Diagnostics for Spatial Models

Mark Otto and David Meek

U.S. Fish and Wildlife Service, Laurel, MD
U.S.Department of Agriculture—Agricultural Research Service, Ames, IA

16 March 2006

1

2

Box-Jenkins Iterative Modeling Procedure

1. Identify the model: transformation, regression, trend, correlation structure
2. Estimate model parameters
3. Check that the model fits the assumptions
4. Repeat 1–3 until diagnostics check out

3

General Concepts

See that the estimated model fits the assumptions. The usual assumption is that the residuals have a zero mean and constant variance.

- Residuals do not show any consistent patterns
  - Fitted values
  - Regression variables
  - Important spatial coordinates
  - Time
- Residuals are white noise: check with the variogram of the residuals

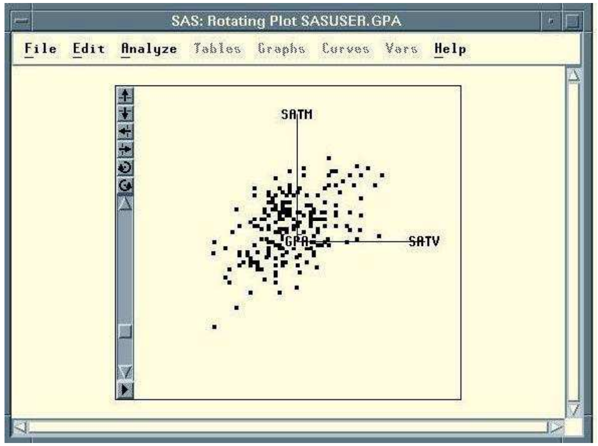
4

Stem Plots

```
stem(rbyc)
 The decimal point is at the |
 -2 | 33333333333333333333333333333333333333333333333333333333333333
 -1 |
 -1 |
 -0 |
 -0 |
  0 | 000000000000000
  0 | 77777777777777
  1 | 11111111111144444
  1 | 666666666668889999
  2 | 1222222222334
  2 | 55667
  3 |
  3 | 8
  4 | 0
```


Spinning

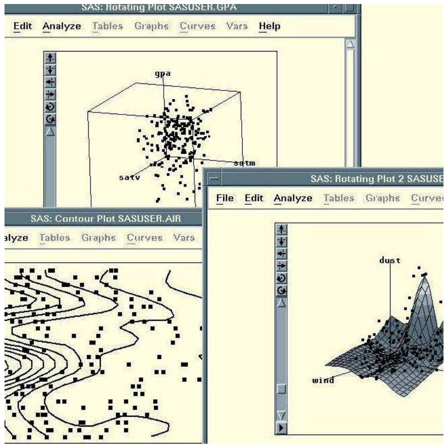
- Spatial data is ≥ 3 dimensional
- Spin map to change point of view



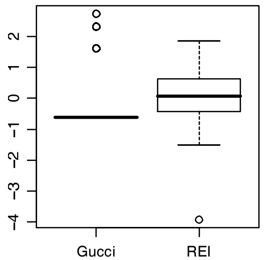
These were done in SAS Insight.

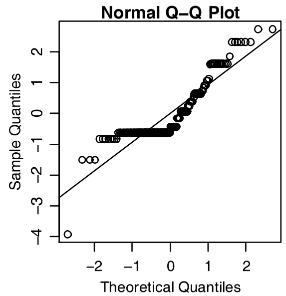
Brushing

- Identify interesting points
- Link to identify in other graphs and tables



Box and Normal Probability Plots





When the independent variables are factors the residuals can be shown in box-plots. Here the levels are different and the variation is much different between the factors. This is on the original data. The lines in the normal probability plot show the strong correlations in the data.

Regression Diagnostics

- Studentized residuals: mean removed, standard variance
- Influence statistics: change due to missing observation
- Plot the above by direction, fitted value, fixed, time

These diagnostics differ from regression diagnostic in that they are transformed by the same linear transformation used on the correlated residuals to make them IID, with $\mathbf{L}^{-1}$.

Transform Correlated Errors to IID

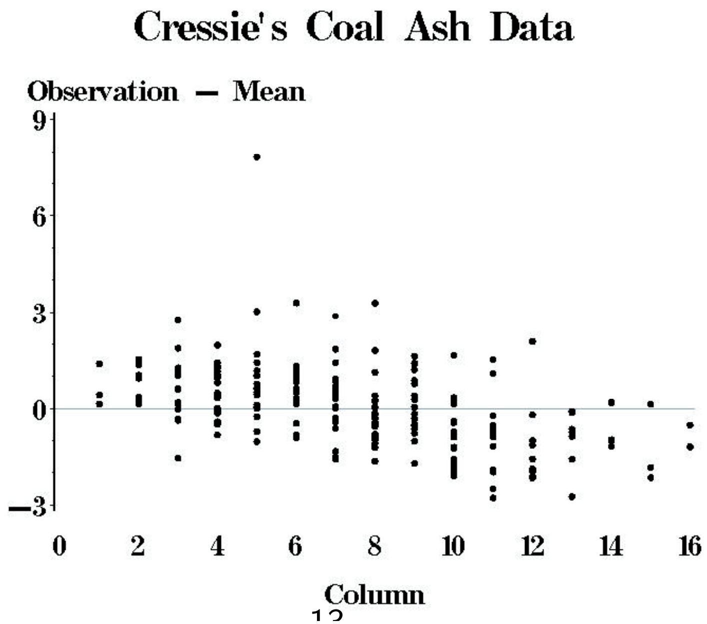
Transform the errors back to IID, $\mathbf{y} \sim N(\mu, \mathbf{V}\sigma^2)$
Transform back to independent normal, $\mathbf{V} = \mathbf{L}\mathbf{L}'$
then

$$\mathbf{L}^{-1}(\mathbf{Y} - \mathbf{X}\beta) \sim \mathrm{N}(\mathbf{0}, \mathbf{I}\sigma^2)$$

Transform Correlated Errors to IID
Transform the errors back to IID, $\mathbf{y} \sim N(\mu, \mathbf{V}\sigma^2)$
Transform back to independent normal, $\mathbf{V} = \mathbf{L}\mathbf{L}'$
then
$\mathbf{L}^{-1}(\mathbf{Y} - \mathbf{X}\beta) \sim \mathrm{N}(\mathbf{0}, \mathbf{I}\sigma^2)$

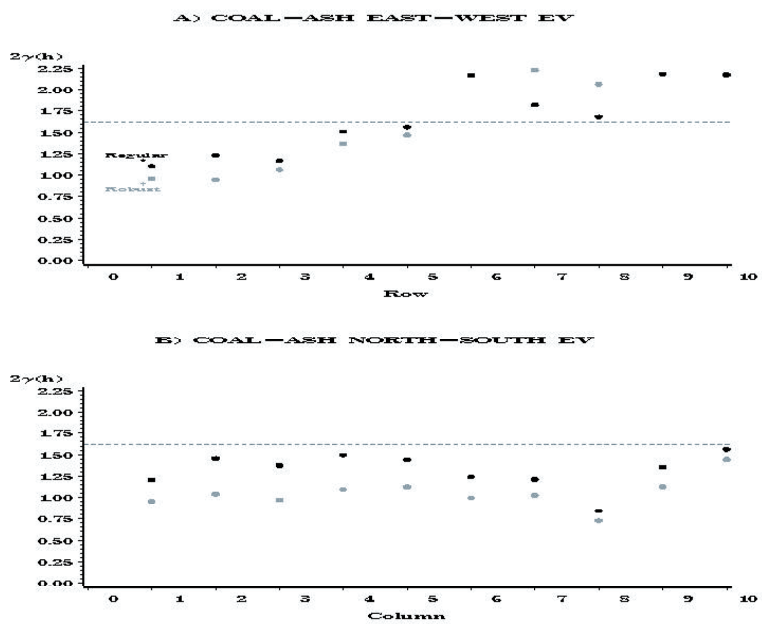
Have talked about how to define correlations in other talks

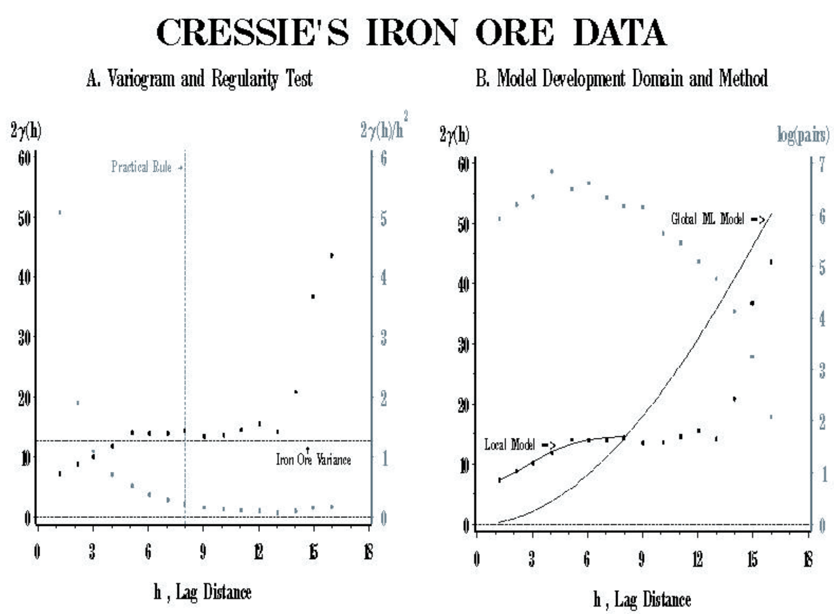
Standardized Residuals

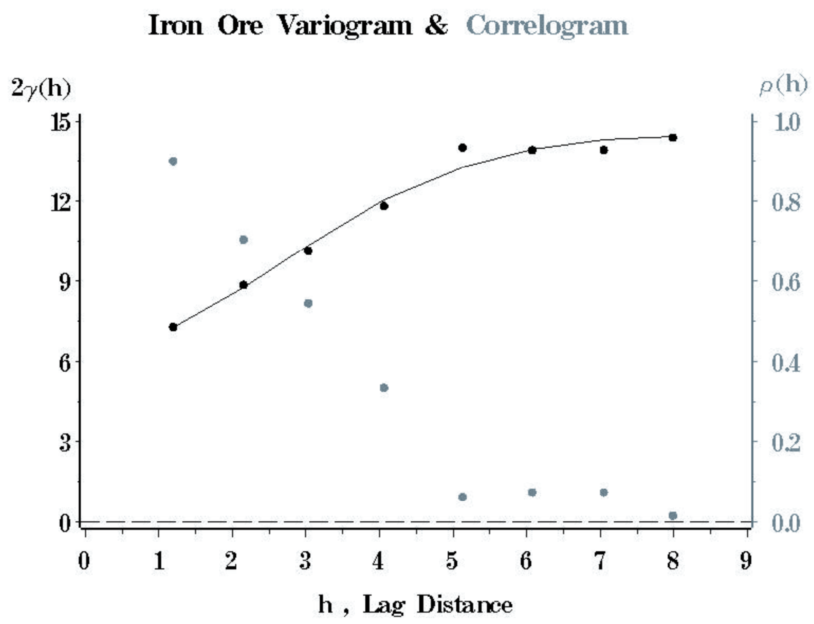


Variogram Plots of Residuals

- Omnidirectional classical, $\gamma(h)$, and robust, $\gamma^{\dagger}(h)$ variograms
- Try directional $\gamma_{\alpha}(h)$ in two (0 and 90 deg) to six (0, 30, ..., 150 deg) depending on the limits of the data
- Pair count vs. h Include a reference line for white noise variogram, $\hat{\gamma} = \sigma^2$
- Regularity test, $\gamma(h)/h^2$ vs. h
- Correlation structure, $\rho(h)$ or $C(h)$ vs. h







Model Mispecification Grid Check

- Ribeiro and Diggle (2004) suggest eye-ball variogram parameters to fit empirical
- Suggest fitting model so residual variogram is white-noise

Outliers

- Two types: point (a single outlying value) and patch (a region at a different level)
- Difficult to identify with correlated data because they don't stand out on their own. It is how the differ from nearby values
- Use a priori knowledge (at least use to confirm)
- Can confound spatial correlation structure

Point Outlier

- Set outlier detection critical value, $t = 3.5$ (p=0.01 controlling for an experiment-wise error rate)
- Identify the fixed effects and initial correlation structure
- Estimate and fix correlation structure
- Add outlier dummy for each observation and estimate
- Add most significant outlier to fixed effects
- Repeat 2–4 until no more significant outliers
- Reassess the fixed effects and correlation structure

Iterative Outlier Detection

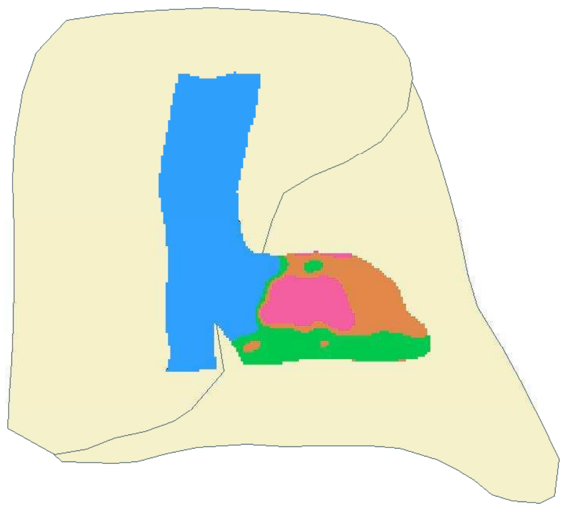
```
Iteration 1.
  Outlier       Est     SE      t
+ Obs47         2.31    0.1322  4.2 *
  Obs141        1.22    0.1411  3.1

Iteration 2.
  Obs47         2.37    0.1072  3.9 *
+ Obs141        2.07    0.1100  3.6 *
```

Outlier at 47 was picked up on the first iteration but 141 was below the critical level. The model was re-estimated with new correlations and variance. An outlier at 141 was then picked up. After the next round no more observation were over the critical value

Region Breaks

- Region a different level
- Too difficult to test every possible patch
- Use ArcGIS patch identification tool (classification algorithm)



Green patch identifies sole fishing area with higher fishing and bycatch

What Diagnostics do not Cover

- Measurement error:
  - Errors caught by repeated measurement at the sample points
  - Errors independent regression variables
  - Inaccuracies in the locations and thus the measures of distance between points
- Micro-scale variation: errors at scales below the smallest distance increment
- Aliasing, variation at periodicities covered by spacings in the data
- Emphasizes the importance of good design to address the important sources of error

Conclusions

- Use the Box-Jenkins iterative modeling approach
- Use Exploratory Data Analysis stem-plots, qq-plots, and brushing for multi-dimensional views of the data
- Look for patterns in residual and influence diagnostics
- Check that classical and robust variograms look like white noise and there is no anisotropy in the residuals
- Iteratively identify point outliers. Use them to look harder at outlying values. Have a priori reasons for including outliers and patches in the model.