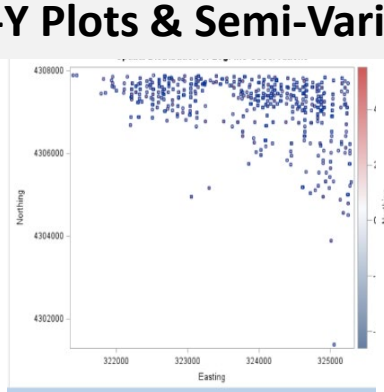


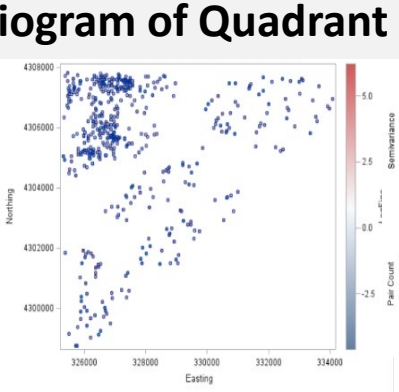


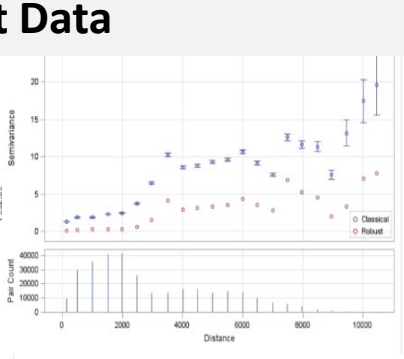


X-Y Plots & Semi-Variogram of Quadrant Data

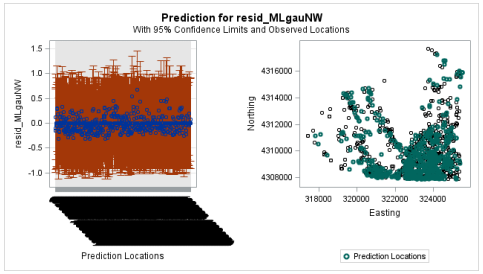




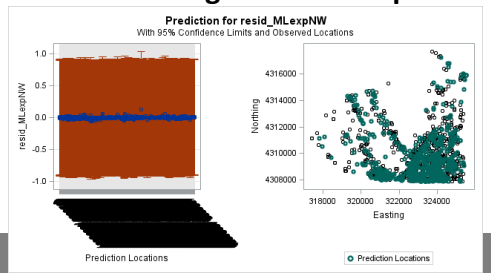




NW Kriged Values gau



NW Kriged Values exp



Gau and Exp Model Errors Comparing Quadrant Data and Initial Data Analysis

REML	*All Data*	*SE Data*	*SW Data*	*NW Data*	*NE Data*
gau MSE	**0.2462179**	0.2467948	0.2513917	0.1809254	0.3409958
gau MAE	**0.3623198**	0.3213962	0.4139875	0.3329316	0.3795189
exp MSE	0.2479724	0.2453665	0.2516336	0.1848097	0.3182412
exp MAE	**0.3620828**	0.3199048	0.4152326	0.3503073	0.3890685
ML	*All Data*	*SE Data*	*SW Data*	*NW Data*	*NE Data*
gau MSE	**0.2461608**	0.2470418	0.2642446	0.1809266	0.3472765
gau MAE	**0.3618963**	0.3217213	0.4334895	0.3329576	0.3808527
exp MSE	0.2478817	0.2457904	0.2544688	0.1848774	0.3114972
exp MAE	**0.3620768**	0.3204898	0.4183122	0.3504072	0.3780639