Colorado potato beetle infestation in plots on a lattice design

Matt Kramer and Don Weber

kramerm@ba.ars.usda.gov, weberd@ba.ars.usda.gov

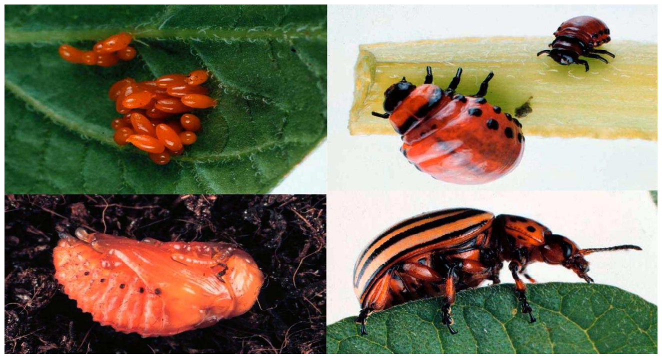
Biometrics (MK) and Insect Biocontrol (DW), ARS/BARC/USDA

Outline

- Introduction to Colorado potato beetles
- Experimental design and plots
- *R* software—*spdep* package
- Spatial weights
- Models and analyses
  - Ignore spatial dependencies
  - SAR
  - CAR
- Diagnostics
- Conclusion

Introduction to Colorado potato beetles

There are 4 life stages: egg, larva, pupa, adult. Larvae and adults feed on leaves.



photographs by Doro Röthlisberger, Zoological Museum, University of Zurich

Introduction to Colorado potato beetles

- Colorado potato beetles (*Leptinotarsa decemlineata*) overwinter as adults and can pass through 2–3 generations in Maryland.
- The data were taken (mid May) when all life stages were present
- CPB is a pest in North America (where it is native) but has also been introduced into Europe, which now suffers damage from it comparable to that in North America.
- CPB attacks plants in the nightshade family (potatoes, eggplants, tomatoes, and their wild relatives).
- Colorado potato beetles have developed resistance to a long succession of different insecticides, and its natural enemies do not reliably control it in current farming practices.
- New practices, in combination with natural enemies, show promise to maintain CPB populations below economic thresholds, reducing the need for pesticide applications.

Introduction to Colorado potato beetles

- In this experiment, tillage practice, planting date, and mulch cover were manipulated.
- We chose these data for a lattice example because the plots are laid out on a lattice, and it is a reasonably small data set. At the onset, we knew there were treatment effects but did not know if there were spatial dependencies.
- The goal of the project is to determine which combination of treatments best reduces CPB infestation
- In addition to treatment effects, we thought there might be block and border effects (and spatial correlation among neigboring plots)
- Sampling occurred in the interior of the plots
- Spatial correlation was suspected because adults and larva are mobile, both walk and adults can fly

Cooperators and administrators

Left to right: Matt Greenstone, Phyllis Johnson, Don Weber, Ron Korcak, John Teasdale, Aref Abdul-Baki, Vinod Kumar

Field team

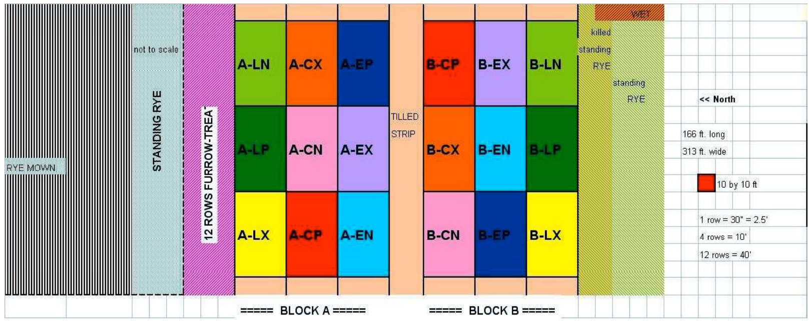
Left to right: Jenn Curtis, Jon Curtis, Eddie Bender, Michael Donovan, Mike Athanas, Greg Benedict

Experimental design and plots

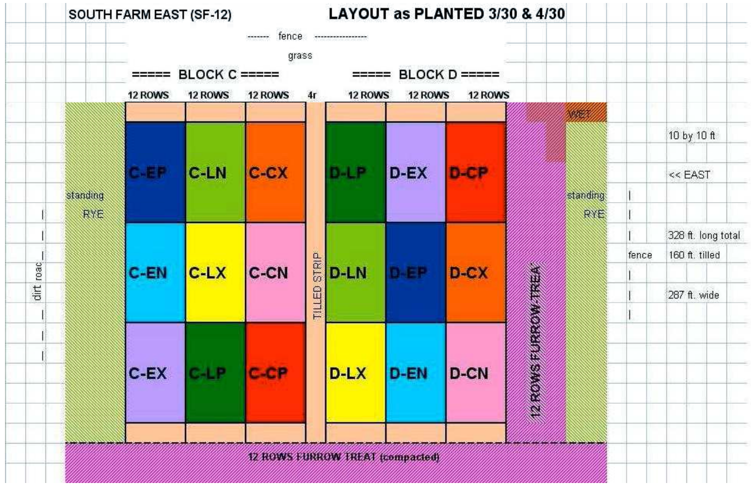
- Treatments were cultivation (whole plot effect)
  - E = early planting, no till
  - L = late planting, no till
  - C = late planting, till

  and amount of mulch used (split plot effect)
  - N = rye cover crop only, none added
  - P = rye cover crop + 1x mulch (straw from rye cover crop)
  - X = rye cover crop + 2x mulch
- The measure of infestation is CPB equivalents per plant stalk = number of adults + $\frac{2}{3}$ of the number of large larvae + $\frac{1}{4}$ of the number of small larvae, averaged over 20 plants per plot
- Split plot design (though not analyzed that way here)
- Four blocks (in two spatially distant sets), nine treatment combinations per plot, so 36 total observations

Experimental design and plots



Experimental design and plots



Experimental design and plots

No till planting into a tall, dense rye cover crop (April)

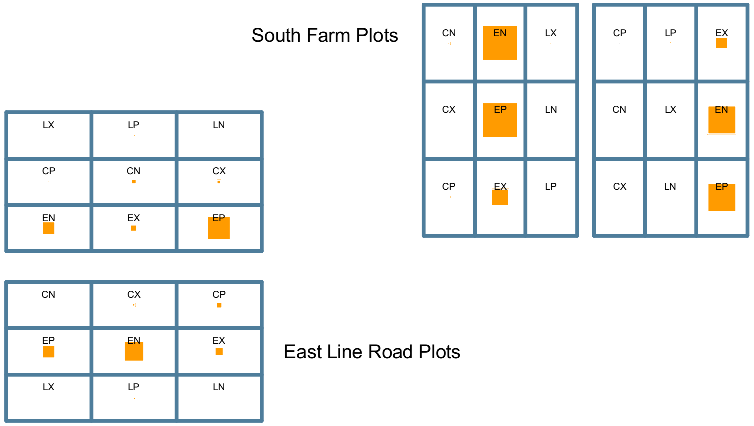
Plots with different mulch treatments (including conventional tilled, background)

Early-planted no-till potatoes with high mulch (front), low mulch, and no added mulch (back) in June

Data collected for Plot A

trt	X	Y	till	plant	CPB	borders	mulch
LN	130	185	no	late	0.12	N, E	none
LP	80	185	no	late	0.00	N	+1x mulch
LX	30	185	no	late	0.02	N, W	+2x mulch
CX	130	155	yes	late	0.33	E	+2x mulch
CN	80	155	yes	late	0.32	-	none
CP	30	155	yes	late	0.19	W	+1x mulch
EP	130	125	no	early	4.10	E, block B	+1x mulch
EX	80	125	no	early	0.67	block B	+2x mulch
EN	30	125	no	early	1.28	W, block B	none

CPB equivalent incidence on plots



R software—spdep package

- *R* software (`http://www.R-project.org`) was used for the analysis
- *spdep* package (main author: Roger Bivand) which has functions for creating spatial weights, tests for spatial autocorrelation (e.g. Moran's I), estimating spatial simultaneous autoregressive (SAR) lag and error models, conditional autoregressive (CAR) models (in a preliminary stage), and includes routines for using sparse matrices
- Installation (on Linux and Windows) of the *spdep* package requires some other *R* packages. For Linux, some of these require compiling C and Fortran code.

Spatial weights

- There are several important decisions to make, e.g. what is a neighbor and how should neighbors be weighted
- *Spdep* can be given the xy coordinates of the middle of each plot and then use a distance cutoff to determine neighbors (weight of 1 for neighbor, 0 if not a neighbor)
- This was tried for various distance cutoffs, and spatial dependence was smaller with a bigger cutoff (bigger neighborhood)
- One can also input a matrix of spatial weights, which could depend on characteristics not directly related to distance (e.g. if plots share a common border). This could be binary (1 if a neighbor, 0 if not a neighbor) or scaled to represent the relationship between neighbors (e.g., length of common border)

Spatial weights

- We used spatial weights that depended on the length of the common border, scaled so the sum of the weights = 36.
- For the same set of residuals, this weighting scheme produced higher estimated spatial dependencies (i.e. seemed to capture more of the spatial correlation)
- Another alternative is to try geostatistical models (e.g. exponential decay, spherical, etc.), these would be based on the distances between the centers of the plots.

Spatial weights for lengths of common border for Block A (not scaled)

plot ID	LN	LP	LX	CX	CN	CP	EP	EX	EN
LN	0	3	0	5	0	0	0	0	0
LP	3	0	3	0	5	0	0	0	0
LX	0	3	0	0	0	5	0	0	0
CX	5	0	0	0	3	0	5	0	0
CN	0	5	0	3	0	3	0	5	0
CP	0	0	5	0	3	0	0	0	5
EP	0	0	0	5	0	0	0	3	0
EX	0	0	0	0	5	0	3	0	3
EN	0	0	0	0	0	5	0	3	0

More on spatial weights

- spatial weights can be symmetric (as in the last example) or asymmetric
- asymmetric weights occur when the spatial weight of the effect of A on B differs from that of B on A. This is reasonable in many circumstances, e.g.,
  - prevailing wind is mostly from one direction
  - the number of neighbors of A is less than that of B, and since B is influenced by many neighbors, the effect of A on B is diluted
- row standardization (i.e. for each observation, the sum of the weights of the neighbors is one) is often suggested, this will lead to asymmetric weights (weights of neighbors will be larger if an observation has fewer neighbors).

Ignore spatial dependencies

- An analysis to determine which treatment, block, and border effects to include in fixed part of model using stepwise regression (based on minimizing AIC)—this is because there were a large number of candidate regressors and only 36 observations to support their estimation.
- Effects were coded as zero-one dummy variables, including some interaction effects
- Since the data were based on counts, a square root transformation was performed. Diagnostics also suggested that this transformation was better than a log or no transformation
- Model: $\sqrt{y} = \boldsymbol{X}\boldsymbol{\beta} + \epsilon$, where
  - $\sqrt{y} =$ square root of Colorado pototo beetle equivalents
  - $\boldsymbol{X}\boldsymbol{\beta} =$ fixed effects
  - $\epsilon =$ uncorrelated random error (noise)

Another look at the data

Data for Block A—border effects (b1–b4) differ depending on block, m1 and m2 represent mulch levels, format for stepwise regression

trt	X	Y	till	pl	CPB	b1	b2	b3	b4	m1	m2
LN	130	185	0	0	0.12	1	0	1	0	0	0
LP	80	185	0	0	0.00	1	0	0	0	1	0
LX	30	185	0	0	0.02	1	0	0	1	0	1
CX	130	155	1	0	0.33	0	0	1	0	0	1
CN	80	155	1	0	0.32	0	0	0	0	0	0
CP	30	155	1	0	0.19	0	0	0	1	1	0
EP	130	125	0	1	4.10	0	1	1	0	1	0
EX	80	125	0	1	0.67	0	1	0	0	0	1
EN	30	125	0	1	1.28	0	1	0	1	0	0

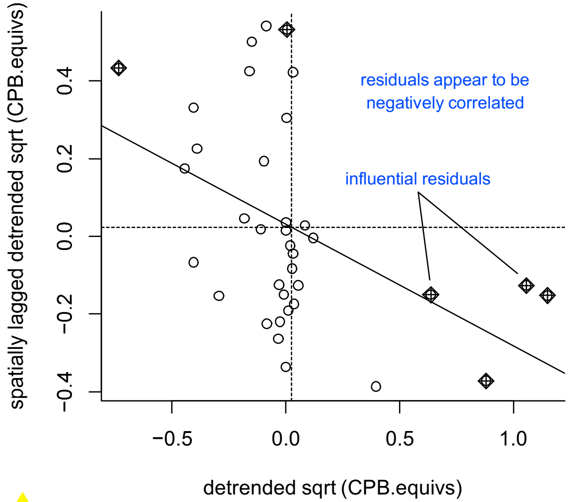
Model from Stepwise regression

```
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)              0.1704     0.0830   2.053 0.051146 .
TILLAGE                  0.6132     0.1114   5.507 1.16e-05 ***
PLANT.TIME.LATE          1.4389     0.1114  12.921 2.67e-12 ***
mulch2                   0.1054     0.1301   0.810 0.426118
b.AB.west               -0.3111     0.1111  -2.800 0.009925 **
block.D                  0.5530     0.1150   4.810 6.73e-05 ***
b.D.west                -0.8014     0.1870  -4.286 0.000255 ***
b.CD                    -0.4682     0.1150  -4.072 0.000439 ***
b.C.east                 0.3205     0.1428   2.245 0.034268 *
b.AB.east                0.1351     0.1004   1.345 0.191332
PLANT.TIME.LATE:mulch2  -0.7913     0.1860  -4.255 0.000276 ***
TILLAGE:mulch2          -0.3239     0.1860  -1.741 0.094438 .

Residual standard error: 0.1938 on 24 degrees of freedom
Multiple R-Squared: 0.9438,     Adjusted R-squared: 0.9181
F-statistic: 36.66 on 11 and 24 DF,  p-value: 2.656e-12
```


Predicted (green) vs. Data (orange)

Moran's I on detrended observations



Which spatial dependency model (for SAR models)?

```
Lagrange multiplier diagnostics for spatial dependence

LMerr =   3.9604, df = 1, p-value = 0.04658
RLMerr = 0.6708, df = 1, p-value = 0.4128
LMlag =   5.7218, df = 1, p-value = 0.01676
RLMlag = 2.4321, df = 1, p-value = 0.1189
SARMA =   6.3925, df = 2, p-value = 0.04092
```

Suggests the lag model might be better than the error model

Simultaneous autoregressive spatial models

There are two basic models fit by *spdep*

- spatial simultaneous autoregressive *error* models

$$y = X\beta + u, u = \lambda W u + \epsilon$$

where

- $y =$ square root of Colorado pototo beetle equivalents
- $X\beta =$ fixed effects
- $u =$ correlated errors with two components
- $\lambda =$ autoregressive error parameter
- $Wu =$ weighted vector of neighboring *residuals* (describes which *residuals* of the neighbors the *residual* of the observation is correlated with and how they are weighted)
- $\epsilon =$ uncorrelated random error (noise)

Simultaneous autoregressive spatial models

- spatial simultaneous autoregressive *lag* models

$$y = \rho W y + X\beta + \epsilon$$

where (for the new terms)

- $\rho =$ autoregressive lag parameter
- $Wy =$ weighted vector of neighbors (describes which neighbors the observation is correlated with and how they are weighted)

Comparison of fixed effects estimates

effect	linear model (SE)	error (SE)	lag (SE)
(Intercept)	0.17 (0.08)	0.18 (0.06)	0.36 (0.10)
TILLAGE	0.61 (0.11)	0.55 (0.08)	0.61 (0.08)
PLANT.LATE	1.44 (0.11)	1.43 (0.08)	1.45 (0.08)
mulch2	0.11 (0.13)	0.08 (0.10)	0.13 (0.10)
b.AB.west	-0.31 (0.11)	-0.31 (0.07)	-0.38 (0.09)
block.D	0.55 (0.12)	0.52 (0.07)	0.65 (0.10)
b.D.west	-0.80 (0.19)	-0.71 (0.13)	-0.85 (0.14)
b.CD	-0.47 (0.12)	-0.41 (0.08)	-0.58 (0.10)
b.C.east	0.32 (0.14)	0.33 (0.09)	0.30 (0.11)
b.AB.east	0.14 (0.10)	0.10 (0.06)	0.09 (0.08)
PLANT.LATE:mulch2	-0.79 (0.19)	-0.73 (0.13)	-0.79 (0.14)
TILLAGE:mulch2	-0.32 (0.19)	-0.20 (0.14)	-0.27 (0.14)

Errorsarlm vs. Lagsarlm

Error model:

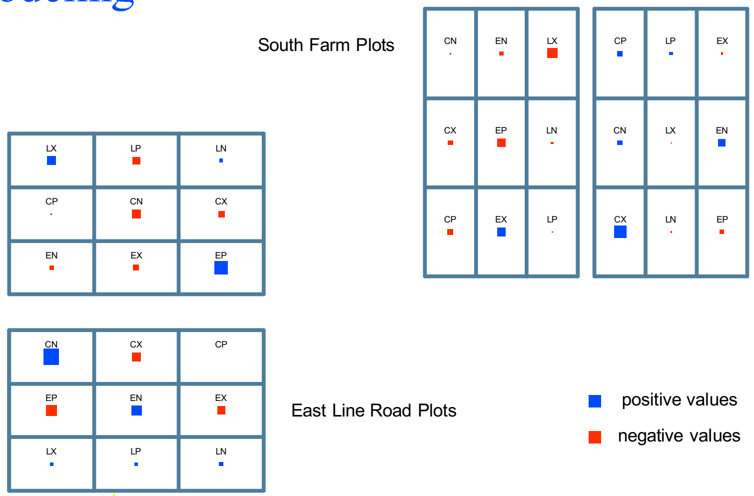
```
Lambda: -0.46916 LR test value: 5.7618 p-value: 0.016379
Asymptotic standard error: 0.14247 z-value: -3.293 p-value: 0.00099118
Log likelihood: 18.17689 for error model
ML residual variance (sigma squared): 0.019354, (sigma: 0.13912)
Number of parameters estimated: 14
AIC: -8.3538, (AIC for lm: -4.592)
```

Lag model:

```
Rho: -0.24002 LR test value: 6.1499 p-value: 0.013142
Asymptotic standard error: 0.091073 z-value: -2.6355 p-value: 0.008401
Log likelihood: 18.37094 for lag model
ML residual variance (sigma squared): 0.020609, (sigma: 0.14356)
Number of parameters estimated: 14
AIC: -8.7419, (AIC for lm: -4.592)
```

Lag: predicted (green), data (orange)

Data minus fixed effects: What $\rho W y$ is modeling



Conditional Spatial Autoregressive Model

In CAR models, an observation's value is conditioned on neighboring values. This is one representation for the model:

$$E(y_i|\mathbf{y}_{*i}) = \mathbf{X}\beta + \lambda\mathbf{W}(\mathbf{y}_{*i} - \mu_{*i})$$

where

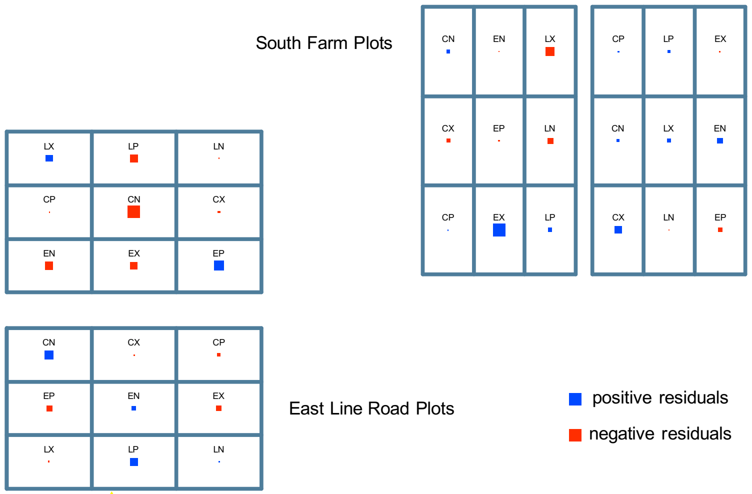
- y_i = square root of Colorado pototo beetle equivalents
- $\mathbf{X}\beta$ = fixed effects for y_i
- $\mathbf{y}_{*i}$ = neighbors of y_i ($*i$ = not including observation i)
- λ = autoregressive parameter
- $\mathbf{W}(\mathbf{y}_{*i} - \mu_{*i})$ = weighted vector of mean adjusted neighbors

Estimation results from CAR model

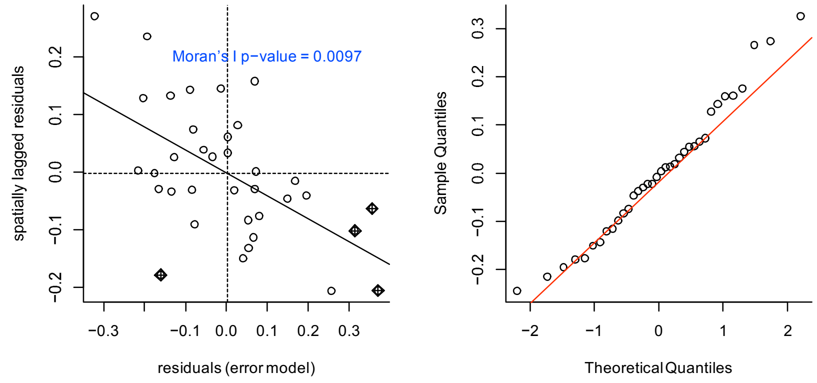
```
                     Estimate Std. Error z value  Pr(>|z|)
(Intercept)          0.176337   0.056978  3.0948 0.0019694 **
TILLAGE              0.560195   0.080908  6.9238 4.395e-12 ***
PLANT.TIME.LATE      1.430348   0.079894 17.9030 < 2.2e-16 ***
mulch2               0.079565   0.100107  0.7948 0.4267288
b.AB.west           -0.309661   0.071309 -4.3426 1.408e-05 ***
block.D              0.523969   0.071477  7.3306 2.292e-13 ***
b.D.west            -0.728253   0.133909 -5.4384 5.375e-08 ***
b.CD                -0.427340   0.081758 -5.2269 1.724e-07 ***
b.C.east             0.326350   0.095643  3.4122 0.0006445 ***
b.AB.east            0.106542   0.064655  1.6479 0.0993820 .
PLANT.TIME.LATE:mulch2 -0.732866   0.137060 -5.3470 8.941e-08 ***
TILLAGE:mulch2      -0.210189   0.142252 -1.4776 0.1395201

Lambda: -0.70764 LR test value: 5.3146 p-value: 0.021147
Log likelihood: 17.95331    AIC: -7.9066
ML residual variance (sigma squared): 0.018909, (sigma: 0.13751)
```

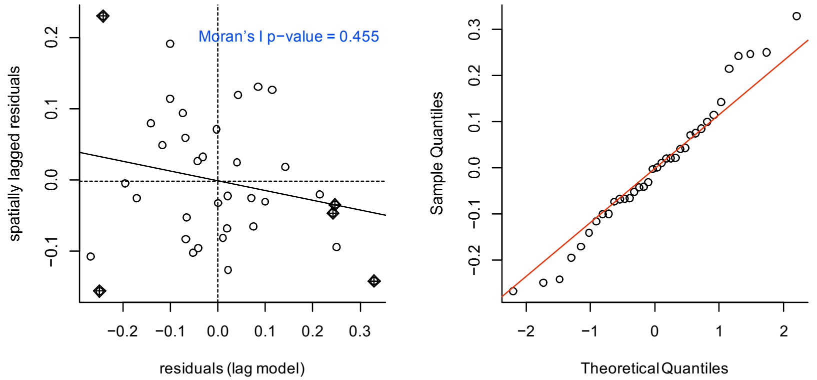
Residuals of CAR model



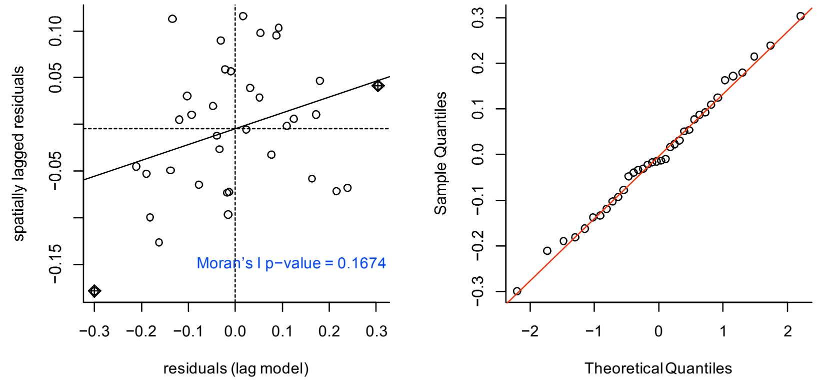
Diagnostics, SAR error model: Moran and QQnorm plots



Diagnostics, SAR lag model: Moran and QQnorm plots



Diagnostics, CAR model: Moran and QQnorm plots



Conclusions I.

- Estimates of fixed effects parameters were similar for all models
- Standard errors of fixed effects parameters were smaller when spatial dependencies were taken into account
- For these data, judging by AIC, the spatial dependencies appeared to be captured adequately by all spatial models discussed, and there is a substantial improvement over the model that ignores spatial dependencies
- The CAR model seems to have better behaved residuals

Conclusions II.

- Why a negative correlation between neighboring plots? Our best guess is that the beetle population is locally redistributing to favorable plots after departing unfavorable ones. So, the relative accumulation of beetle numbers on a particular treatment combination depends on which neighbors it has.
- In field season 2006 we will be looking at individual beetle behavior including arrival and residence time in different treatments, which should yield insight into this spatial pattern.