Occupancy modeling with spatial dependence

Iryna Dronova and Lu Liang

October 22, 2012

Spatial autocorrelation & spatial dependence

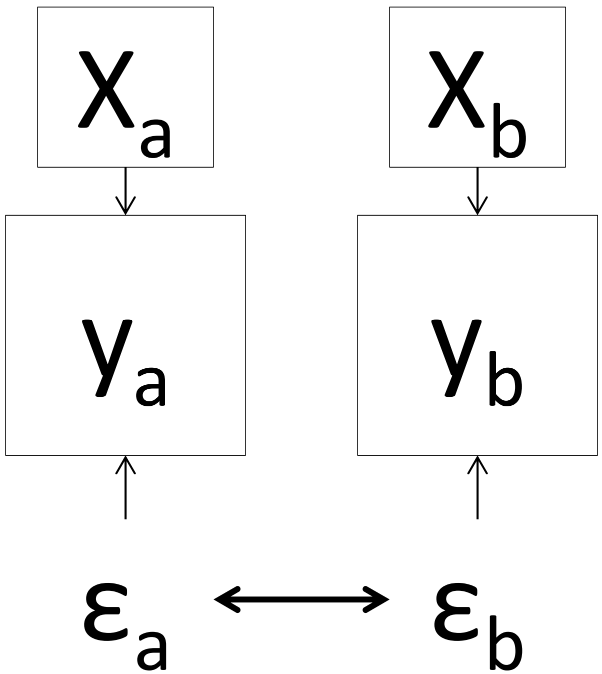
- Correlation between values of the same variable at different spatial locations
- (Dis)similarity in values among locations varies with distance among them

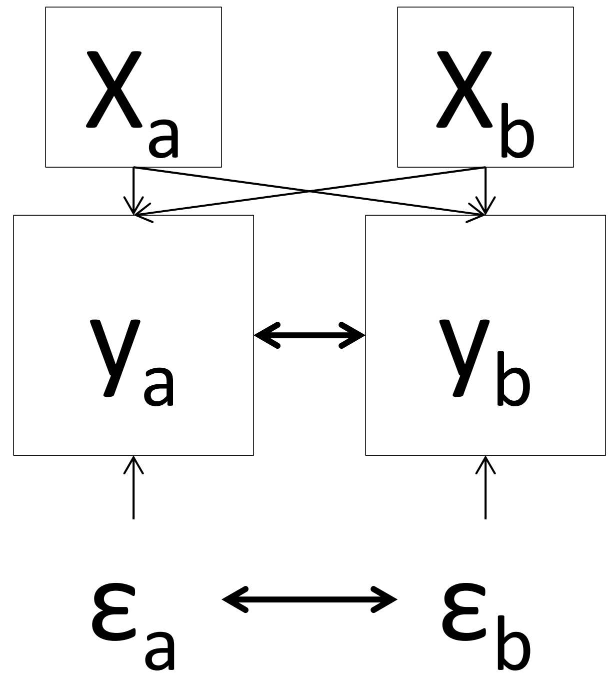
But: not the same as trend due to covariate!

- **Tobler's First Law of Geography:**

"Everything is related to everything else, but near things are more related than distant things"

2 common types of spatial dependence in the context of simple regression models:





- Spatially structured variables are missing from the model
- Inadequate sampling units a & b

- Second-order spatial process
- Interaction between locations varies with distance/adjacency

Why important:

- Violates the assumptions of independent and identically distribute errors in standard analyses
- If unaccounted for, regression slopes/model coefficient estimates may be unreliable

But

- May illuminate real spatial processes such as dispersal and competition
- May be used to generate new hypotheses

Spatial dependence in occupancy:

- Probability of presence, occupancy and even detectability may depend on whether or not neighbor sites are occupied
- Underlying mechanisms of connectivity, dispersal or competition
- Distance between sites may affect the chance of detecting the same individual versus different ones
- Habitat properties may be spatially correlated, too

Hines et al. 2010. Tigers on trails

- Spatial dependence in non-random sampling design with spatial replication within a sampling unit

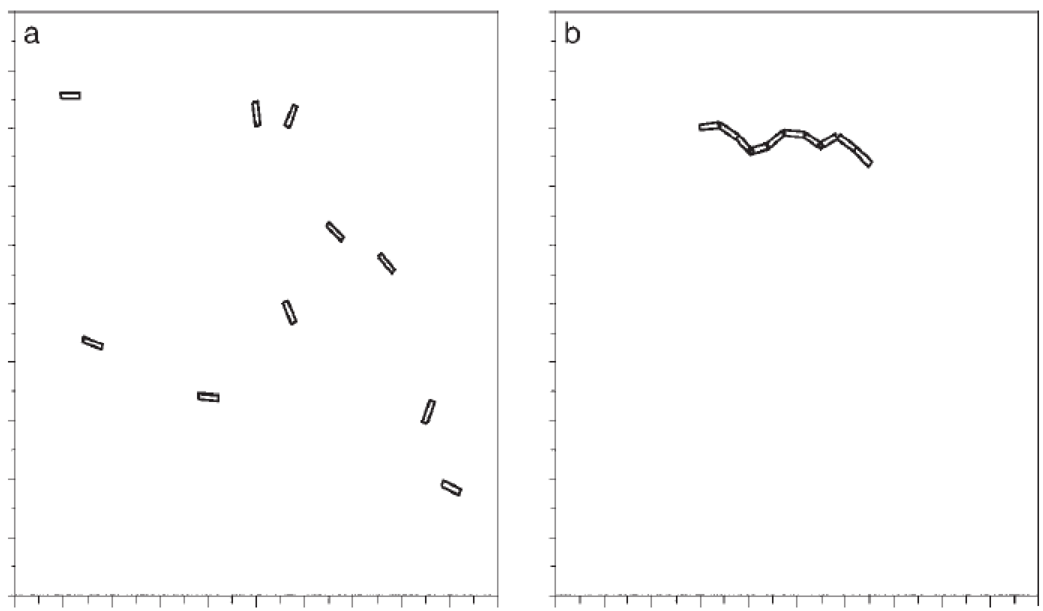


FIG. 1. Two sampling designs employing spatial replication to draw inference about occupancy and detection probabilities: (a) design 1 depicts an example of random sampling; (b) design 2 depicts sampling of segments along a trail, likely producing correlated spatial replicates.

- Applicable to other studies, replace "previous segment" by "nearest neighbor" (even in random sampling)

Single-season with spatial dependence

- Data are likely to be correlated on adjacent sample sites (trail segments)

  [ψ(.), θ(.), θ'(.), p(.)]

- Probability of presence with spatial dependence:

  θ': Pr (presence | present at previous segment)

  θ: Pr (presence | <u>not present at previous segment</u>)

True absence at previous segment

Present but not detected at previous segment

Simple case: 2 segments on a trail

$$\mathbf{Pr(h_j{=}11){=}\ \psi\underline{\theta}p\theta'p}$$

Does this mean that segment 1 can only have θ because it does not have a “previous” segment to it?

$$\mathbf{Pr(h_j{=}10)}{=}\ \psi\theta p[\underbrace{(1-\theta')}_{\text{Truly absent from segment2}}+\underbrace{(1-p)\theta'}_{\text{Present on segment 2 but not detected there}}]{=}\ \mathbf{\psi\theta p(1-\theta' p)}$$

$$\mathbf{Pr(h_j{=}01){=}\ \psi[\underbrace{\theta(1-\theta)}_{\text{Truly absent from segment 1}}+\ \underbrace{(1-p)\theta'}_{\text{Present on segment 1 but not detected there}}]p}$$

2 segments on a trail (continued)

Pr(h_j=00)= 5 possibilities =

1) Site not occupied ($1-\psi$)
2) Segment 1 occupied but not detected; segment 2 not occupied
3) Segment 1 occupied, but not detected, segment 2 occupied, but not detected
4) Segment 1 not occupied, segment 2 not occupied
5) Segment 1 not occupied, segment 2 occupied but not detected

$$=(1-\psi)+\psi[\overbrace{\theta(1-p)}^{\text{Segment 1 occupied but not detected}}\overbrace{(\theta'(1-p)+(1-\theta'))}^{\text{then segment 2 either not detected or not occupied, use }\theta'}+\overbrace{(1-\theta)}^{\text{Segment 1 not occupied}}\overbrace{(\theta(1-p)+(1-\theta)}^{\text{then segment 2 either not detected or not occupied, use }\theta}]=$$

$$\mathbf{=(1-\psi)+\psi[\theta(1-p)(1-\theta' p)+(1-\theta)(1-\theta p)]}$$

5 segments example

For detection history h_j=01011

$$Pr(h_j=01011)=$$

$$=\psi[(1-\theta)\theta+\theta(1-p)\theta']\times p[(1-\theta')\theta+\theta'(1-p)\theta']p\theta' p$$

Detected at segment 4

Present on segment 2 but truly absent from segment 1

Present on both segments 1 & 2 but not detected at segment 1

Detected at segment 2

Truly absent from segment 3 though detected at segment 2, and present at segment 4

Present on both segments 3 & 4 but not detected at segment 3

Detected at segment 5 given presence at segment 4

Adapting for a more general case of one nearest neighbor (no segments):

- Unlike in example with segments, the site and neighbor site are now **different sites/sampling units** and no longer have “equal” roles (hence no more θ for the site p or (1-p))
- What is the Pr(presence) at the site given presence or non-presence at the neighbor site?

θ': Pr (presence | present at neighbor site)

θ: Pr (presence | not present at neighbor site)

p: assume constant for all sites

Pr(h_j=site, neighbor site):

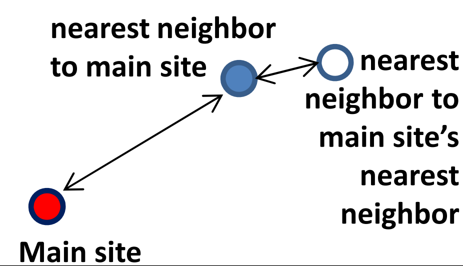
$Pr(h_j=1,1)= \psi p\theta' p$

$Pr(h_j=1,0)= \psi p[\theta+(1-p)\theta']$

$Pr(h_j=0,1)= (1- \psi) + \psi(1-p)\theta' p$

$Pr(h_j=0,0)= (1- \psi) + \psi(1-p)[\theta + (1-p)\theta']$

Note: the “main site” may or may NOT be the nearest neighbor to its own nearest neighbor site:



What do you think?

Note: this is not the way spatial model is used in PRESENCE, the software follows Hines et al. framework.

Even more general set-up (Hines et al.)

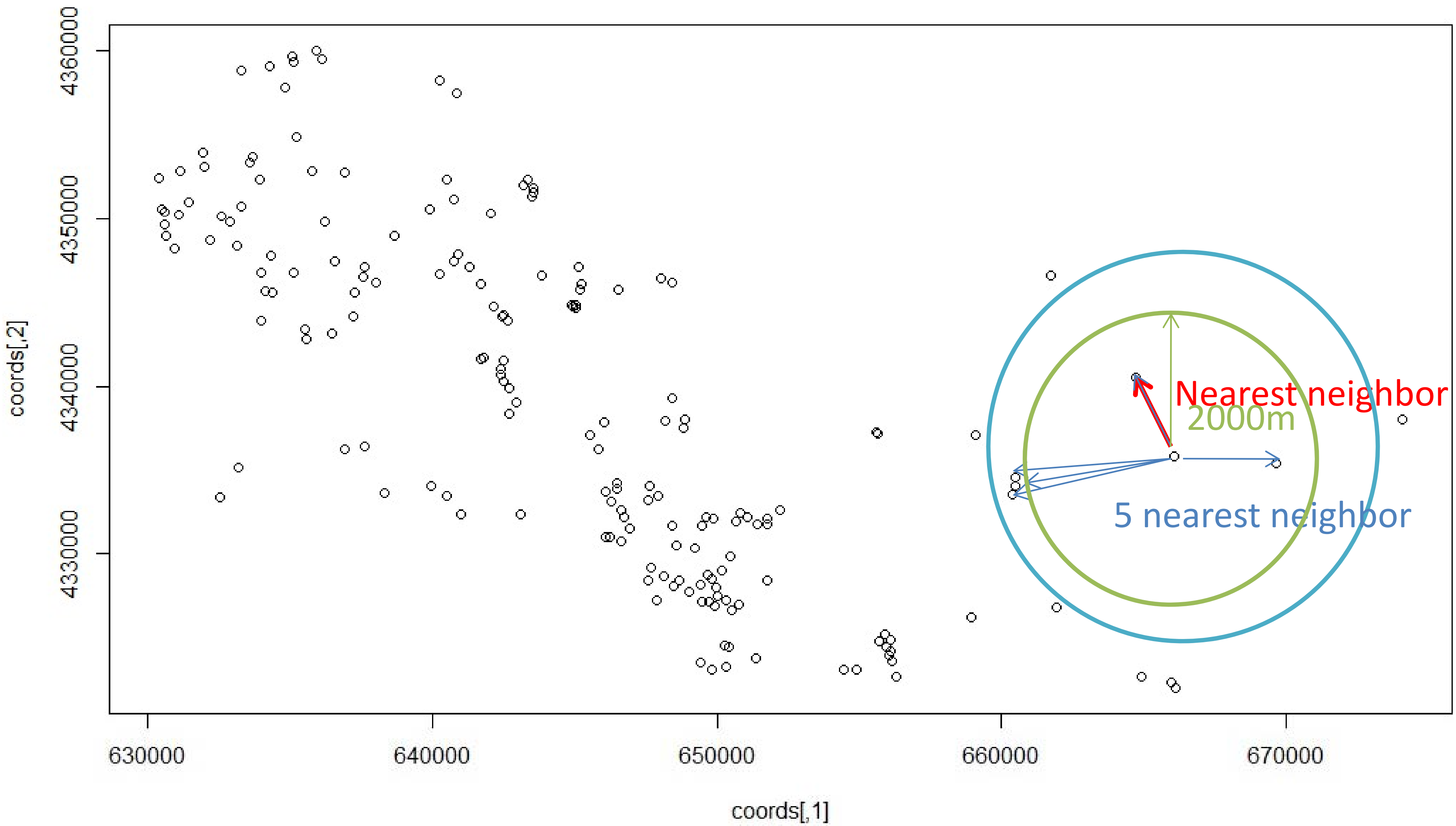
p=Pr (detection | sample unit occupied, species not detected at previous segment)

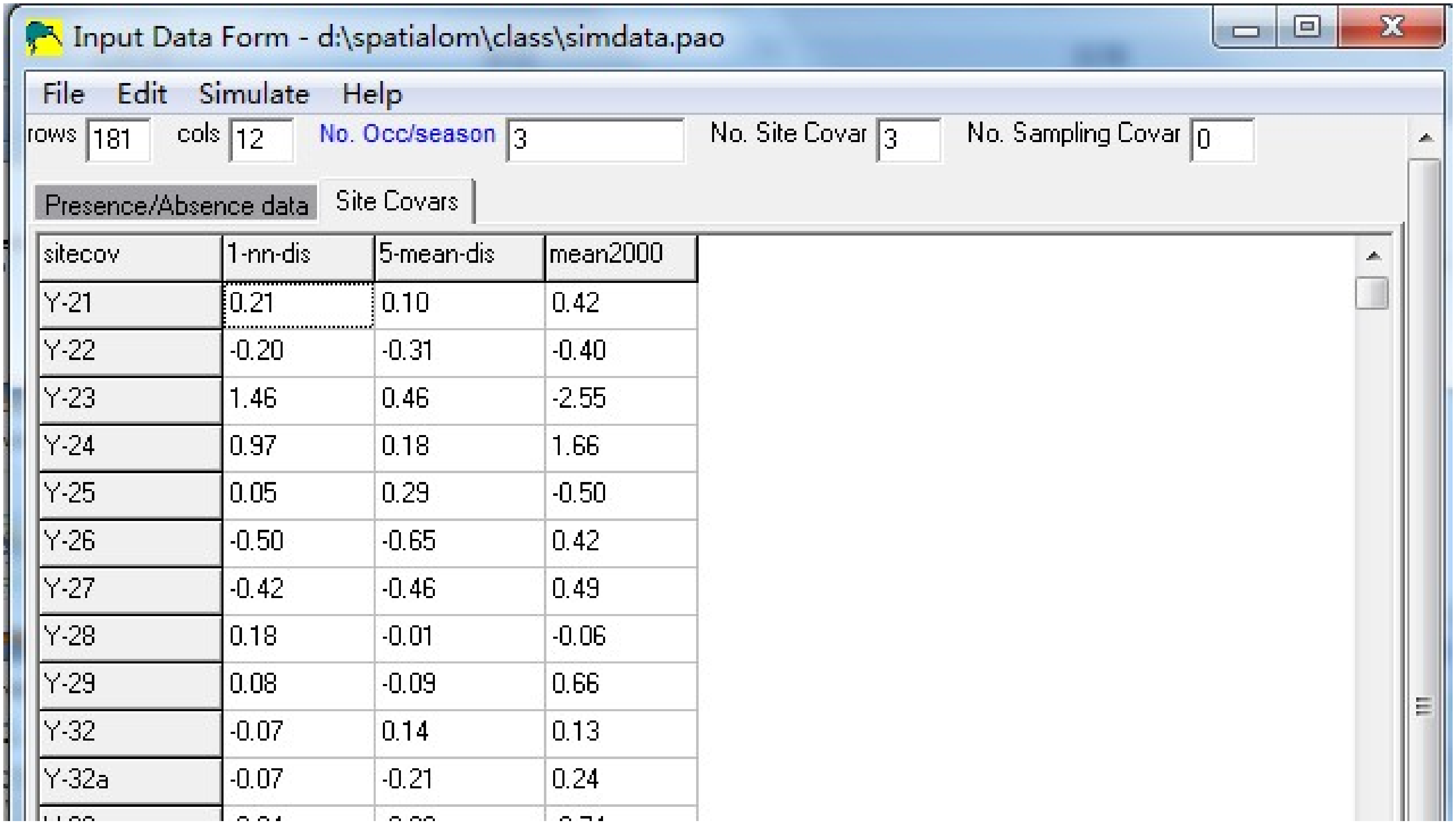
p'=Pr (detection | sample unit occupied, species detected at previous segment

Pr (hj=01011)=ψ(1-p)p(1-p')pp'

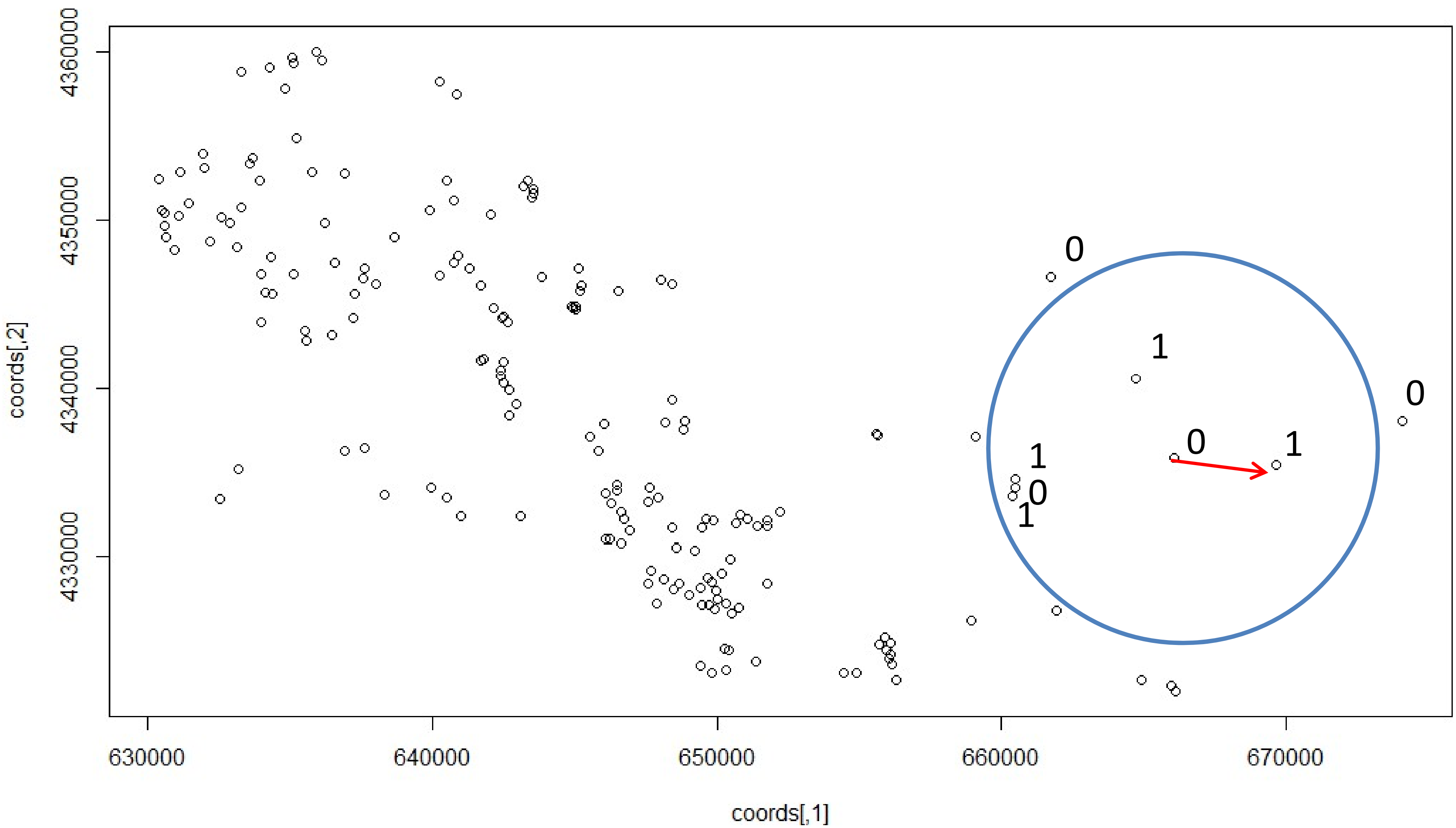
Pr (hj=00000)=(1- ψ)+ $\prod_{t=1}^{5}(1-p)$

Using spatial distance as covariates




Input Data Form - d:\spatialom\class\simdata.pao
File
Edit
Simulate
Help
rows 181
cols 12
No. Occ/season 3
No. Site Covar 3
No. Sampling Covar 0
Presence/Absence data
Site Covars
sitecov 1-nn-dis 5-mean-dis mean2000
Y-21 0.21 0.10 0.42
Y-22 -0.20 -0.31 -0.40
Y-23 1.46 0.46 -2.55
Y-24 0.97 0.18 1.66
Y-25 0.05 0.29 -0.50
Y-26 -0.50 -0.65 0.42
Y-27 -0.42 -0.46 0.49
Y-28 0.18 -0.01 -0.06
Y-29 0.08 -0.09 0.66
Y-32 -0.07 0.14 0.13
Y-32a -0.07 -0.21 0.24

Using distance to define neighbors



Detection history:

- Using nearest neighbor: 01
- Using five nearest neighbor: 101110

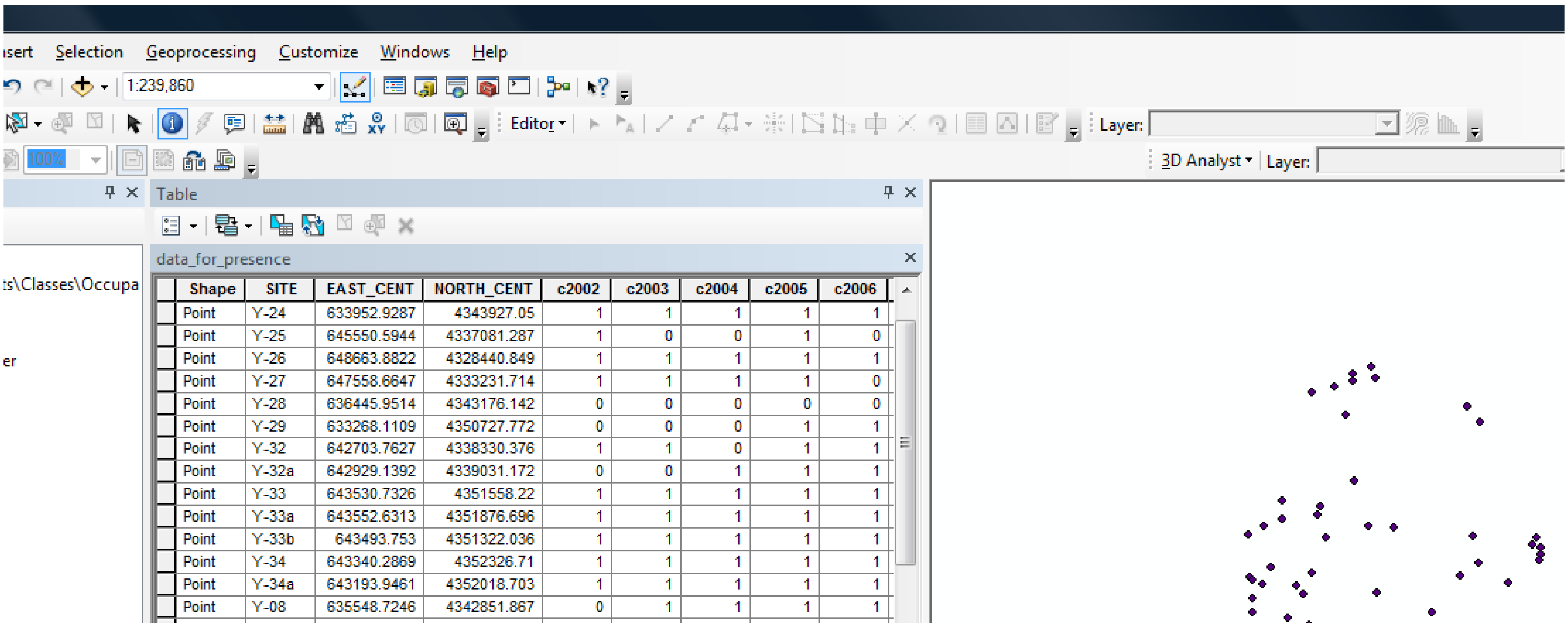
Models with neighborhood/connectivity covariates

- Considering **only occupied nearest neighbors**
  - Distance to nearest neighbor
  - Buffer metrics (summaries of occupied patch areas within a fixed-distance neighborhood around each site)
  - Incidence function (take into account distances to all source populations)

- Moilanen, A. and M. Nieminen. 2002. Simple connectivity measures in spatial ecology. Ecology 84:1131-1145.
- Assume dispersal distance ($1/\alpha$) for Rails 8 km, $0.5/\alpha$=4 km

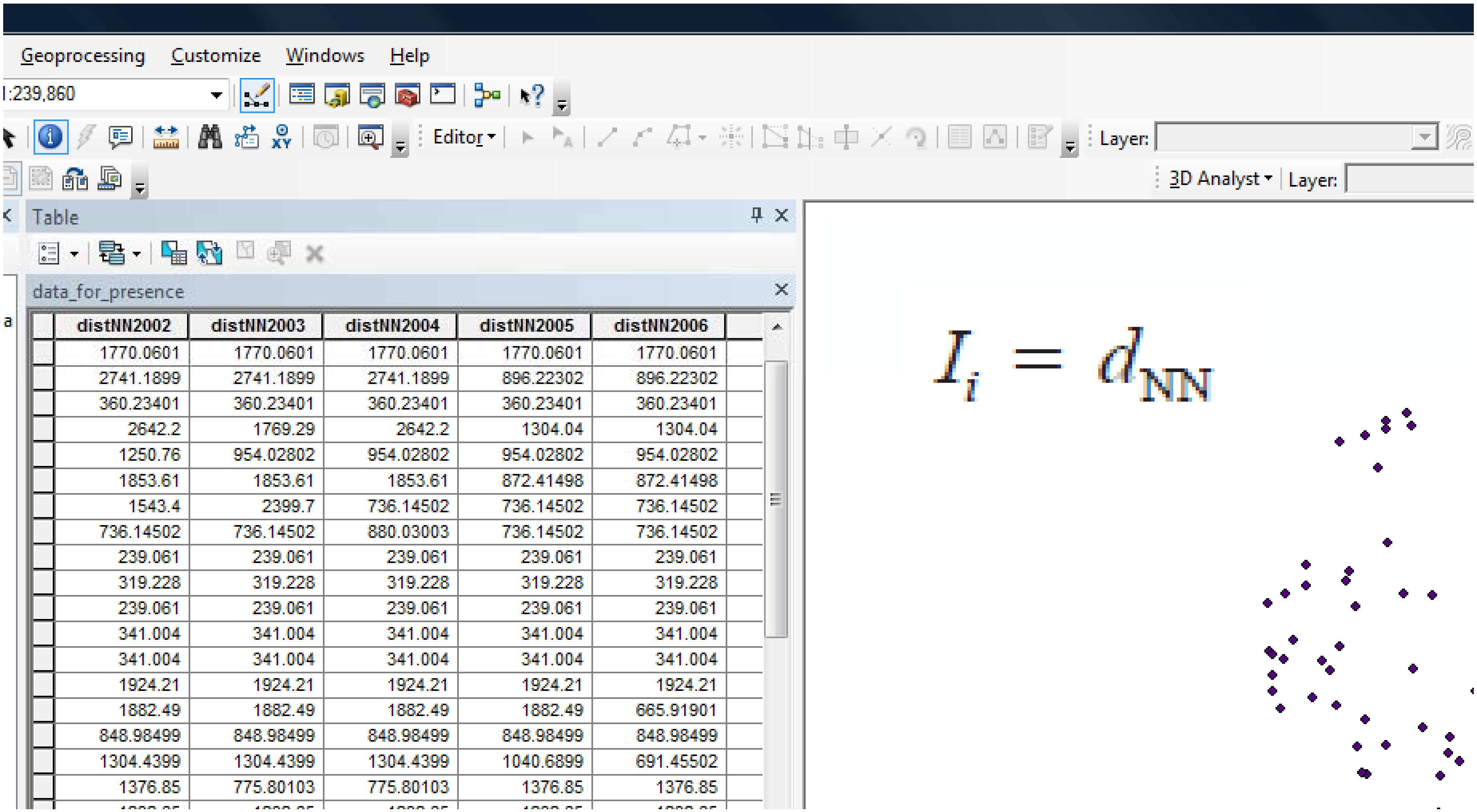
Preprocessing in ArcGIS 10

- Ignore within-year site visits, consider only whether species detected at least once a year or not



Shape	SITE	EAST_CENT	NORTH_CENT	c2002	c2003	c2004	c2005	c2006
Point	Y-24	633952.9287	4343927.05	1	1	1	1	1
Point	Y-25	645550.5944	4337081.287	1	0	0	1	0
Point	Y-26	648663.8822	4328440.849	1	1	1	1	1
Point	Y-27	647558.6647	4333231.714	1	1	1	1	0
Point	Y-28	636445.9514	4343176.142	0	0	0	0	0
Point	Y-29	633268.1109	4350727.772	0	0	0	1	1
Point	Y-32	642703.7627	4338330.376	1	1	0	1	1
Point	Y-32a	642929.1392	4339031.172	0	0	1	1	1
Point	Y-33	643530.7326	4351558.22	1	1	1	1	1
Point	Y-33a	643552.6313	4351876.696	1	1	1	1	1
Point	Y-33b	643493.753	4351322.036	1	1	1	1	1
Point	Y-34	643340.2869	4352326.71	1	1	1	1	1
Point	Y-34a	643193.9461	4352018.703	1	1	1	1	1
Point	Y-08	635548.7246	4342851.867	0	1	1	1	1

Calculate distance to the closest **occupied** nearest neighbor



distNN2002	distNN2003	distNN2004	distNN2005	distNN2006
1770.0601	1770.0601	1770.0601	1770.0601	1770.0601
2741.1899	2741.1899	2741.1899	896.22302	896.22302
360.23401	360.23401	360.23401	360.23401	360.23401
2642.2	1769.29	2642.2	1304.04	1304.04
1250.76	954.02802	954.02802	954.02802	954.02802
1853.61	1853.61	1853.61	872.41498	872.41498
1543.4	2399.7	736.14502	736.14502	736.14502
736.14502	736.14502	880.03003	736.14502	736.14502
239.061	239.061	239.061	239.061	239.061
319.228	319.228	319.228	319.228	319.228
239.061	239.061	239.061	239.061	239.061
341.004	341.004	341.004	341.004	341.004
341.004	341.004	341.004	341.004	341.004
1924.21	1924.21	1924.21	1924.21	1924.21
1882.49	1882.49	1882.49	1882.49	665.91901
848.98499	848.98499	848.98499	848.98499	848.98499
1304.4399	1304.4399	1304.4399	1040.6899	691.45502
1376.85	775.80103	775.80103	1376.85	1376.85

$$I_i = d_{\mathrm{NN}}$$

Generate buffers at ½ dispersal distance (4km) & sum areas of occupied neighbors within a buffer

Buffer measures:

$$S_i = \sum_{\substack{j \neq i \\ d_{ij} \leq r}} A_j^b$$

$$S_i = A_i^c \sum_{\substack{j \neq i \\ d_{ij} \leq r}} A_j^b$$

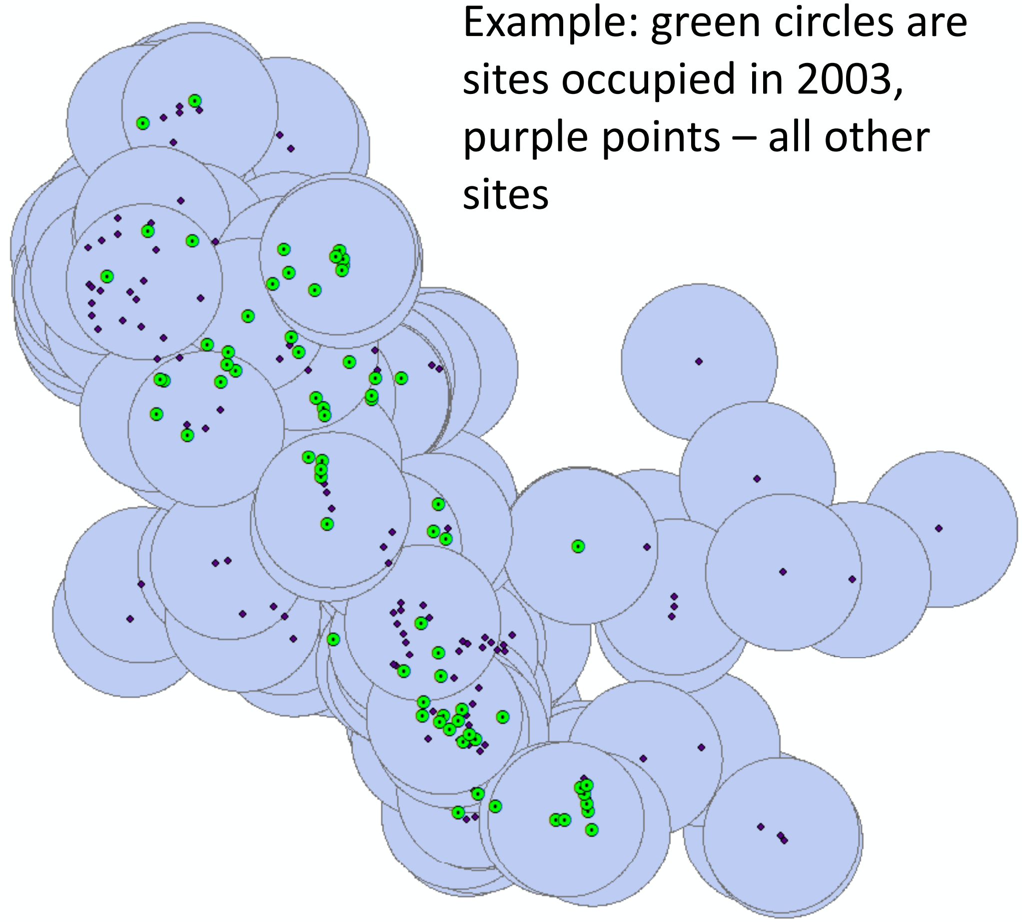
Area of site i

Area of neighbor patch j

Incidence functions:

$$S_i = \sum_{j \neq i} \exp(-\alpha d_{ij}) A_j^b$$

$$S_i = A_i^c \sum_{j \neq i} \exp(-\alpha d_{ij}) A_j^b$$



Example: green circles are sites occupied in 2003, purple points – all other sites

Single-season with connectivity covariates (2002)

Program PRESENCE version 5.2 <121012.0912> (data2002ss.pa3)

File View Run Tools Help

Model	AIC	deltaAIC	AIC wgt	Model Likelihood	no.Par.	-2*LogLike
psi(incfn2),p(nndist)	202.08	0.00	0.9717	1.0000	4	194.08
psi(incfn),p(nndist)	209.47	7.39	0.0241	0.0248	4	201.47
psi(buffer),p(nndist)	214.24	12.16	0.0022	0.0023	4	206.24
psi(incfn2),p(.)	215.88	13.80	0.0010	0.0010	3	209.88
psi(incfn2),p(incfn)	217.29	15.21	0.0005	0.0005	4	209.29
psi(incfn2),p(buffer)	217.83	15.75	0.0004	0.0004	4	209.83
psi(incfn),p(.)	221.33	19.25	0.0001	0.0001	3	215.33
psi(NNDIST),p(.)	223.25	21.17	0.0000	0.0000	3	217.25
psi(.),p(INCFN2)	229.02	26.94	0.0000	0.0000	3	223.02
psi(buffer),p(.)	230.25	28.17	0.0000	0.0000	3	224.25
psi(.),p(.)	231.99	29.91	0.0000	0.0000	2	227.99

notepad C:\Users\Irisha\Documents

pres_psi(incfn2)_p(nndist).out - Notepad

File Edit Format View Help

```
Custom Model:

Number of parameters          = 4
Number of significant digits  =   6.9

Model has been fit using the logistic link.


Number of parameters          = 4
Number of function calls          = 148
-2log(likelihood)             = 194.0835
AIC                           = 202.0835
LikeNRSig=6 eps=0.01 ETA=1e-013

Untransformed Estimates of coefficients for covariates (Beta's)
======================================================================
                                          estimate     std.error
A1    psi                             :    0.129155     0.327465
A2    psi.INCF202                     :    0.059454     0.025938
B1    p1                              :    2.357007     0.299099
B2    p1.NNDIST02                     :   -0.000701     0.000256

============================================================

   Individual Site estimates of <psi>
                 Site              estimate   Std.err    95% conf. interval
psi              1 site 1         :  0.7243    0.0567      0.6009 - 0.8210
psi              2 site 2         :  0.9758    0.0321      0.7373 - 0.9983
psi              3 site 3         :  0.7535    0.0600      0.6187 - 0.8520
psi              4 site 4         :  1.0000    0.0002      0.8787 - 1.0000
psi              5 site 5         :  0.5682    0.0709      0.4275 - 0.6987
psi              6 site 6         :  0.6367    0.0568      0.5199 - 0.7393
psi              7 site 7         :  0.7147    0.0559      0.5942 - 0.8108
psi              8 site 8         :  0.5374    0.0799      0.3821 - 0.6857
```