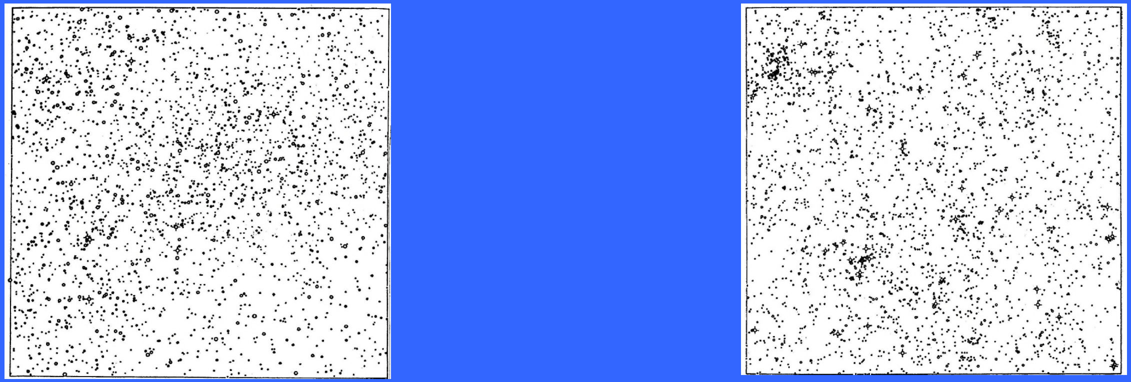


Synthetic plots: history and examples

David R. Brillinger
Statistics Department
University of California, Berkeley

www.stat.berkeley.edu/~brill
brill@stat.berkeley.edu

SECTIONS

I. Model appraisal methods

II. Synthetic plots

III. Spatial p.p. – *galaxies*

IV. Time series – *river flow*

V. Spatial-temporal p.p. -*wildfires*

VI. Trajectories – *seals, elk*

VII. Summary and discussion

?. Explanatories

I. MODEL APPRAISAL

Science needs appraisal methods

Cycle:

Model construction <--> model appraisal

Is model compatable with the data?

Classical chi-squared (df correction)

The method of synthetics
Neyman et al

II. SYNTHETIC PLOTS

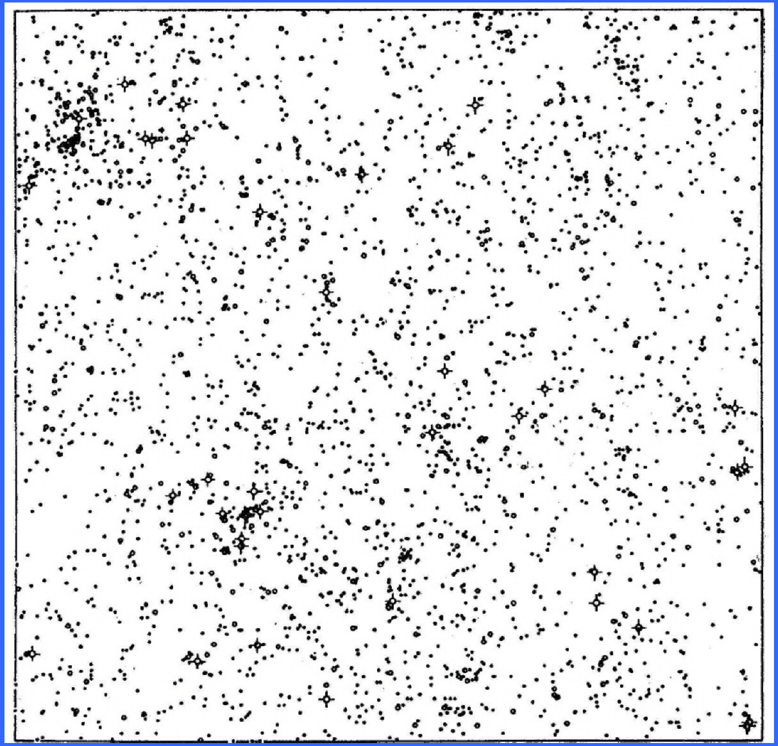
Simulate realization of fitted model

Put real and synthetic side by side

Assessment

- Turing test?
- Compute same quantity for each?

III. SPATIAL P. P. - *galaxies*



Neyman, Scott and Shane (1953) On the spatial distribution of galaxies … *Astr J*, 117, 92-133

"… Figure 1 was constructed assuming …, the Poisson law …."

"… it was decided to produce a synthetic plot …"

"When the calculated scheme of distribution was compared with the actual distribution of galaxies recorded in Shane's photographs of the sky [see page 192], it became apparent that the simple mechanism could not produce a distribution resembling the one we see. In the real universe there is a much more pronounced tendency for galaxies to be grouped in clusters."

Neyman, Scott & Shane (1954) On the index of clumpiness … *Astr. J. Suppl.* 1, 269-294.

“In the third paper …, it was shone that the visual appearance of a ‘synthetic’ photographic plate, obtained by means of a large scale sampling experiment conforming exactly with the assumptions of the theory, is very similar to the actual plate. The only difference noticed between the two is concerned with the small-case clumpiness of images of galaxies.”

In summary:

Data $\{(x_j, y_j)\}$, spatial point process

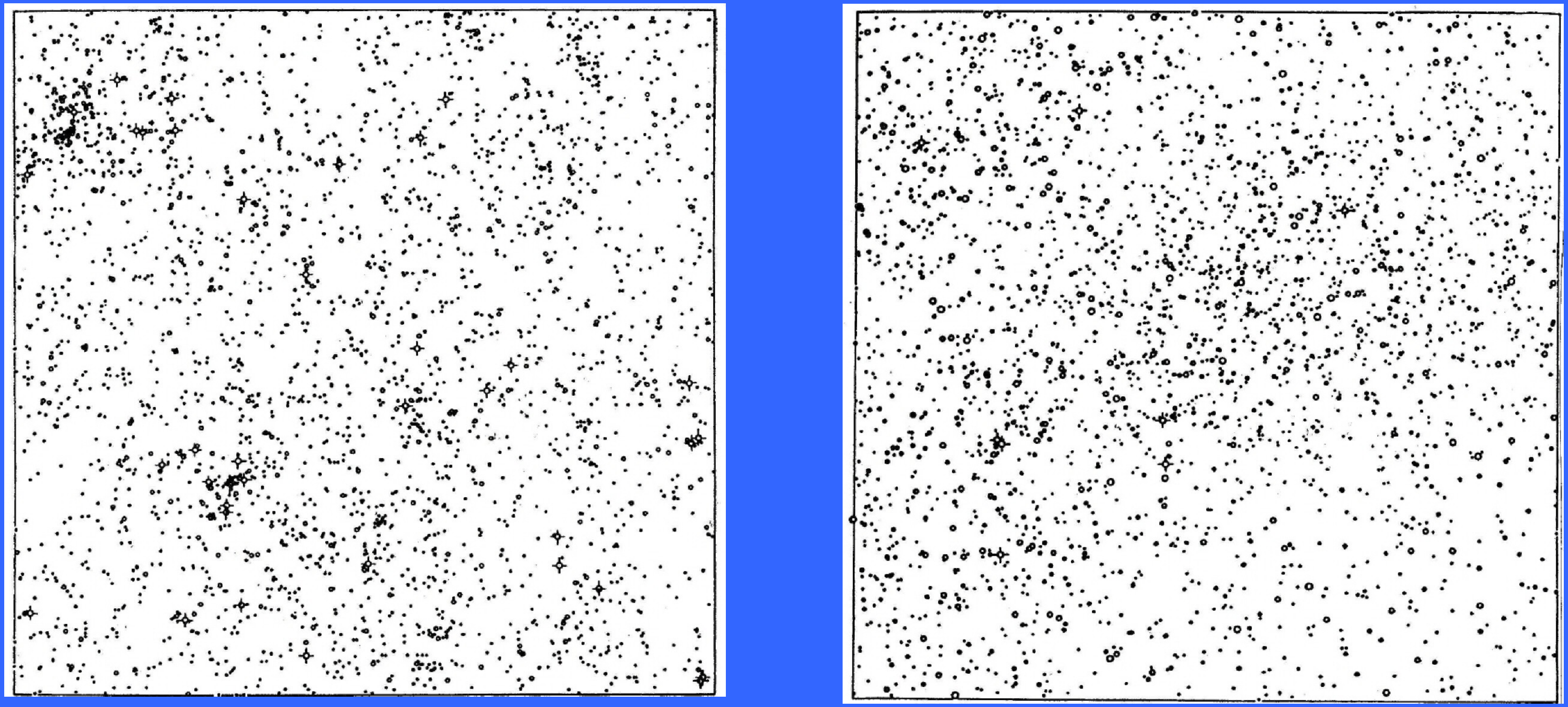
1. Poisson
   rejected (visually)

2. Clustering (Neyman-Scott process)
   rejected (visually)

3. More clustering

Detail: counting error, variation in limiting magnitude

Results – *Scientific American* 1956



Turing test?

Comparison – Scott et al (1953)

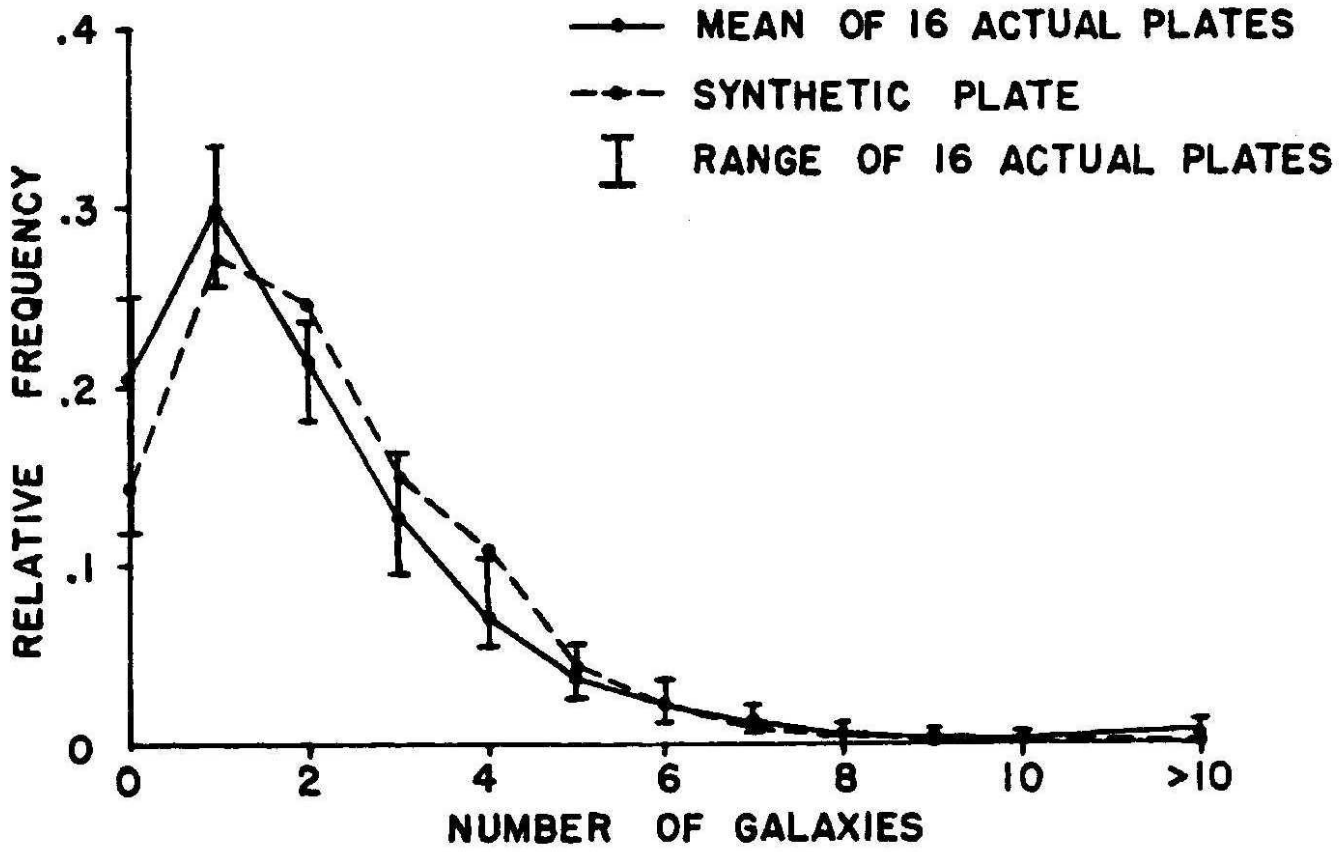
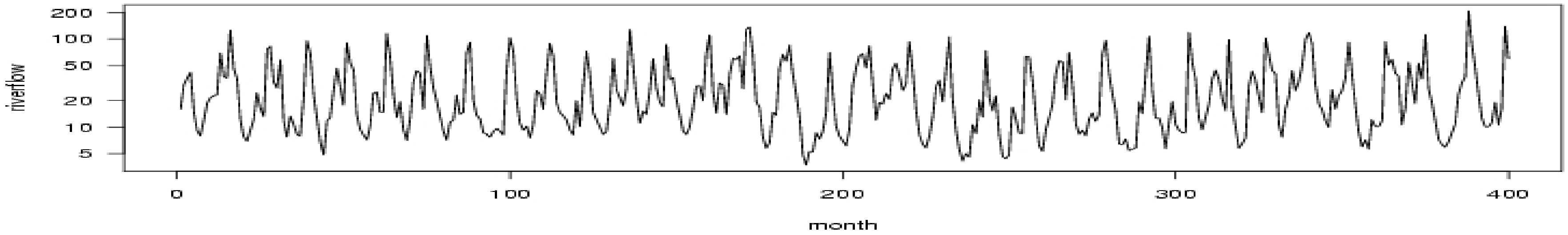


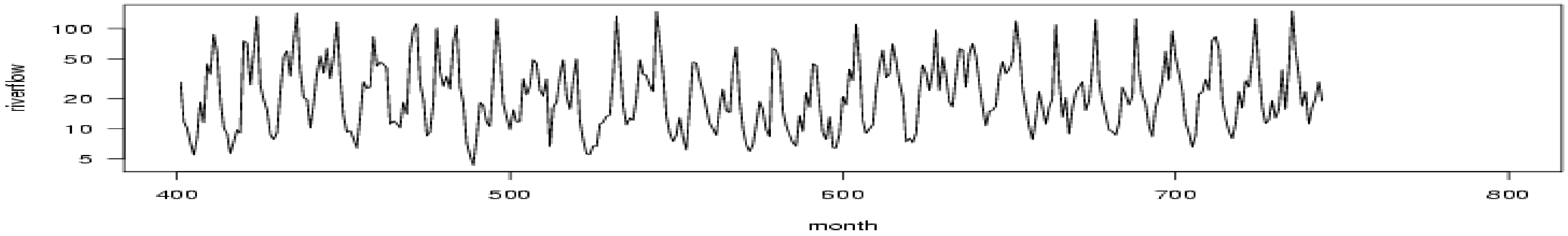
FIG. 8.—Relative frequency of 10′ × 10′ cells containing 0, 1, 2, . . . , galaxies. Comparison of synthetic plate with mean of 16 actual plates.

IV. TIME SERIES – Saugeen River

Average monthly flow 1915 – 1976

Walkerton





Data: *{y(t)}, time series*

Hippel and McLeod

Periodic autoregression (PAR)

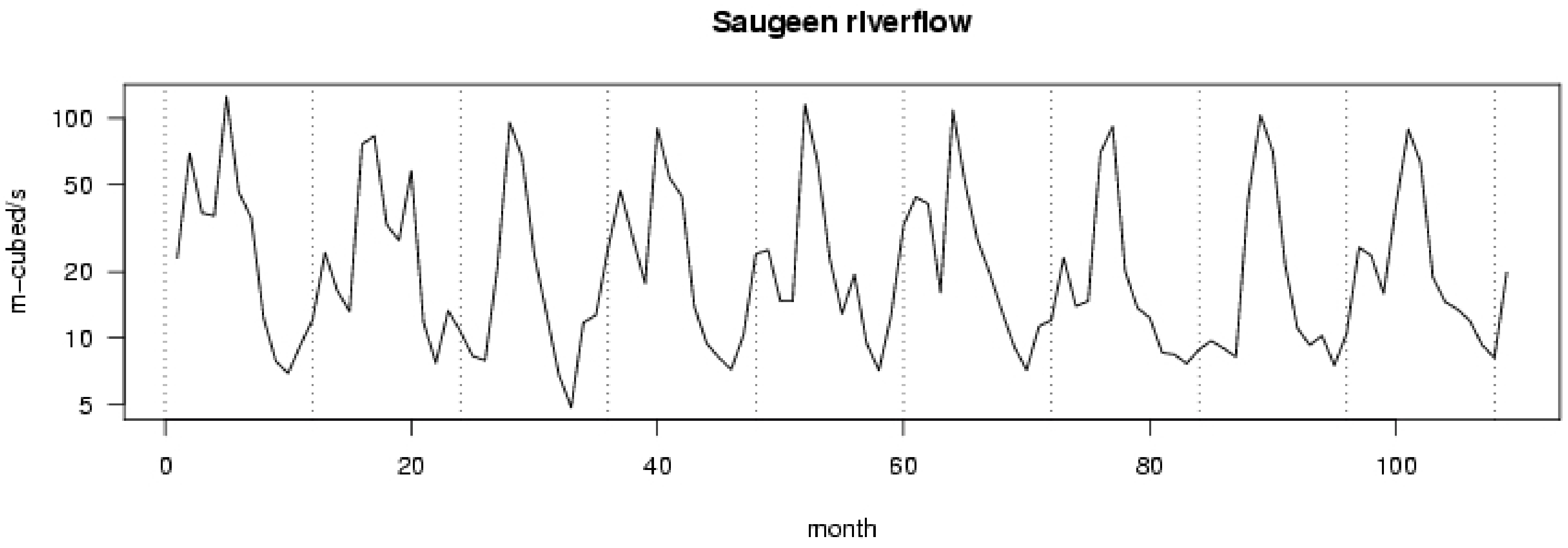
Stack years - 62 by 12 matrix (Buys-Ballot)

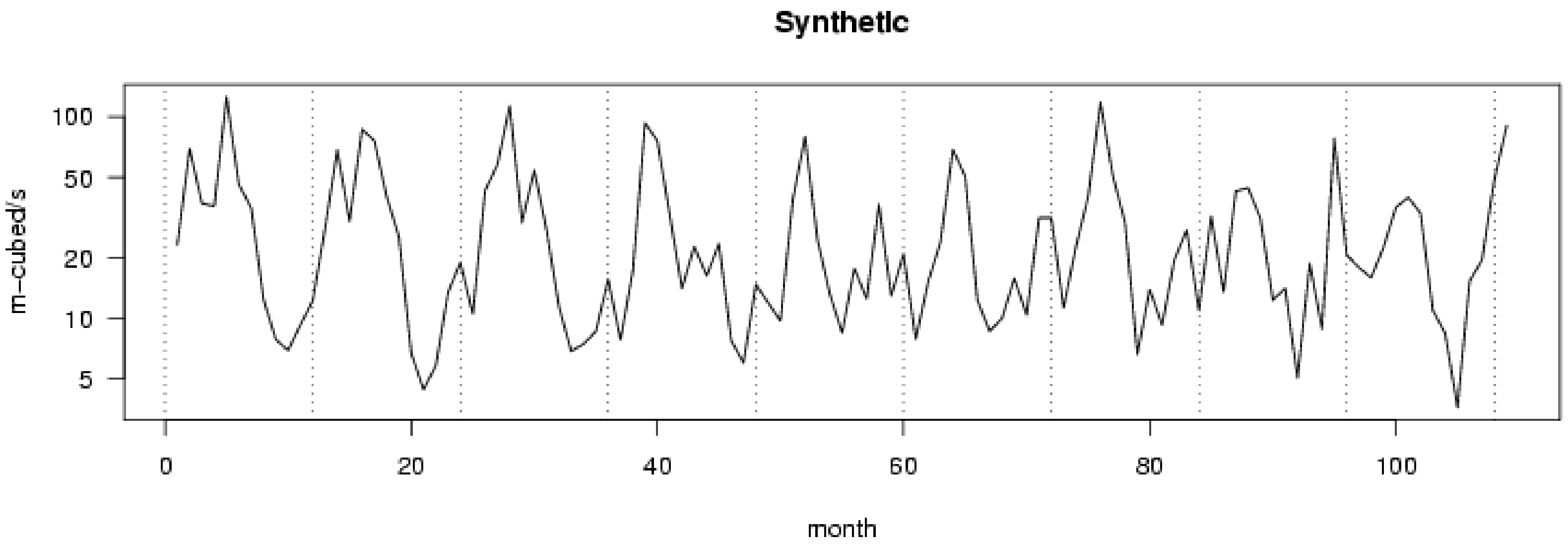
AR(1):

$$X_{ij} = \rho_j X_{i,j-1} + \varepsilon_{ij}$$

Nonstationary

Fit, generate synthetic series

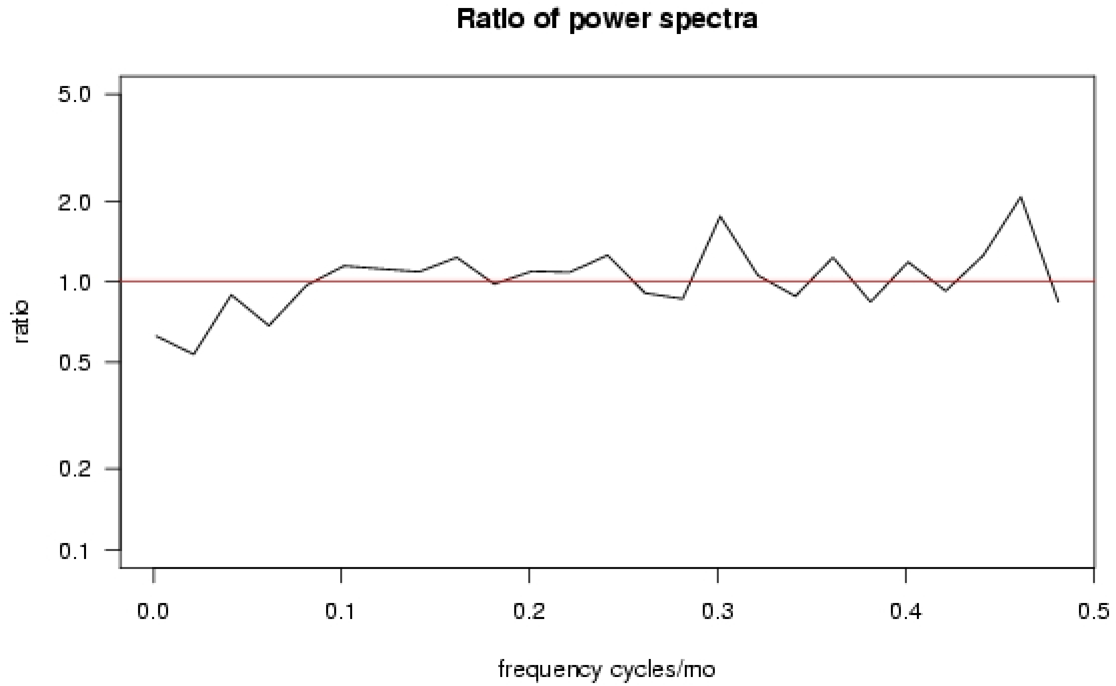

Saugeen riverflow
m-cubed/s
100
50
20
10
5
0
20
40
60
80
100
month

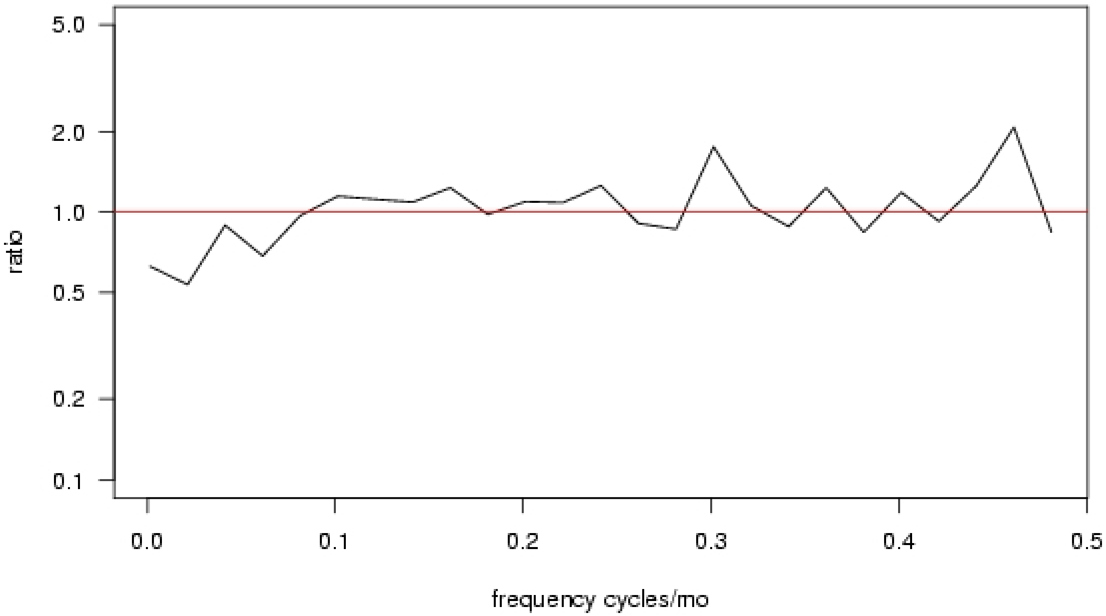

Synthetic
m-cubed/s
100
50
20
10
5
0
20
40
60
80
100
month

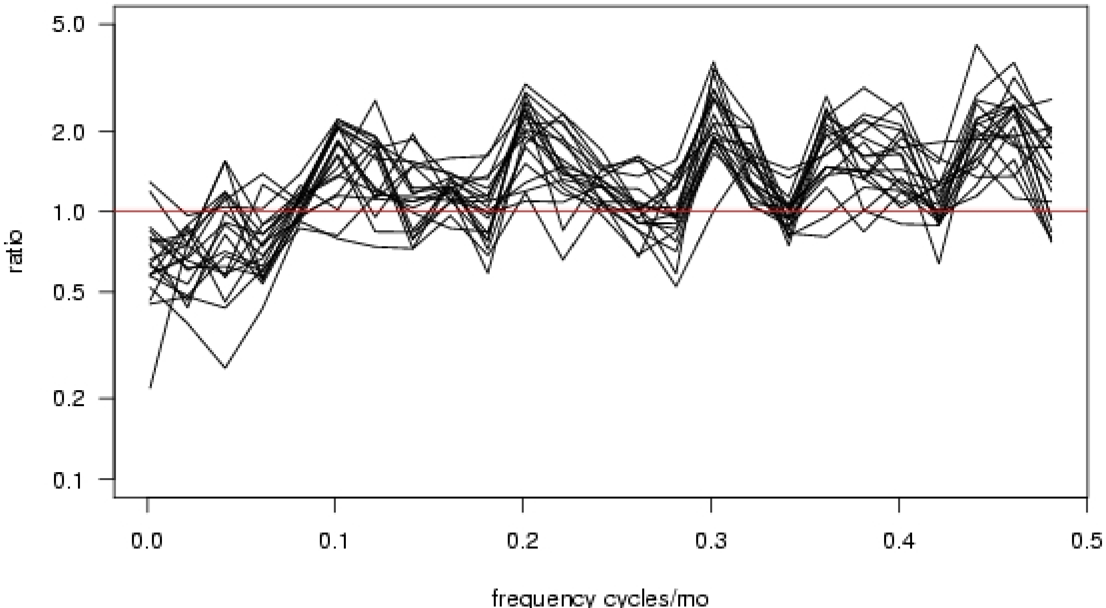
Turing test?

Comparison

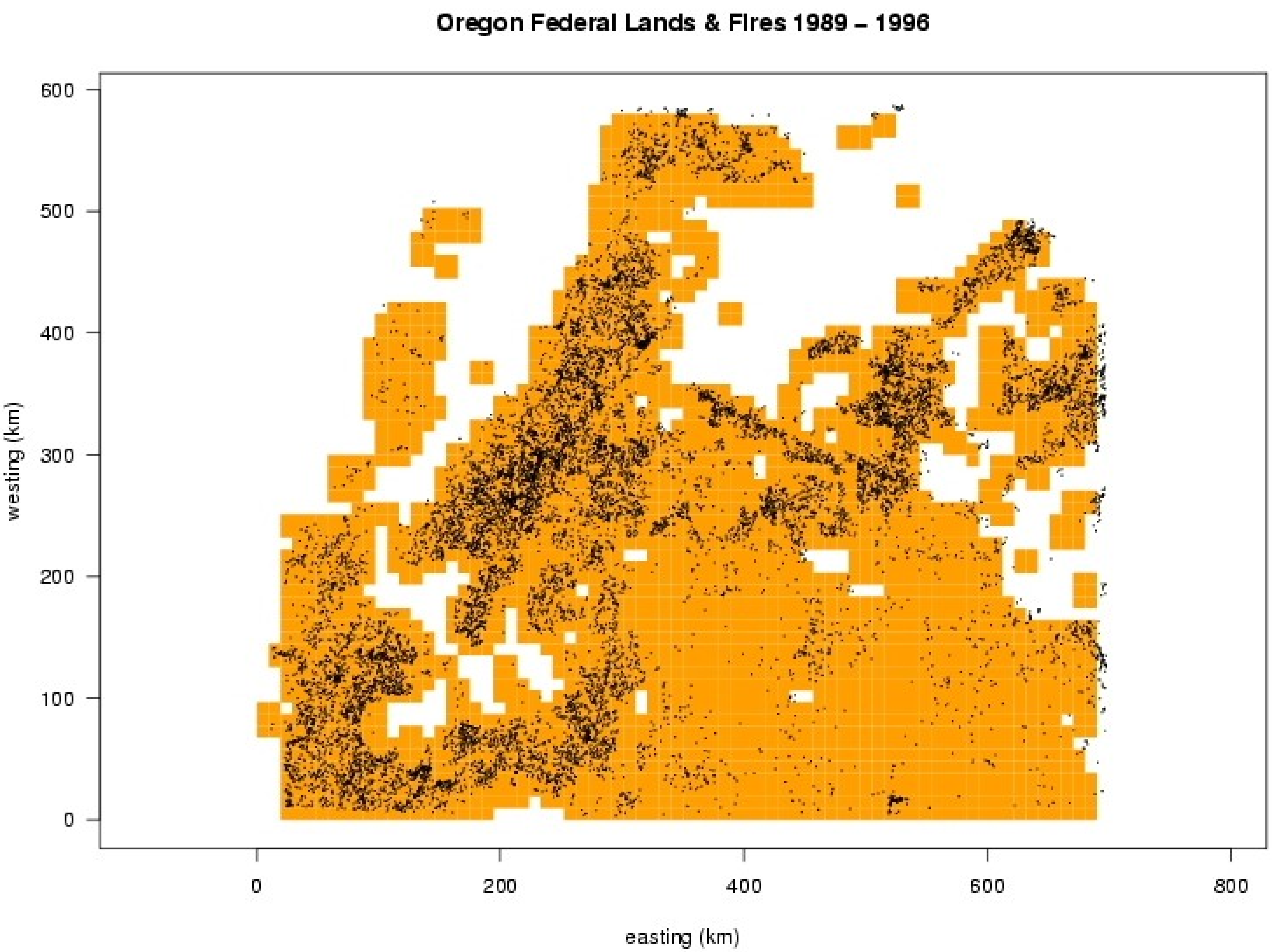
Spectral ratio

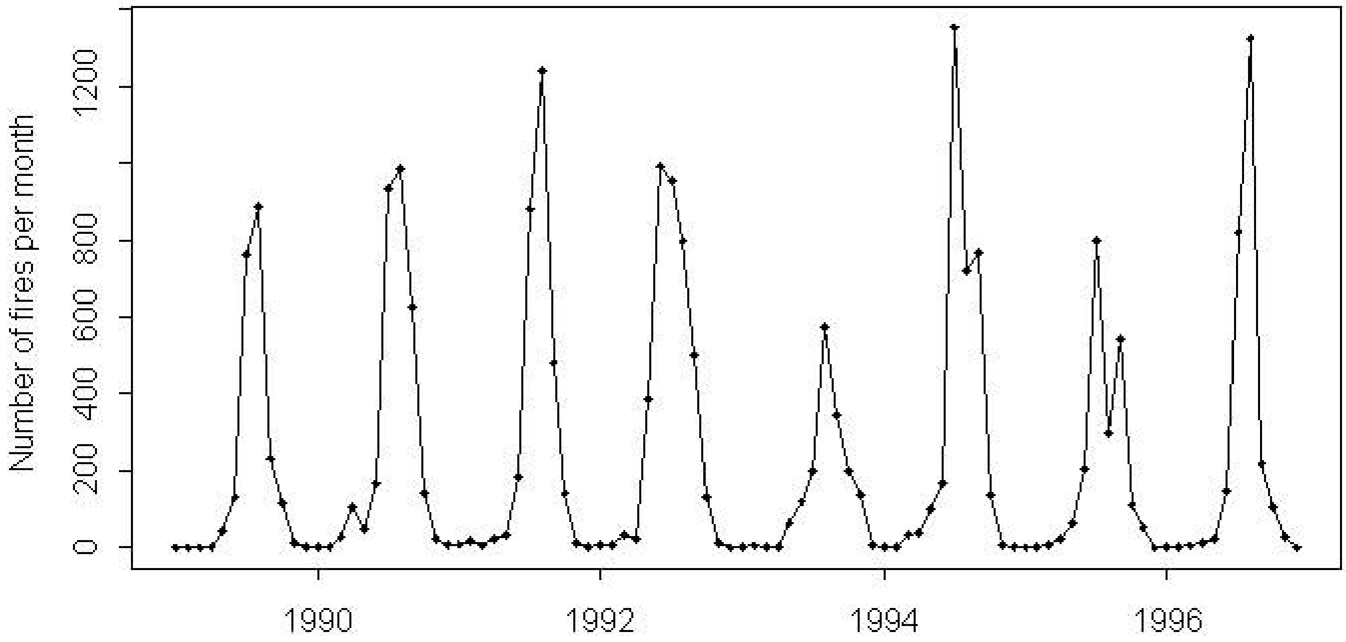

Ratio of power spectra
5.0
2.0
1.0
0.5
0.2
0.1
ratio
0.0
0.1
0.2
0.3
0.4
0.5
frequency cycles/mo


Ratio of power spectra


Ratio of power spectra

V. SPATIAL-TEMPORAL P. P.- *wildfires*




Number of fires per month
0
200
400
600
800
1200
1990
1992
1994
1996

Risk analysis

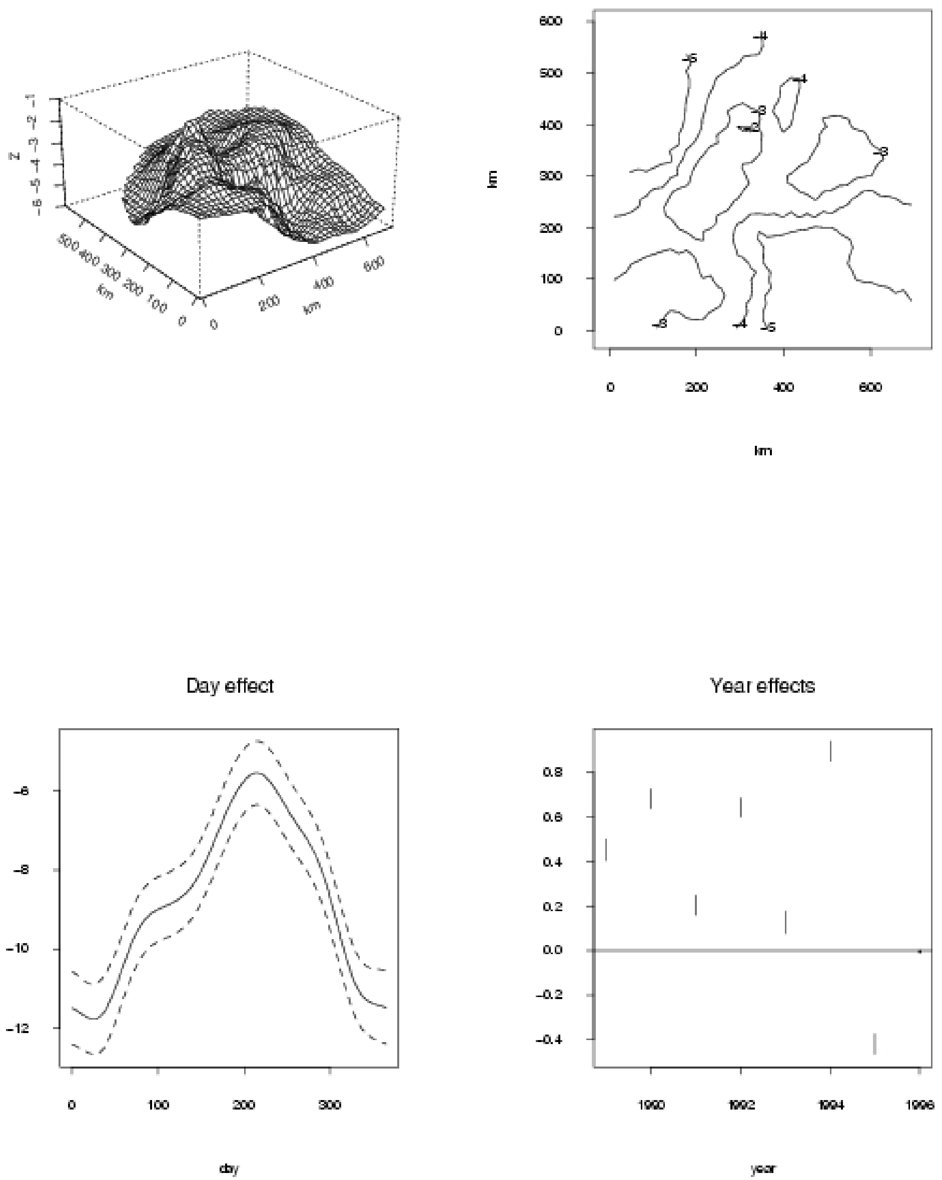
Pixel model, $\{(x_j, y_j, t_j)\}$

$\text{logit } P\{N_{xyt} = 1\} = g_1(x,y) + g_2(<t>) + h_{[t]}$

(x,y): location, <t>: day, [t]: year,

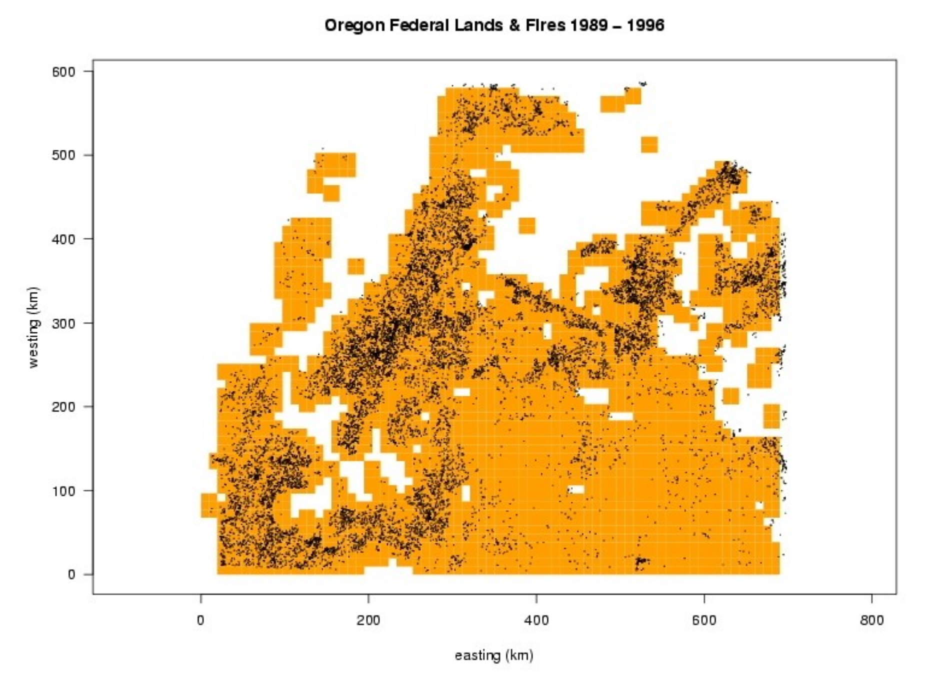
g_1, g_2 smooth

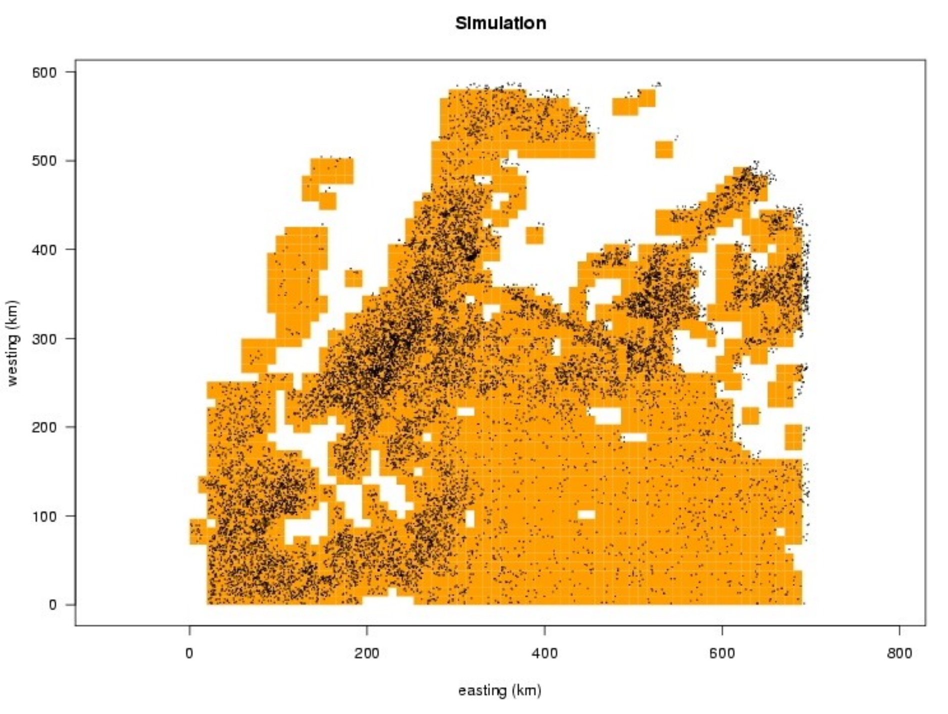
Sampled 0's



Z
-6 -5 -4 -3 -2 -1
500 400 300 200 100 0
km
0 200 400 600
km
600
500
400
300
200
100
0
km
-5
-4
-4
-3
-2
-3
-3
-4
-5
0 200 400 600
km
Day effect
-6
-8
-10
-12
0 100 200 300
day
Year effects
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
1990 1992 1994 1996
year

Original and Bernoulli simulation

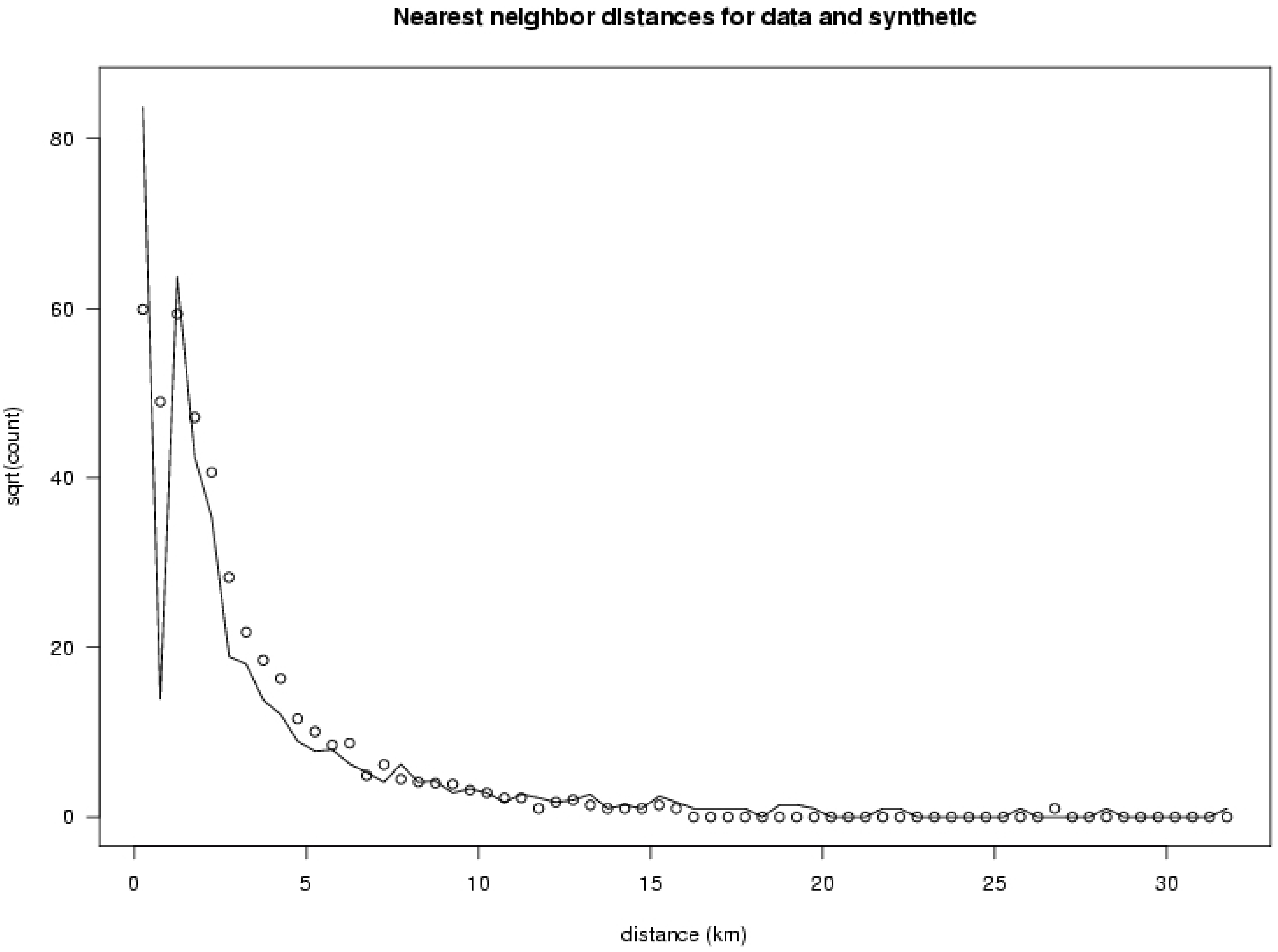

Oregon Federal Lands & Fires 1989 – 1996


Simulation

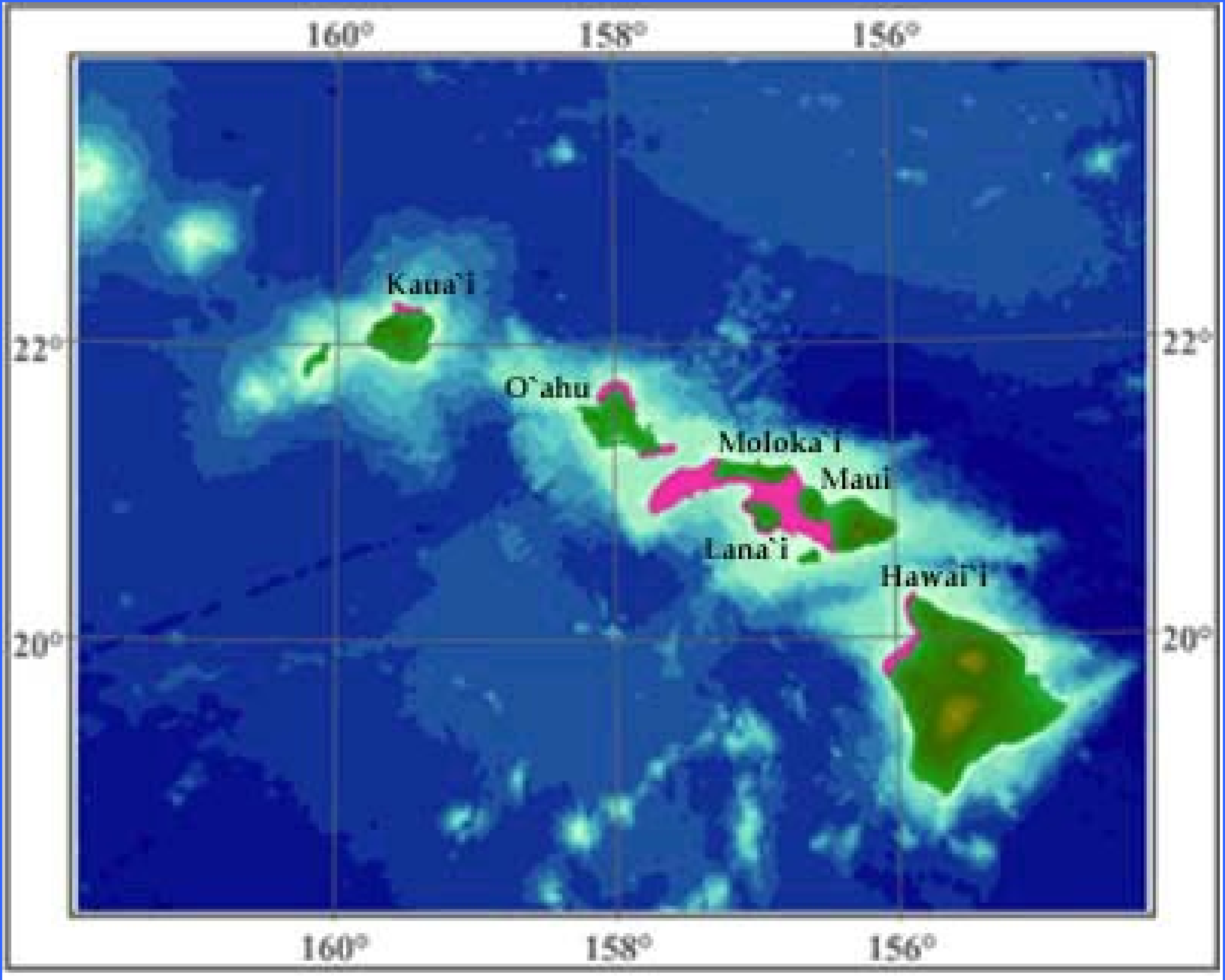
Turing test?

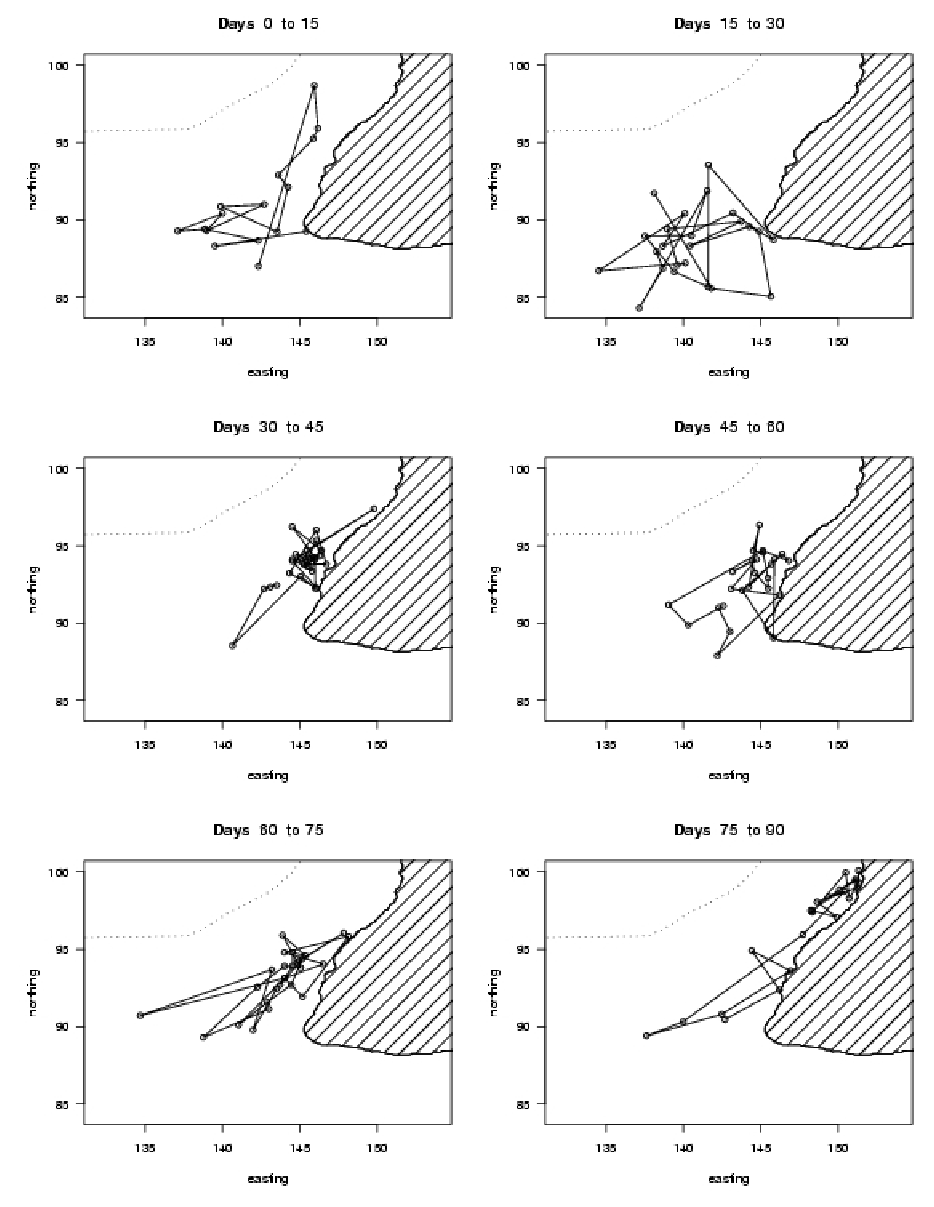
Comparisons

Nearest neighbour distribution


Nearest neighbor distances for data and synthetic
sqrt(count)
0
20
40
60
80
0
5
10
15
20
25
30
distance (km)

VI. TRAJECTORIES – *Hawaiian monk seal,* endangered


160°
158°
156°
22°
20°
Kaua`i
O`ahu
Moloka`i
Maui
Lana`i
Hawai`i



Foraging, resting, …

DEs. Newtonian motion

Scalar potential function, H

Planar case, location $\mathbf{r} = (x,y)'$, time t

$$d\mathbf{r}(t) = \mathbf{v}(t)\,dt$$

$$d\mathbf{v}(t) = -\beta\, \mathbf{v}(t)\,dt - \beta\, \text{grad}\, H(\mathbf{r}(t),t)\,dt$$

$\mathbf{v}$: velocity β: coefficient of friction

$$d\mathbf{r} = -\,\text{grad}\, H(\mathbf{r},t)\,dt = \boldsymbol{\mu}(\mathbf{r},t)\,dt, \quad \beta \gg 0$$

Examples of H.

Point of attraction

$$H(\mathbf{r}) = .5*\sigma^2 \log r - \delta r$$

Point of repulsion

$$H(\mathbf{r}) = C/r$$

Attraction/repulsion

$$H(\mathbf{r}) = \alpha(1/r^{12} - 1/r^6)$$

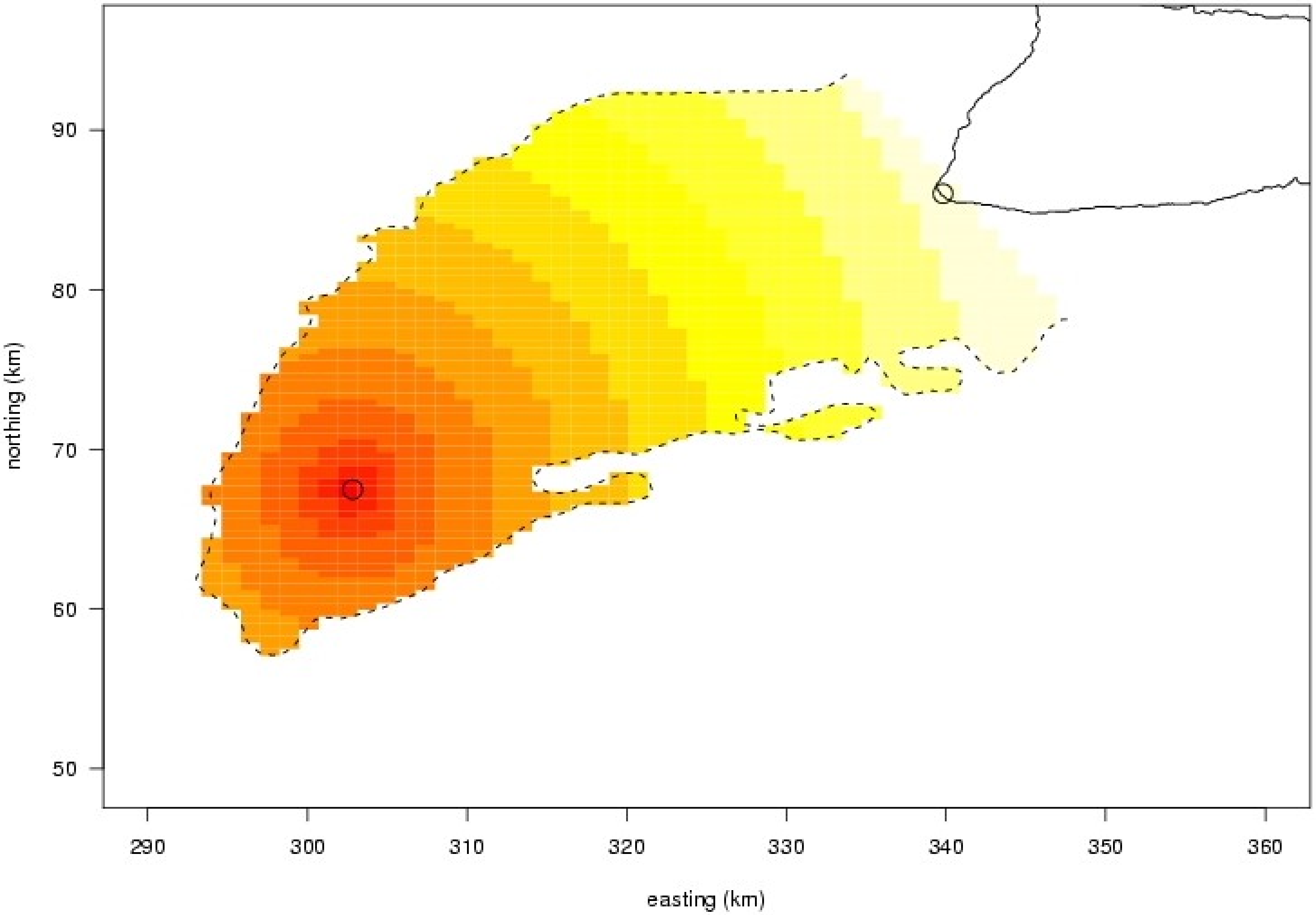
General parametric

$$H(\mathbf{r}) = \beta_{10}x + \beta_{01}y + \beta_{20}x^2 + \beta_{11}xy + \beta_{02}y^2$$

Nonparametric

spline expansion

Potential function for outbound journey



SDEs.

$$d\mathbf{r}(t) = \boldsymbol{\mu}(\mathbf{r}(t),t)dt + \boldsymbol{\sigma}(\mathbf{r}(t),t)d\mathbf{B}(t)$$

$\boldsymbol{\mu}$: drift, $-\mathrm{grad}\, H$

$\boldsymbol{\sigma}$: diffusion

$\{\mathbf{B}(t)\}$: bivariate Brownian

Data: $\{((x(t_j),y(t_j)),t_j)\}$

Solution/approximation

$$(\mathbf{r}(t_{i+1})-\mathbf{r}(t_i))/(t_{i+1}-t_i) =$$

$$\boldsymbol{\mu}(\mathbf{r}(t_i),t_i) + \boldsymbol{\sigma}(\mathbf{r}(t_i),t_i)\ \mathbf{Z}_{\mathbf{i+1}}/\sqrt{(t_{i+1}-t_i)}$$

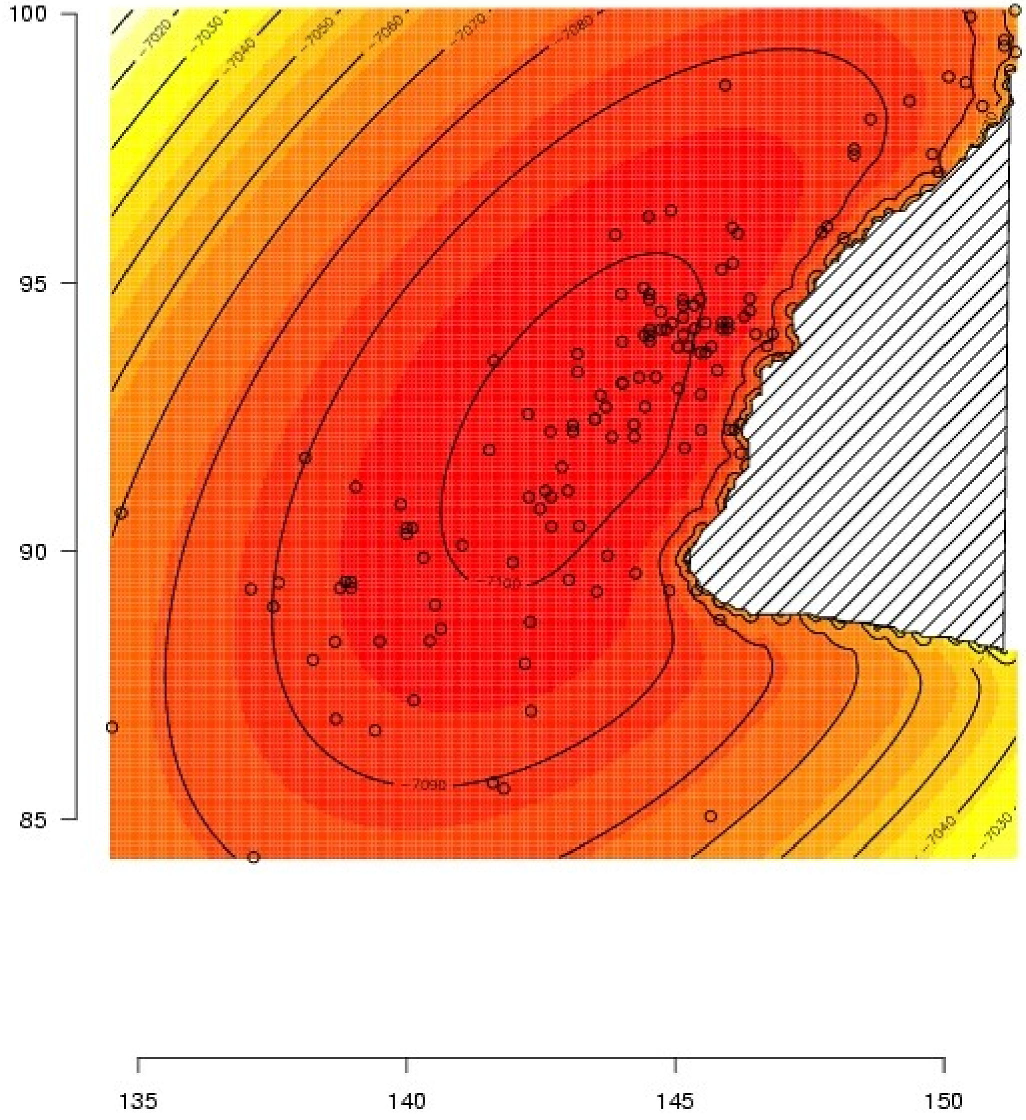
Euler scheme

Approximate likelihood

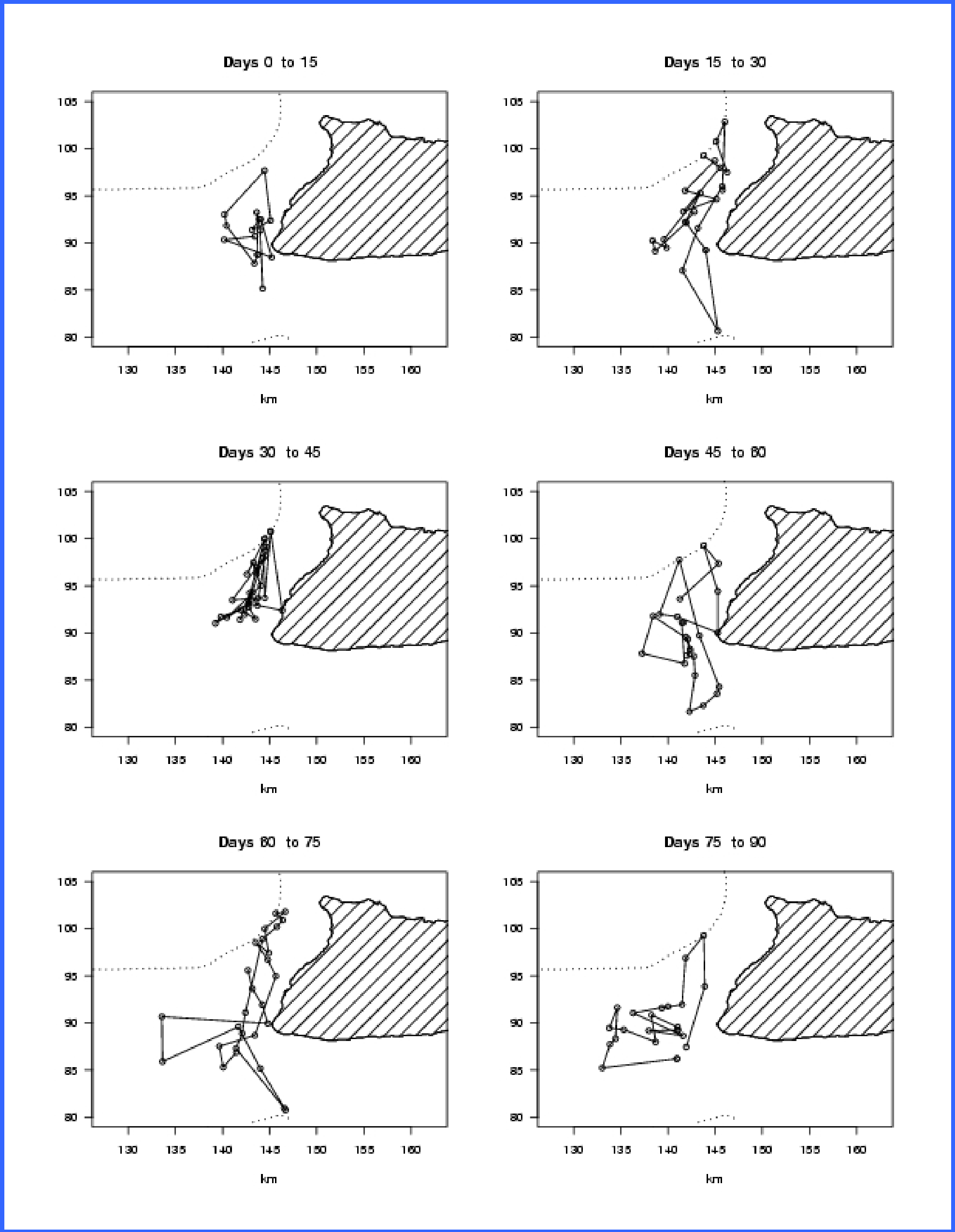
Boundary, startup effects

Fitted potential:

general parametric
attraction &
repulsion



Synthetic



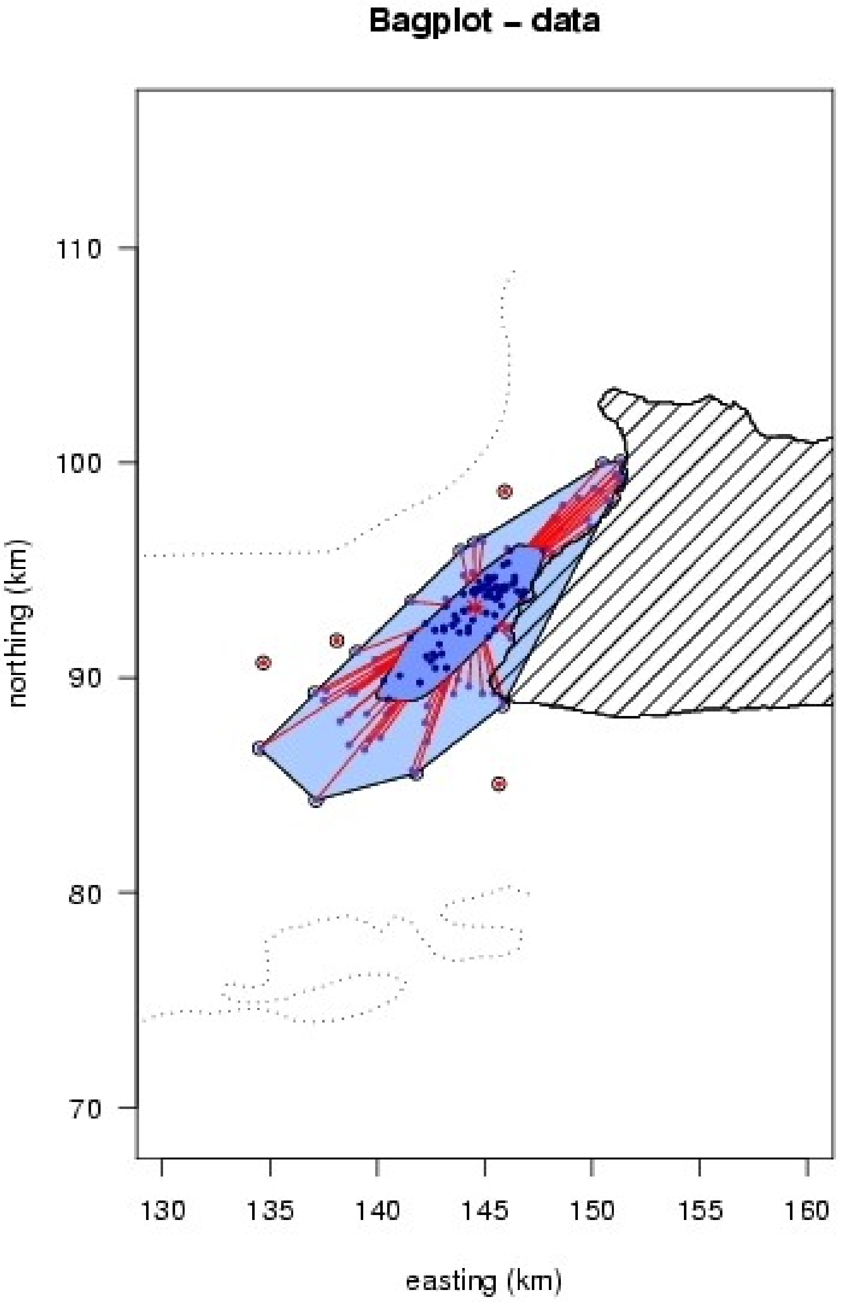
Turing test?

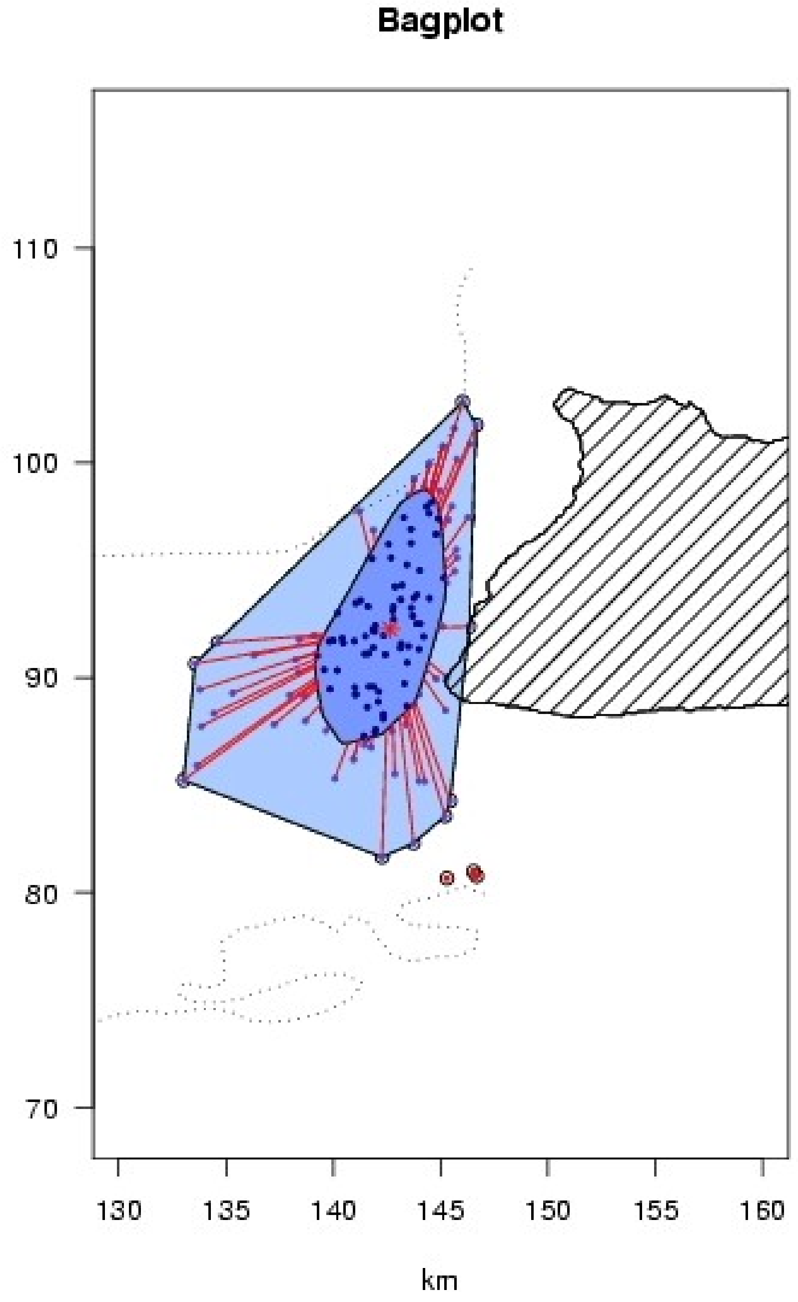
Comparison

Bagplot

- cp. Boxplot
- centre is bivariate median
- “bag” contains 50% with greatest depth
- fence – inflate bag by 3

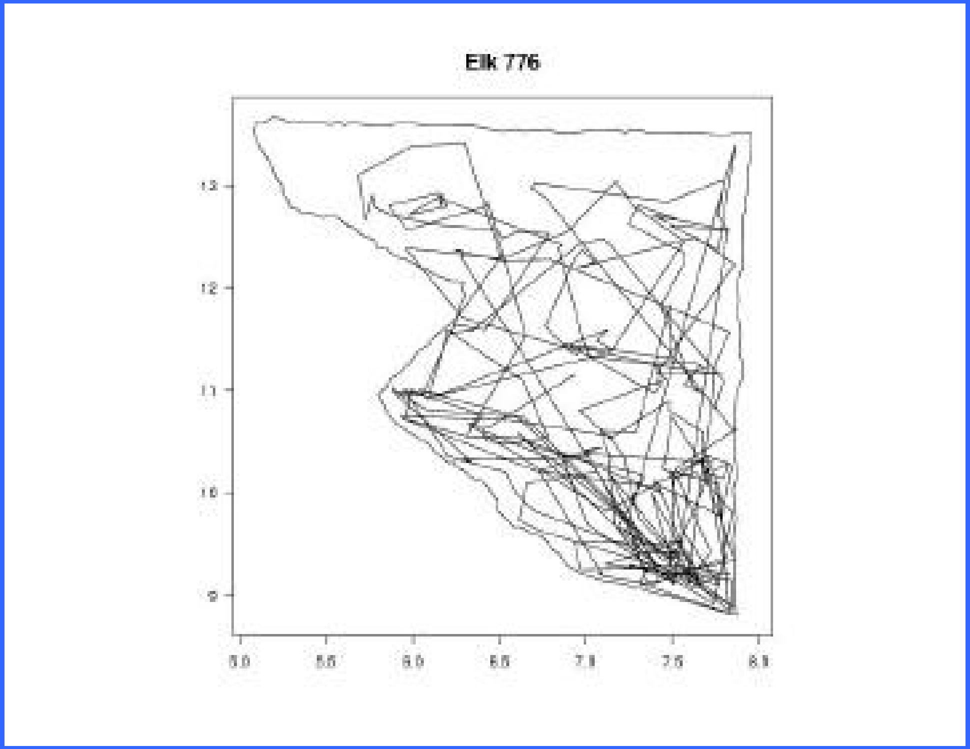
Rousseuw, Rutts, Tukey (1999)


Bagplot - data
110
100
90
80
70
northing (km)
130
135
140
145
150
155
160
easting (km)


Bagplot
110
100
90
80
70
130
135
140
145
150
155
160
km

TRAJECTORIES - elk/wapiti

Rocky Mountain elk (*Cervus elaphus*) Banff
Starkey Reserve, Oregon
Joint usage possible?



Data: $\{(x(t_j), y(t_j)), t_j)\}$

8 animals, Δt = 2hr

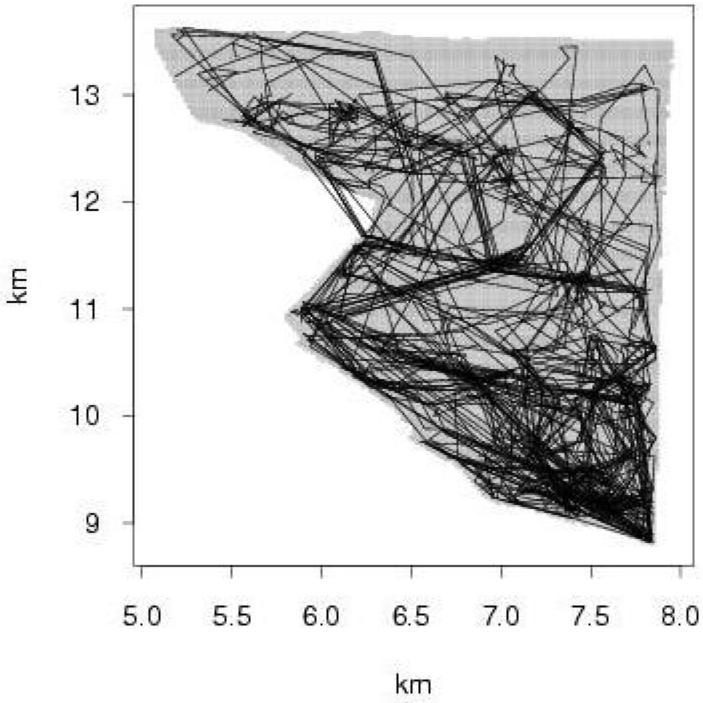
Foraging, resting, hiding, …

Model.

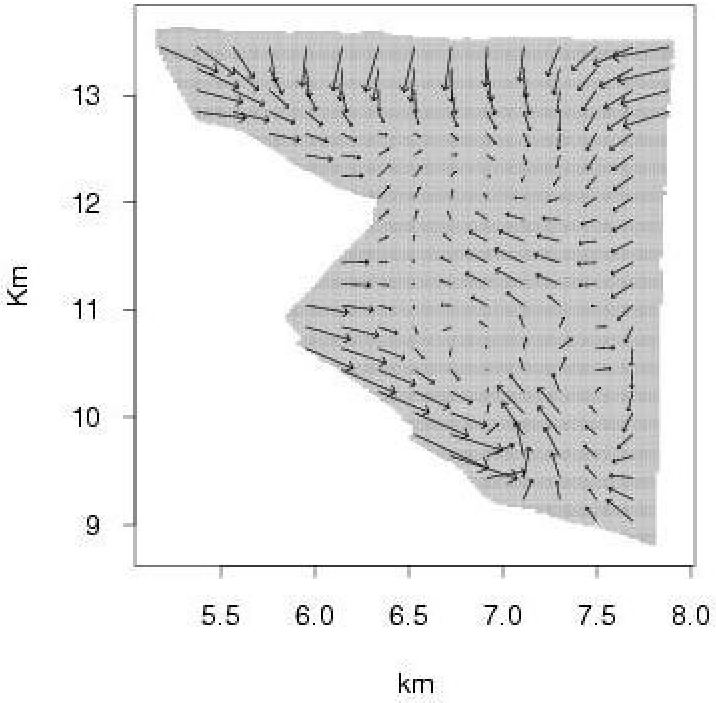
$$d\mathbf{r} = \boldsymbol{\mu}(\mathbf{r})dt + \boldsymbol{\sigma}d\mathbf{B}(t)$$

$\boldsymbol{\mu}$ smooth - geography

velocity field



Elk on C-control days



Velocity field

Boundary (NZ fence)

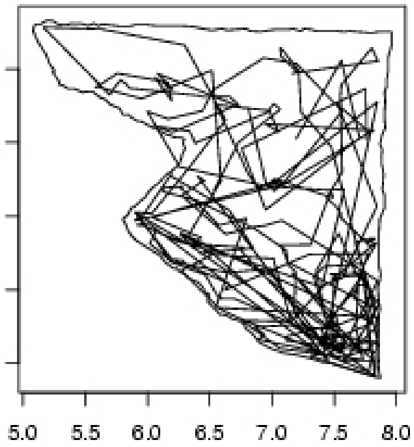
$$d\mathbf{r}= \boldsymbol{\mu}(\mathbf{r})dt + \boldsymbol{\sigma}(\mathbf{r})d\mathbf{B}(t) + d\mathbf{A}(\mathbf{r})$$

A**,** support on boundary**,** keeps particle constrained

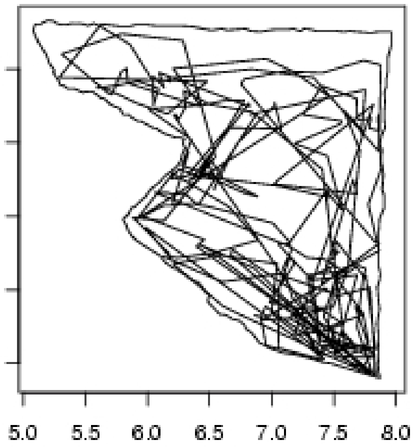
Synthetic paths.

If generated point outside**,** keep pulling bac
by half til inside

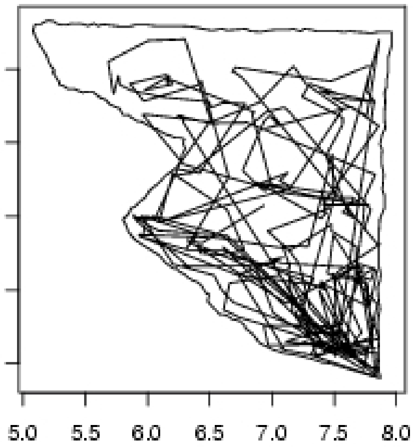
Elk 770



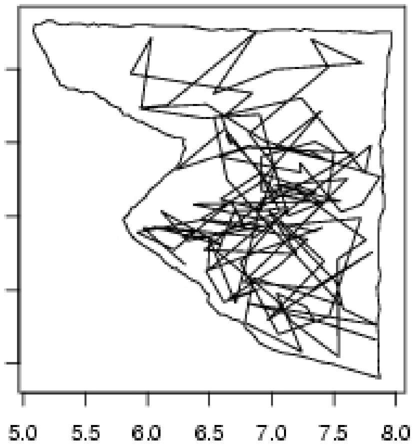
Elk 771



Elk 776



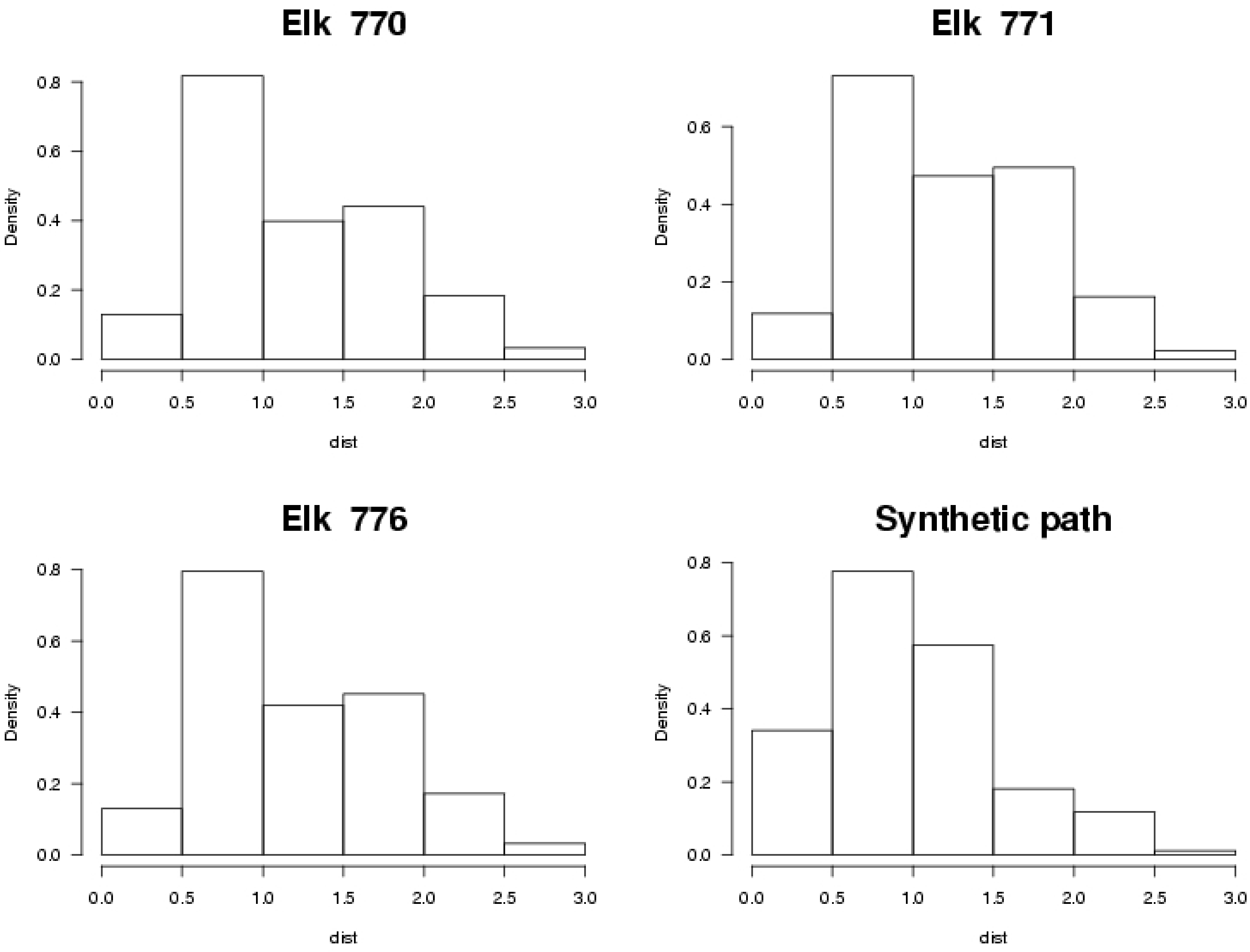
Synthetic path



Turing test?

Comparisons

Distribution of distances to “centre”


Elk 770
Density
0.0
0.2
0.4
0.6
0.8
0.0
0.5
1.0
1.5
2.0
2.5
3.0
dist
Elk 771
Density
0.0
0.2
0.4
0.6
0.0
0.5
1.0
1.5
2.0
2.5
3.0
dist
Elk 776
Density
0.0
0.2
0.4
0.6
0.8
0.0
0.5
1.0
1.5
2.0
2.5
3.0
dist
Synthetic path
Density
0.0
0.2
0.4
0.6
0.8
0.0
0.5
1.0
1.5
2.0
2.5
3.0
dist

VII. SUMMARY & DISCUSSION

Synthetic plots: method for appraising complex data-based models via Monte Carlo

Criteria: EDA, formal

Four examples: time series, spatial-temporal p.p., trajectories

Found inadequacies in each case

Corrections like Pearson's for chi-squared

Difficulties:

- land mask
- irregular times
- large time differences
- simulations based on same data

Acknowledgements.

Aager, Littman, Preisler, Stewart

NSF, FS/USDA

REFERENCES

E. Nelson (1967). *Dynamical Theories of Brownian Motion*. Princeton

H. S. Niwa (1996). Newtonian dynamical approach to fish schooling. *J. theor. Biol.*

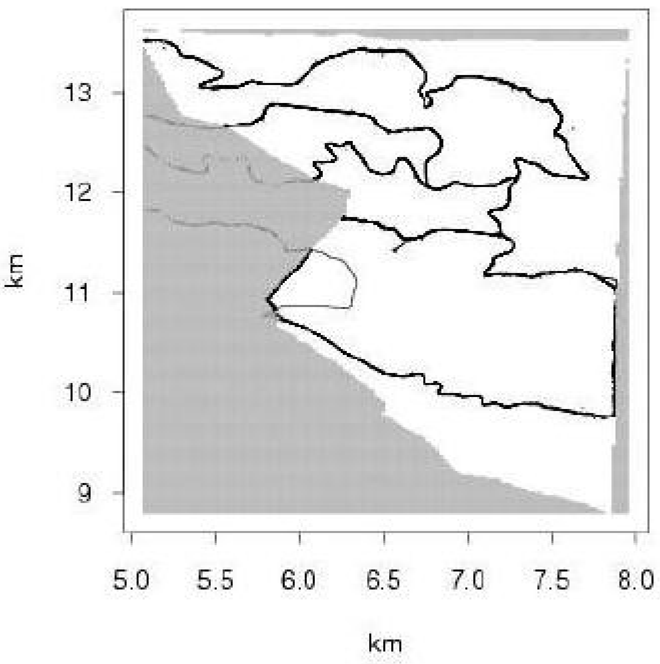
Part B.

Experiment with explanatory

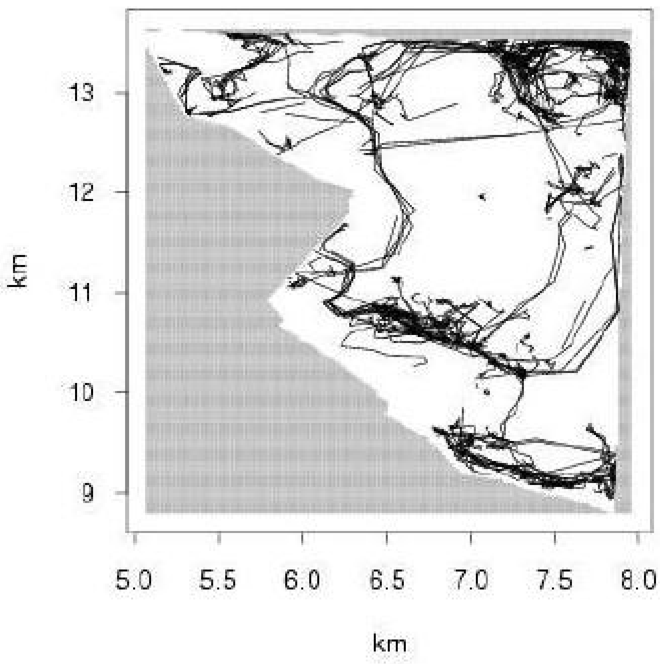
Same 8 animals

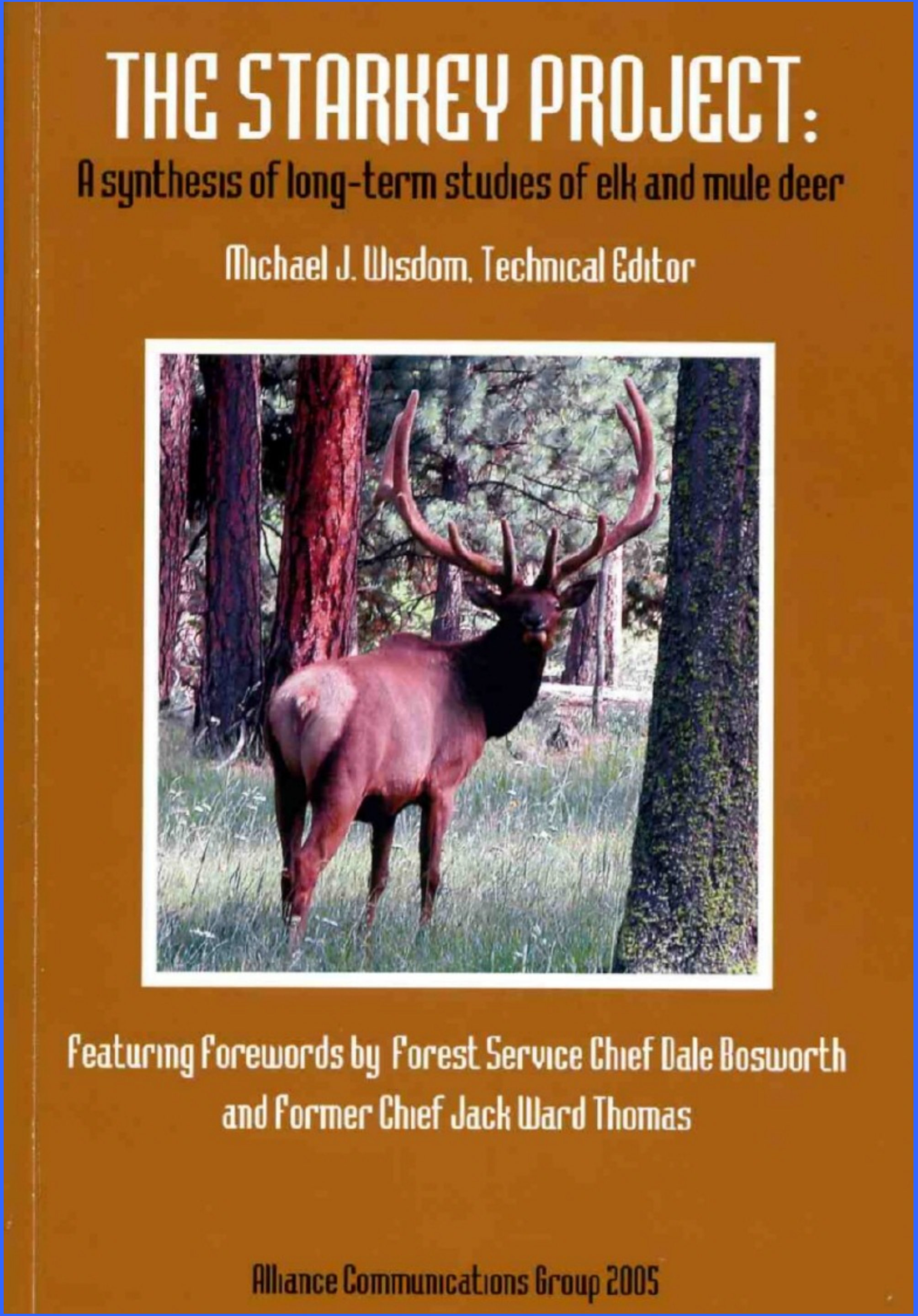
ATV days, Δt = 5min

ATV paths



Elk on ATV days





THE STARKEY PROJECT:

A synthesis of long-term studies of elk and mule deer

Michael J. Wisdom, Technical Editor

Featuring forewords by Forest Service Chief Dale Bosworth
and former Chief Jack Ward Thomas

Alliance Communications Group 2005

Next project

Whale shark feeds passively on small prey by swimming with its mouth open. A snorkeler watches from above.