1. Forecasting wildfires.

2. Forecasting earthquakes.

3. Forecasting global temperature.

(from Alexander & Cruz, 2013)

Table 1
Classification of surface fire spread models (1946–2000)

Reference	Type	Origin
Foens [4]	Theoretical	United States
Emmons [5]	Theoretical	United States
Hottel et al. [6]	Theoretical	United States
McArthur [7]	Empirical[a]	Australia
Van Wagner [8]	Theoretical	Canada
Thomas [9]	Theoretical	United Kingdom
McArthur [10]	Empirical[a]	Australia
Anderson [11]	Theoretical	United States
Frandsen [12]	Semiempirical	United States
Rothermel [13]	Semiempirical[a]	United States
Pagni and Peterson [14]	Theoretical	United States
Telisin [15]	Theoretical	Russia
Steward [16]	Theoretical	United States
Konev and Sukhinin [17]	Theoretical	Russia
Cekirge [18]	Theoretical	United States
Fujii et al. [19]	Theoretical	Japan
Grishin et al. [20]	Theoretical	Russia
Griffin and Allan [21]	Semiempirical	Australia
Huang and Xie [22]	Theoretical	United States
Sneeuwjagt and Peet [23]	Semiempirical	Australia
Albini [24,25]	Theoretical	United States
De Mestre et al. [26]	Theoretical	Australia
Weber [27]	Theoretical	Australia
Borrows et al. [28]	Semiempirical	Australia
Forestry Canada Fire Danger Group [29]	Empirical[a]	Canada
Croba et al. [30]	Theoretical	Greece
Marsden-Smedley and Catchpole [31]	Semiempirical	Australia
Grishin [32]	Theoretical	Russia
Dupuy [33]	Theoretical	France
Santoni and Balbi [34]	Theoretical	France
Linn [35]	Theoretical	United States
Catchpole et al. [36]	Semiempirical	Australia
Catchpole et al. [37]	Semiempirical	Australia
Fernandes [38]	Semiempirical	Portugal
Vega [39]	Semiempirical	Spain
McCaw [40]	Semiempirical	Australia
Viegas et al. [41]	Empirical	Portugal
Cheney et al. [42]	Empirical	Australia
Larini et al. [43]	Theoretical	France
Margerit and Guillaume [44]	Theoretical	France
Burrows [45,46]	Semiempirical	Australia
Hargrove et al. [47]	Empirical[a]	United States

[a] Models that constitute the basis of operating tools actually used in forestry agencies.

(from Pastor et al., 2003)

2) Currently used evaluation methods for spread models.

a) Is the problem of modeling spread solved?

Many simplifying assumptions in current models.

Simulations can do very well retrospectively,
on data used to make the sims.

b) Methods for model assessment

Evaluate a particular statistic, like surface temperature or spread rate.

Comparison of final burn areas.

Error matrices, Cohen's *K,* Sørensen's *Q*.

Most rate of fire spread models have the following kinds of limitations and should not be expected to predict what they do not pretend to represent (after Albini 1976a):

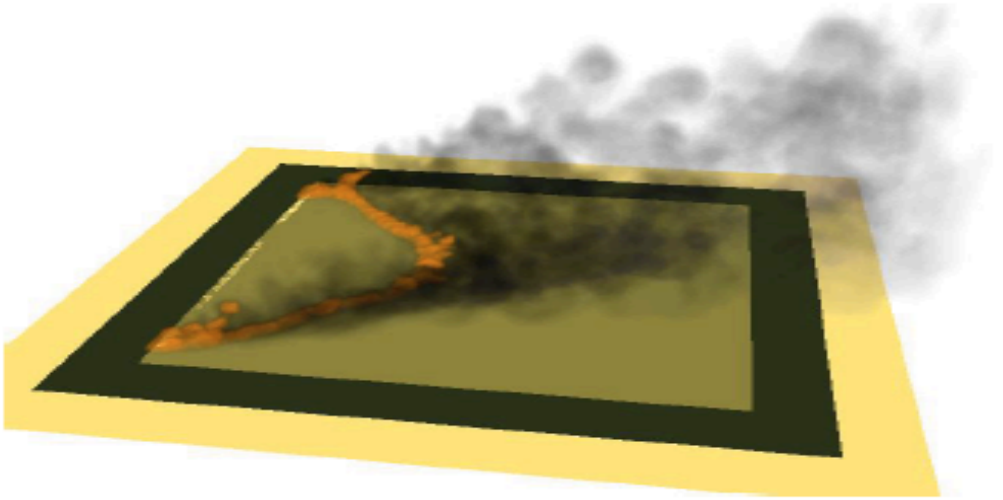
1. ***The fuel complex is assumed to be continuous, uniform, and homogeneous.*** The more the actual fuel situation departs from this idealized assumption, the more likely the prediction will not match the observed fire behaviour. While this issue is a matter of scale, subsequent research (e.g., Frandsen and Andrews 1979, Catchpole *et al.* 1989) and other innovations (Fujioka 1985, Finney 2003) such as the two-fuel model concept (Rothermel 1983, Martin 1988), as well as geographic information system (GIS)-based fire growth models (Beck 2000, Finney 2004, Tymstra *et al.* 2010) have not substantially reduced this problem. It thus remains a continuing research challenge (Parsons *et al.* 2011) and involves both the physical fuel characteristics as well as fuel moistures, including differences due to topographic features such as slope exposure (Cheney 1981).

2. ***Some models assume that the fuel bed is a single layer and is contiguous to the ground.*** In other words, there is no distinct gap between fuel layers (e.g., a forest stand with ground/surface fuels and crown or aerial fuels). As Van Wagner (1985) has so emphatically stated, "The fire world would beat a path to the door of the modeller who could account for vertical gradients and interruptions in moisture content and fuel density." Much progress has been made in

(Alexander & Cruz, 2013)

Mell et al. (2007), **retrospective.** The experiment shown here was used in model construction.

(a)



(b)

Figure 4: (a) Photograph of experimental fire F19 at $t = 56$ s. (b) Snapshot of WFDS simulation of experimental fire F19 at $t = 56$ s.

Observed vs. Predicted spread rate (Cruz & Alexander, 2013), **<u>retrospective:</u>**

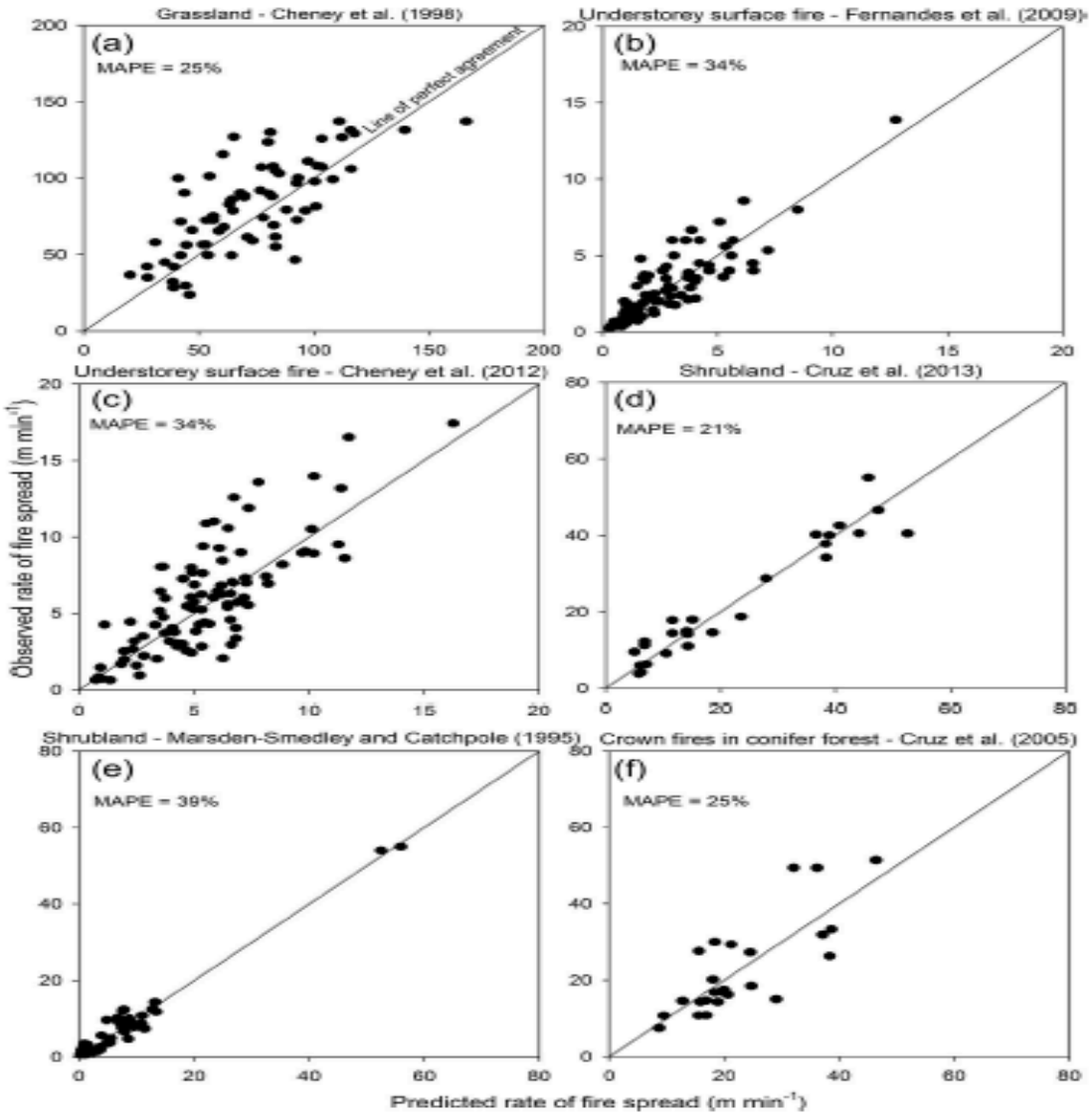


Fig. 3. Observed rates of spread used in model development versus model predictions for experimental surface fires in (a) grasslands fires (Cheney et al., 1998), (b) surface fire in maritime pine forest (Fernandes et al., 2009), (c) surface fire in dry eucalypt forest (Cheney et al., 2012), (d) surface and crown fires in mallee-heath shrublands (Cruz et al., 2013), (e) buttongrass moorlands (Marsden-Smedley and Catchpole, 1995), and (f) crown fires in conifer forests (Cruz et al., 2005). MAPE = mean absolute percent error.

Observed vs. Predicted spread rate (Cruz & Alexander, 2013), **prospective:**

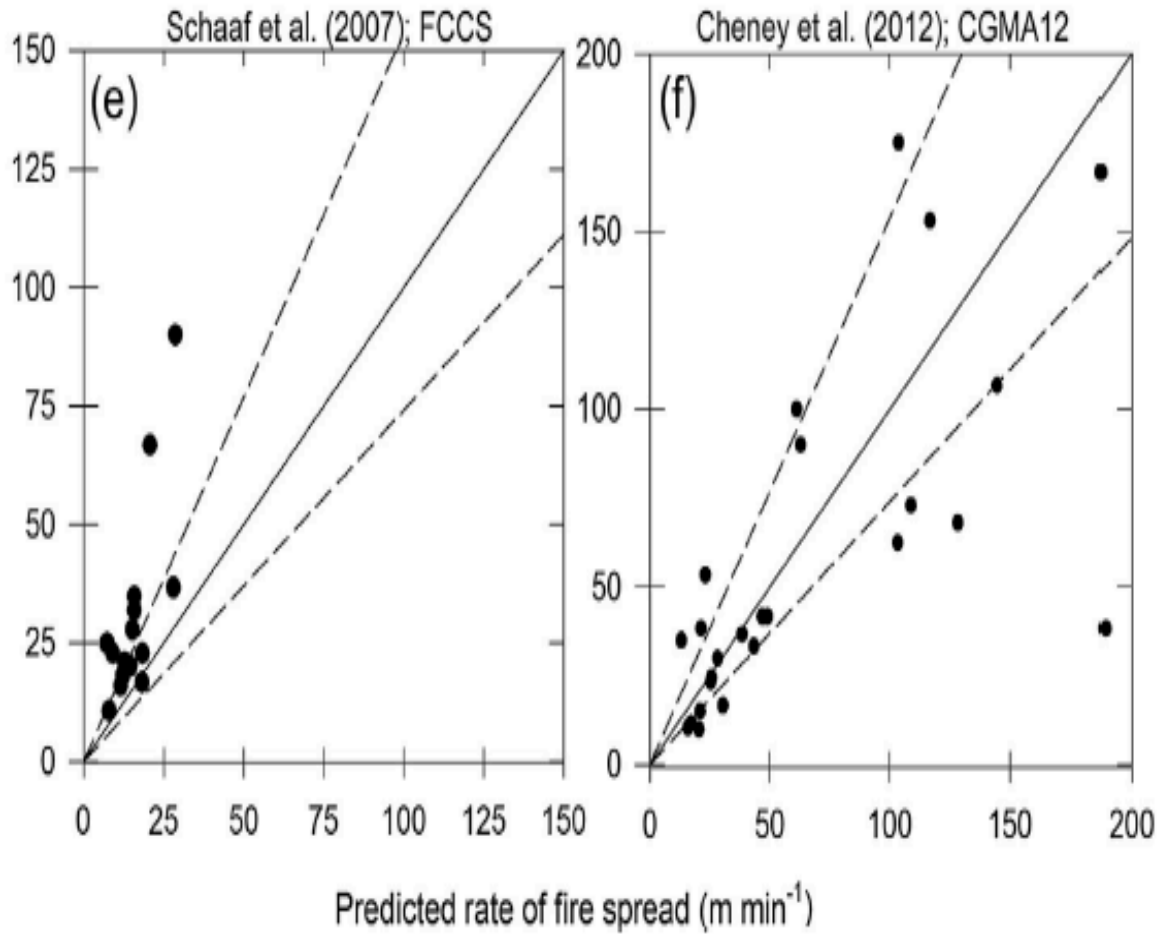


Fig. 6. Observed rates of spread versus model predictions for a selection of studies presented in Table 1 featuring both 'moderately slow' and 'exceedingly fast' spreading fires (i.e. upwards of ~150 m min^{-1}). The dashed lines around the line of perfect agreement indicate the ±35% error interval. Refer to Table 2 for the mean of the model abbreviations (e.g. RCR72 = Rothermel, 1972). MAPE = mean absolute percent error.

Observed vs. Predicted spread rate (Alexander & Cruz, 2013)

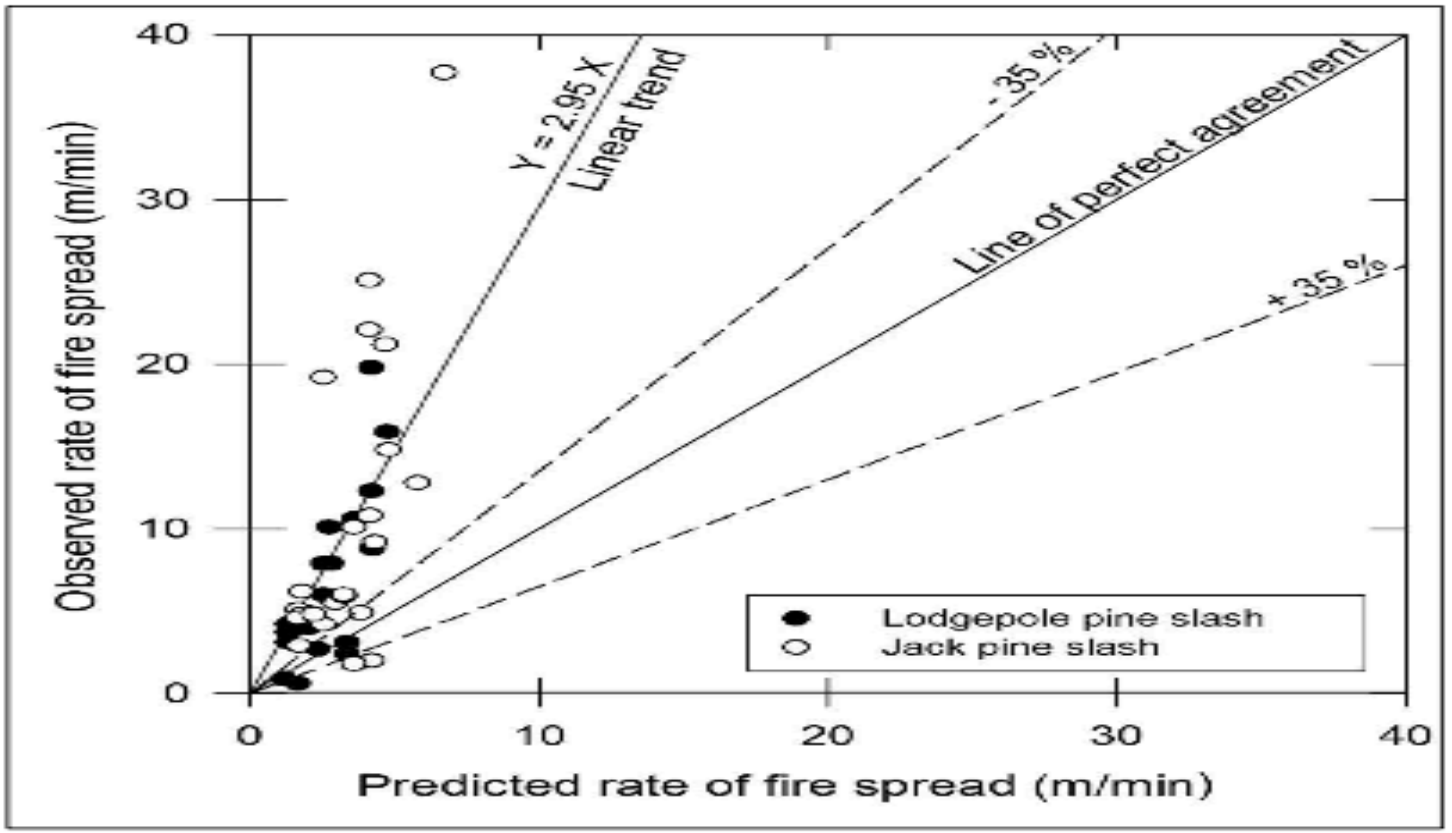


Fig. 3. Observed rates of spread for experimental fires in lodgepole pine logging slash in southwestern Alberta (Quintilio 1972) and jack pine logging slash in northeastern Ontario (Stocks and Walker 1972) versus predictions from Rothermel's (1972) surface fire rate of spread model for Fuel Model 12 – Medium Logging Slash (Anderson 1982) using a wind adjustment factor of 0.4 (Andrews 2012) (adapted from Cruz and Alexander 2013). The dashed lines around the line of perfect agreement indicate the ±35% error interval.

Observed and Predicted surface temperature (de Mestre et al., 1989, from Weber 1991)

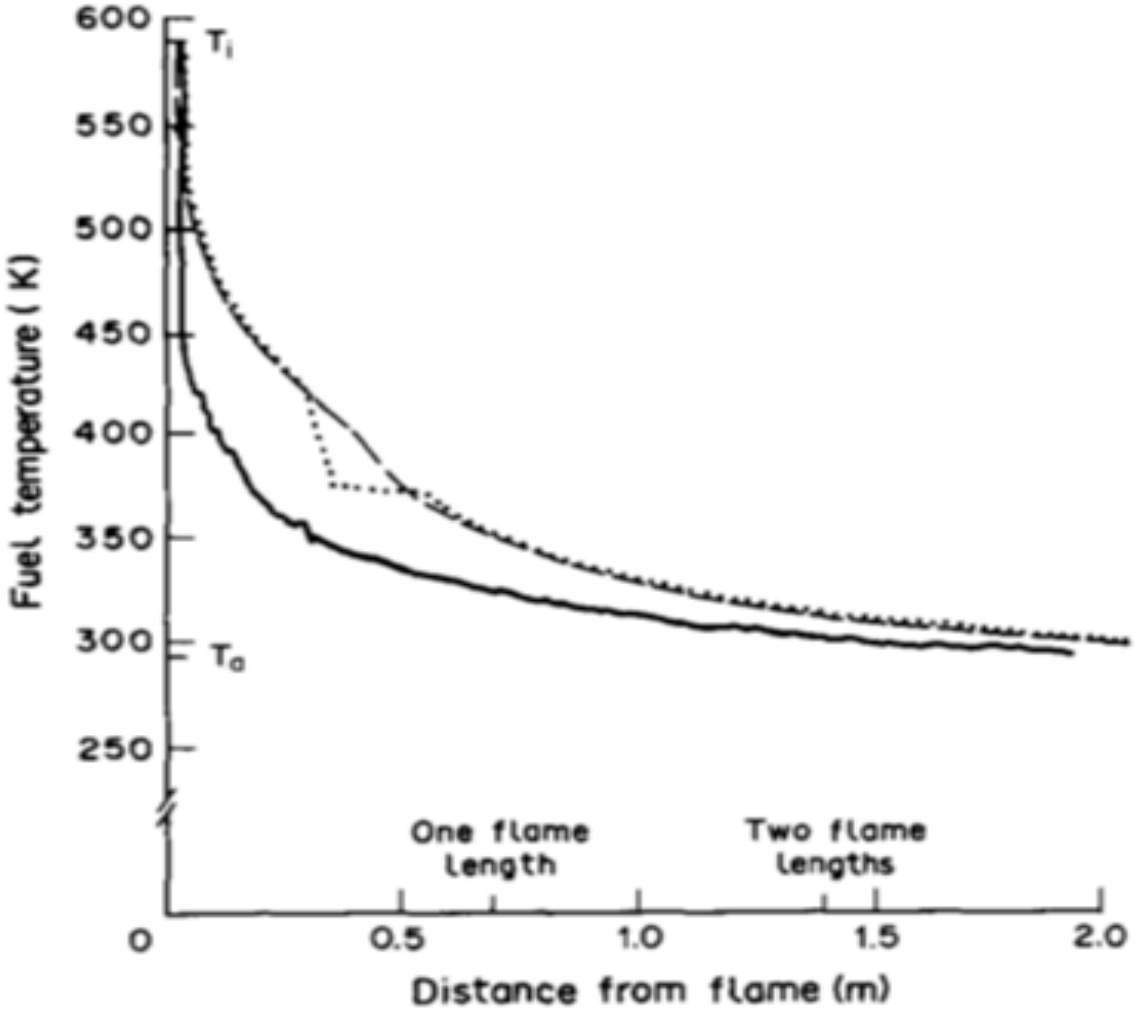


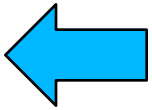
FIG. 2. Fuel surface temperature ahead of the fire front. Experimental (solid line), 'frozen' moisture model (dotted line), and averaged moisture model (dashed line) are shown. From de Mestre *et al.*[3]

Errors in windspeed measurements (Sullivan and Knight, 2001)

Table 2. Errors in estimating 5-min wind averages at a fire front as a percentage of the measured wind within 68% confidence limits (1 SD) assuming a standard deviation of 27% of the mean wind speed.

No. of anemometers	Fire width			
	0–40 m	≈80 m	≈160 m	≈300 m
1	±38%	±33%	±30%	±28%
2	±33%	±27%	±23%	±21%
4	±30%	±23%	±19%	±16%
8	±28%	±21%	±16%	±13%

One anemometer can predict the wind speed at another location (say at the front of a small fire) with a probability curve associated with the convolution of two 26.7% curves, i.e., from eq. 8, $s_f = 37.8\%$ of the mean wind speed. In other words there is a 68% chance of the wind speed at the fire front being within ±37.8% of the measured wind speed. This is not particularly good.



Observed and Simulated burn polygons (Arca et al., 2007)

Time step	Observed ROS	Simulated ROS
1	7.0	6.5
2	12.4	10.3
3	6.6	7.4
whole area	8.1	8.1

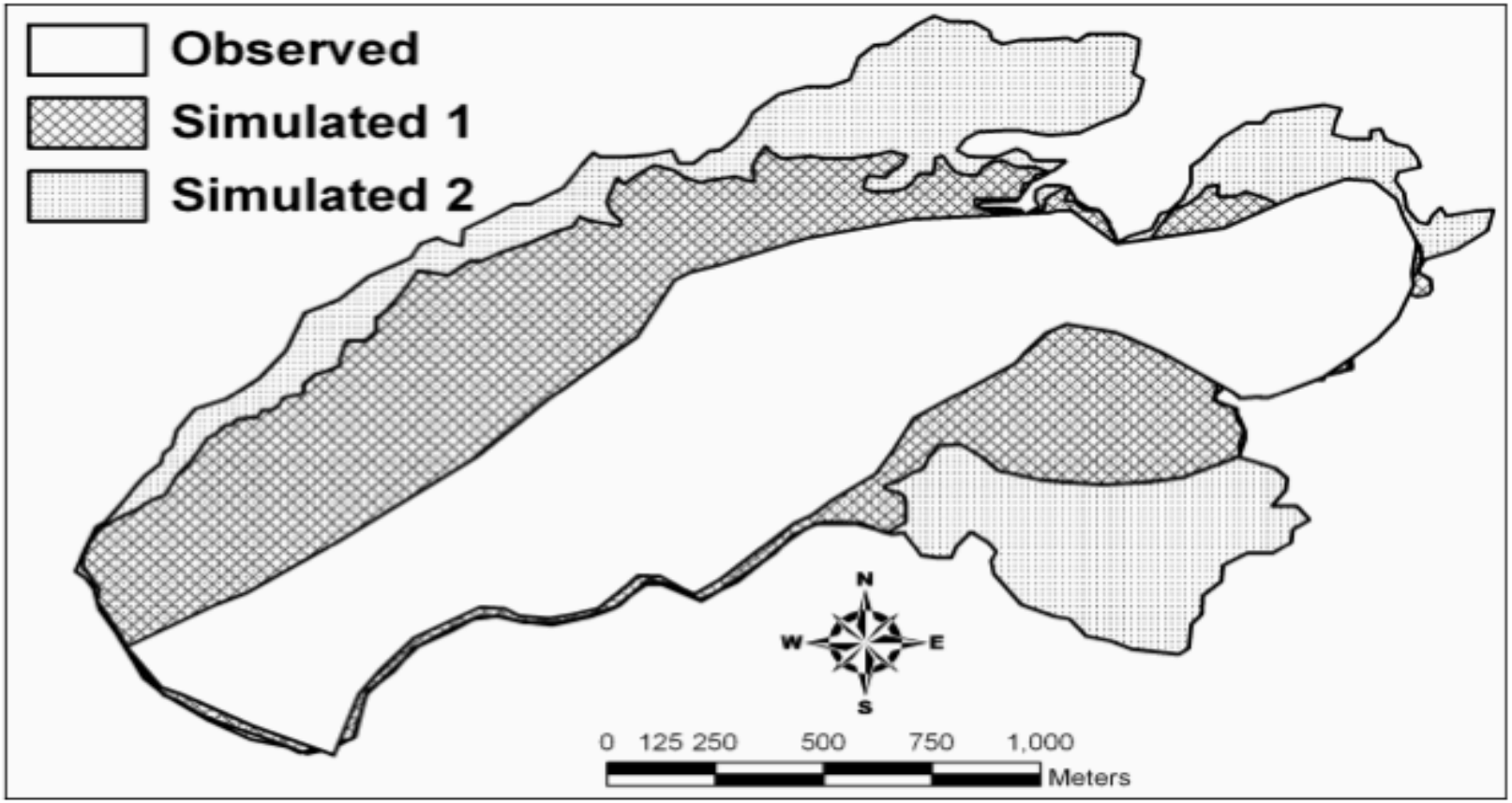


Figure 2 – Comparison between observed and simulated fire areas from the simulation n. 4 (custom fuel model CM28) using raster wind maps (Simulated 1) and constant wind field (Simulated 2).

Observed and Simulated burn polygons (Fujioka, 2002)

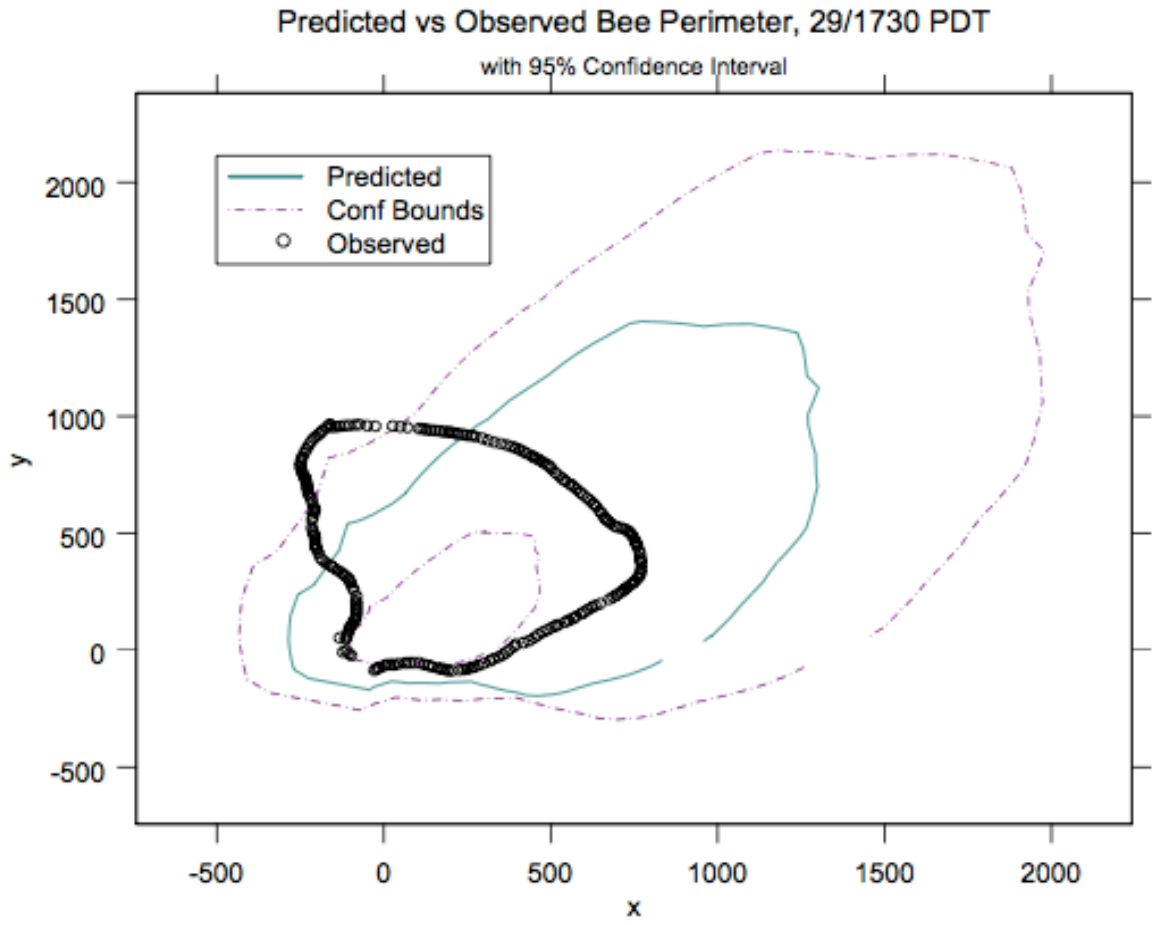


Figure 7. Position and error corrected perimeter prediction for the Bee Fire, 29 June 1996, 1730 PDT, and approximate 95% confidence interval for the true perimeter.

Remote Sensing techniques.

Error matrix (Congalton 1991)

Table 1. An Example Error Matrix

	Reference Data				
	D	C	BA	SB	row total
D	65	4	22	24	115
C	6	81	5	8	100
BA	0	11	85	19	115
SB	4	7	3	90	104
column total	75	103	115	141	434

Land Cover Categories

D = deciduous

C = conifer

BA = barren

SB = shrub

OVERALL ACCURACY = 321/434 = 74%

PRODUCER'S ACCURACY

D = 65 / 75 = 87%
C = 81 / 103 = 79%
BA = 85 / 115 = 74%
SB = 90 / 141 = 64%

USER'S ACCURACY

D = 65 / 115 = 57%
C = 81 / 100 = 81%
BA = 85 / 115 = 74%
SB = 90 / 104 = 87%

Cohen's *K* (Congalton 1991)

$$\hat{K} = \frac{N \sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} (x_{i+} * x_{+i})}{N^2 - \sum_{i=1}^{r} (x_{i+} * x_{+i})},$$

Sørensen's original formula was intended to be applied to presence/absence data, and is

$$QS = \frac{2C}{A+B} = \frac{2|A \cap B|}{|A| + |B|}$$

Histogram of over-prediction and under-prediction of areas (Cruz and Alexander, 2013).

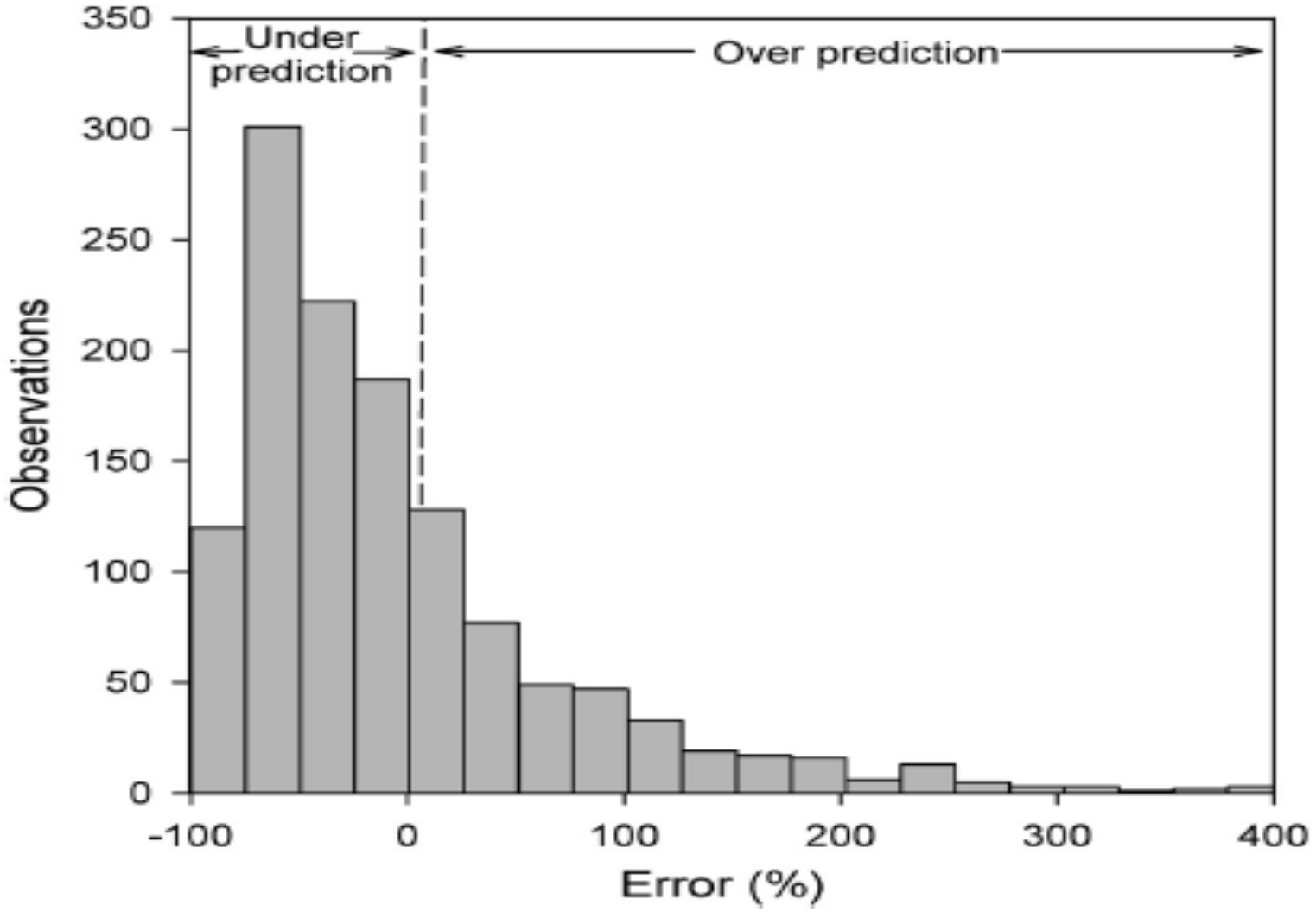
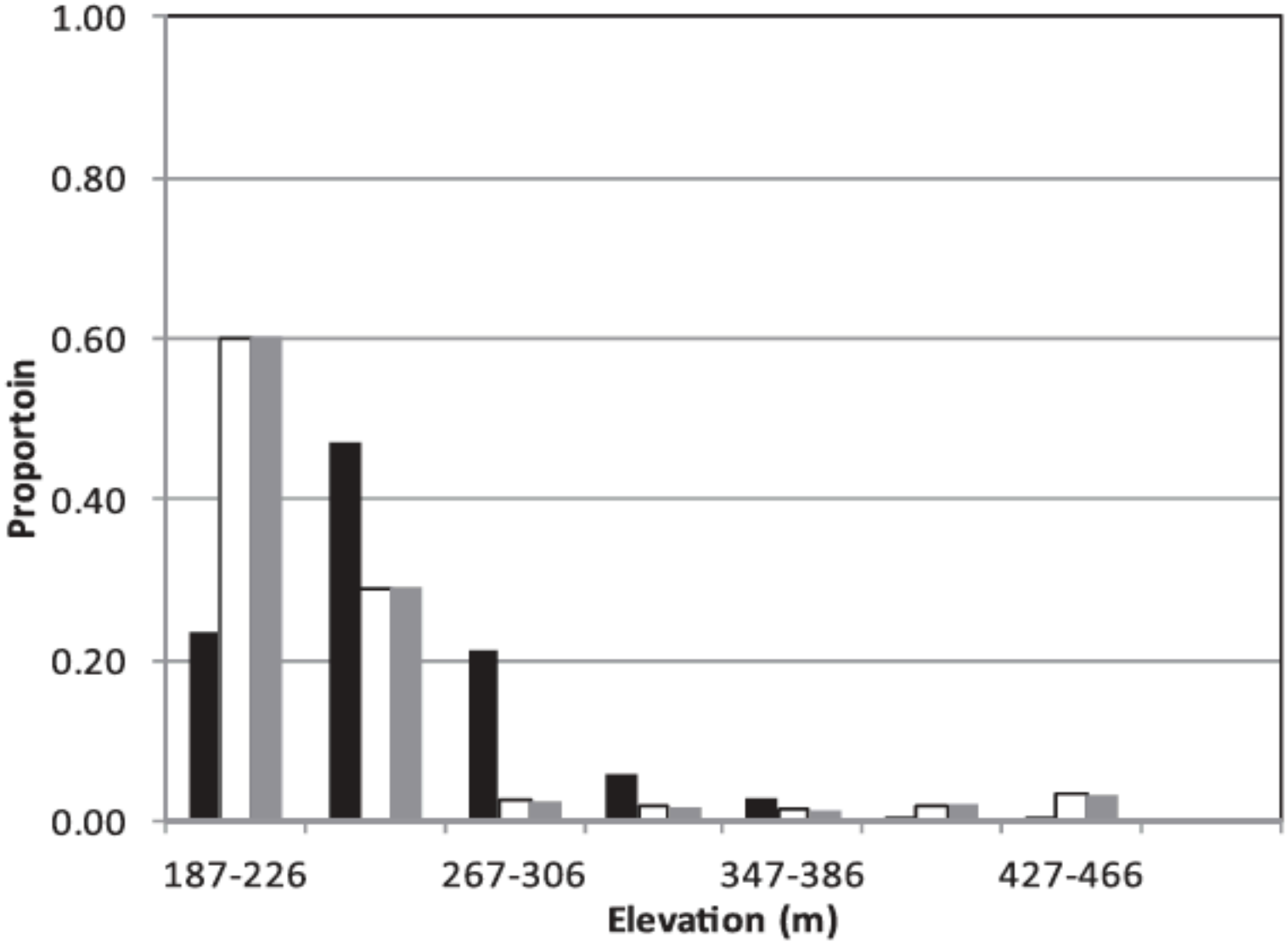


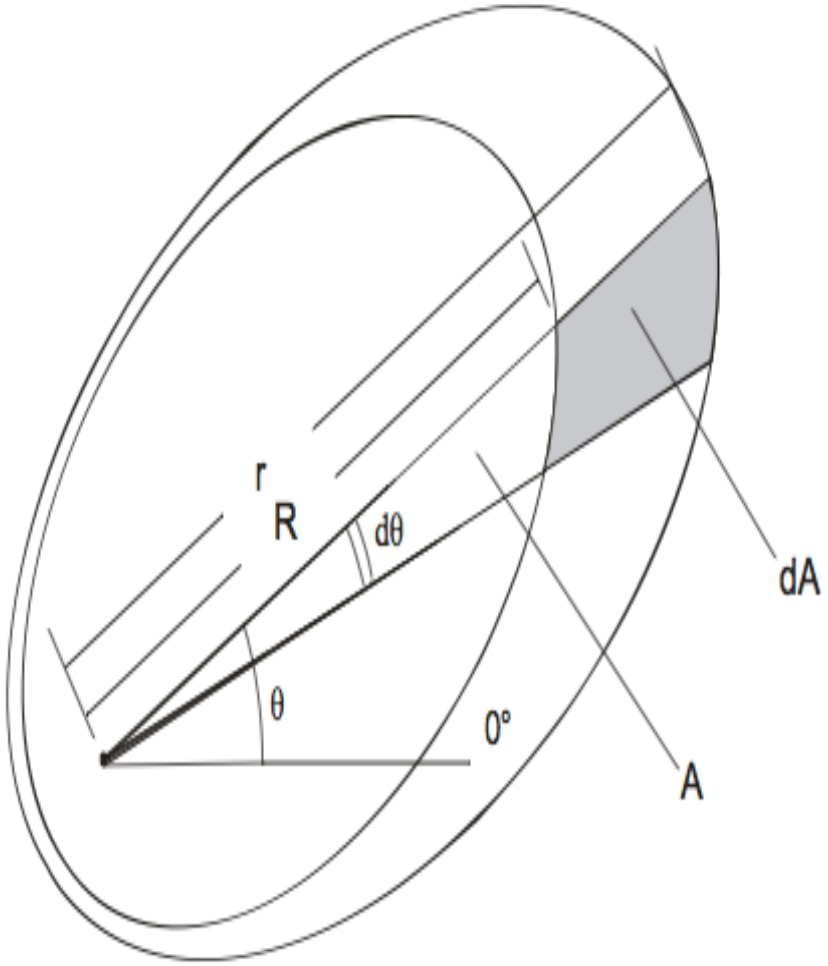
Fig. 7. Distribution of under and over-prediction expressed as a percent errors for the fires associated with Table 1.

Overlap proportion (Duff et al., 2013). Black = overlapping burn area,
white = burn area predicted but not observed, grey = burn area observed but not predicted.

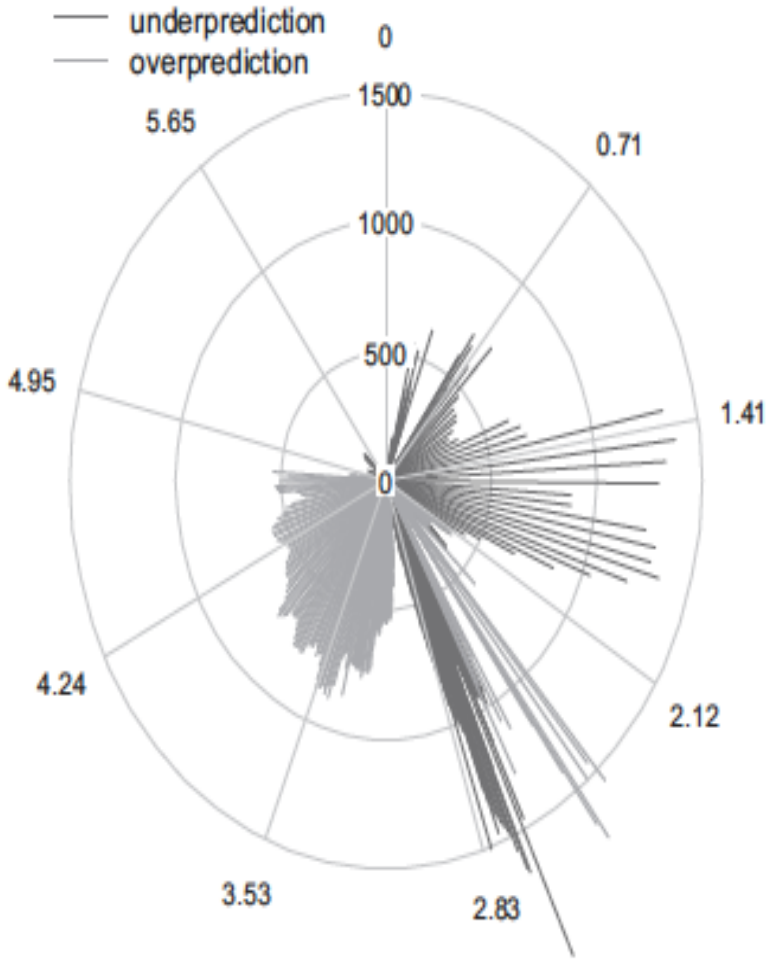


Difference between burn areas, evaluated radially as a function of θ

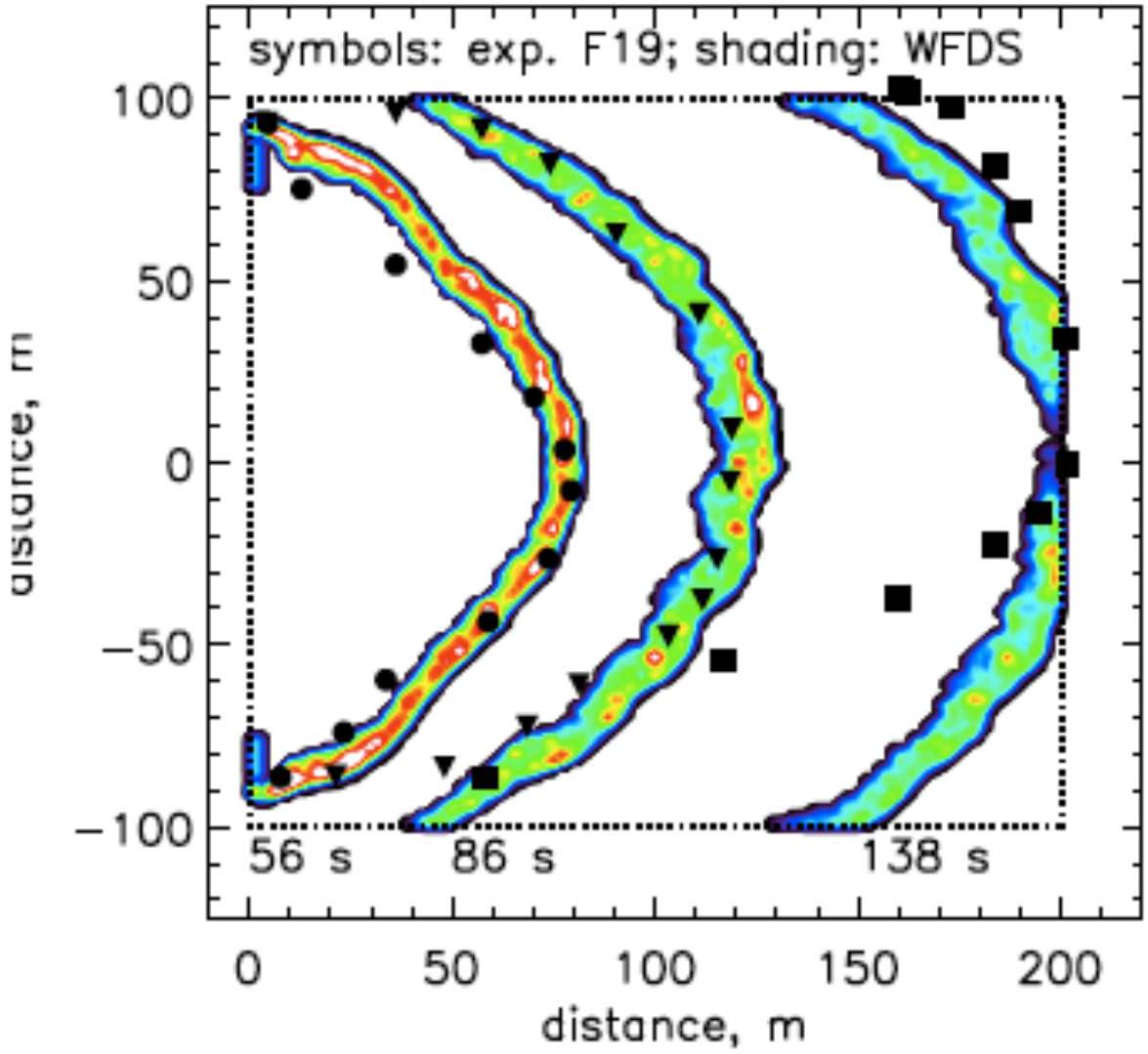
(Cui & Perera, 2010)



(Duff et al., 2013)



Few studies look at evolution of predicted & observed burn areas over time (Mell et al. 2007)



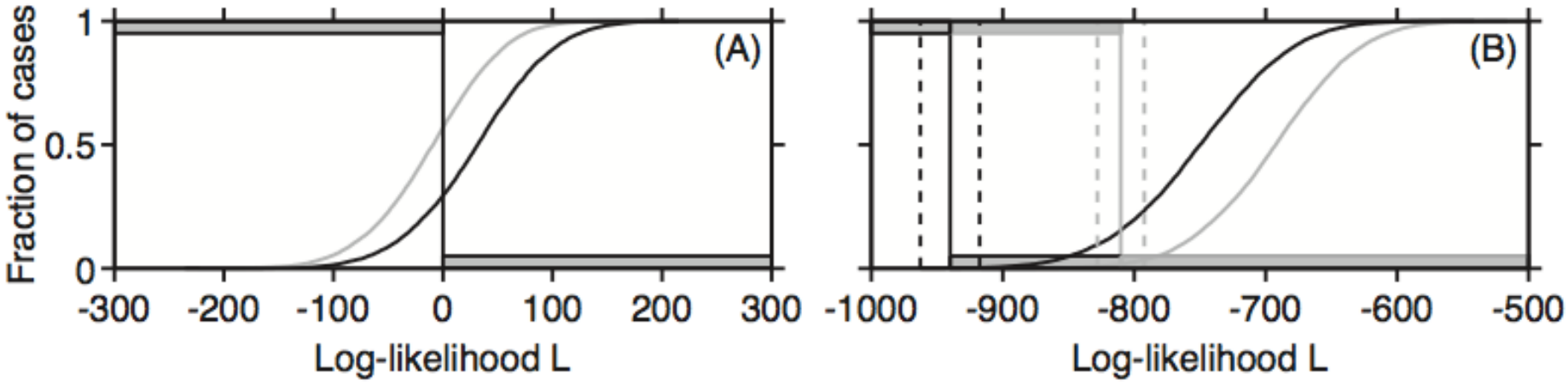
Likelihood methods for spread models.

Model evaluation may be simplified using probabilistic spread models.

Can simply use the **likelihood** as a measure of fit.

If model A routinely assigns lower likelihoods to the observed wildfires than model B, then model B should be preferred.

* Simple, straightforward, meaningful comparisons of multiple competing models.
* Less potential for subtle disagreements between model and data to be amplified.
* L-test can be used to see if discrepancies for 1 model are statistically significant.
  (Simulate, and see if the likelihood for the data is in the middle 95% range.)



Schorlemmer et al. 2007

Currently used evaluation methods for forecasts of occurrence.

Percentage of days with wildfires vs. Index (Viegas et al., 1999)

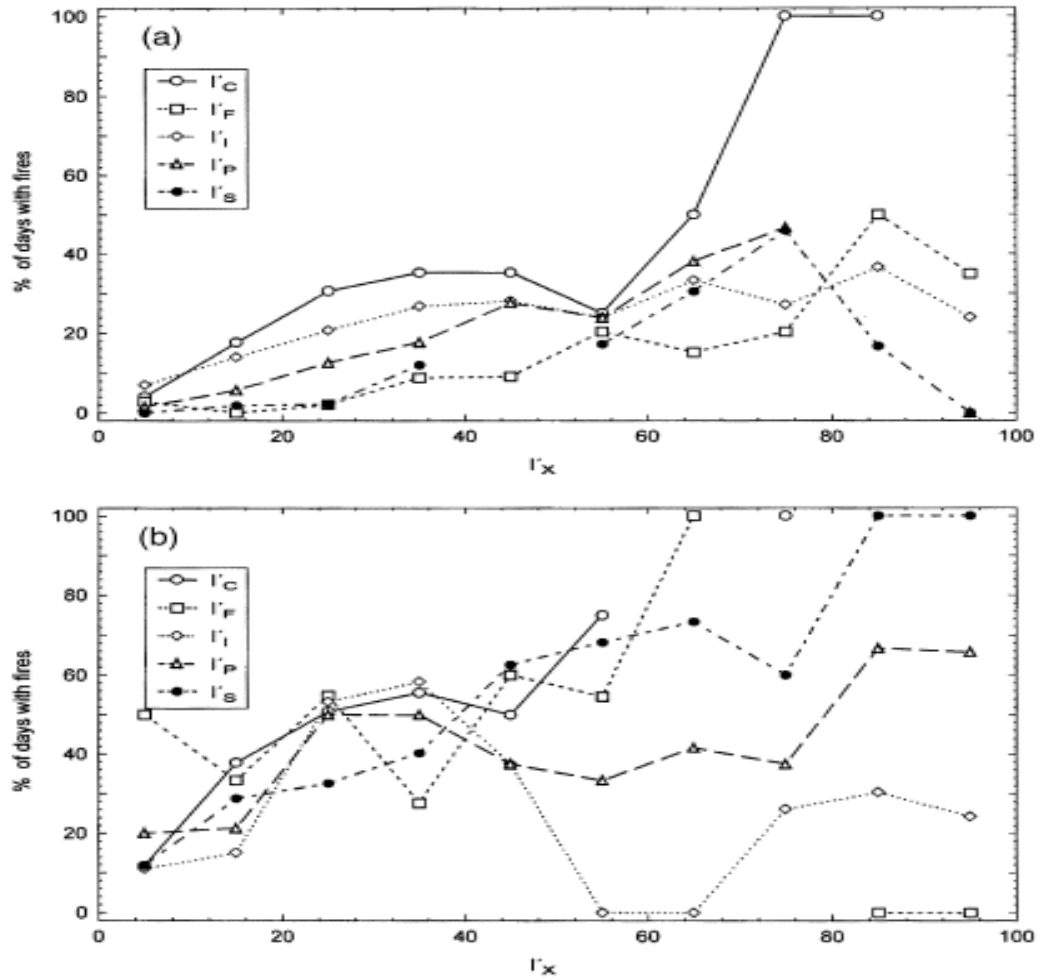


Figure 3. Percentage y_6 of days with fires: (a) A. H. Provence; (b) Veneto.

Mean area burned per day vs. Index (Viegas et al., 1999)

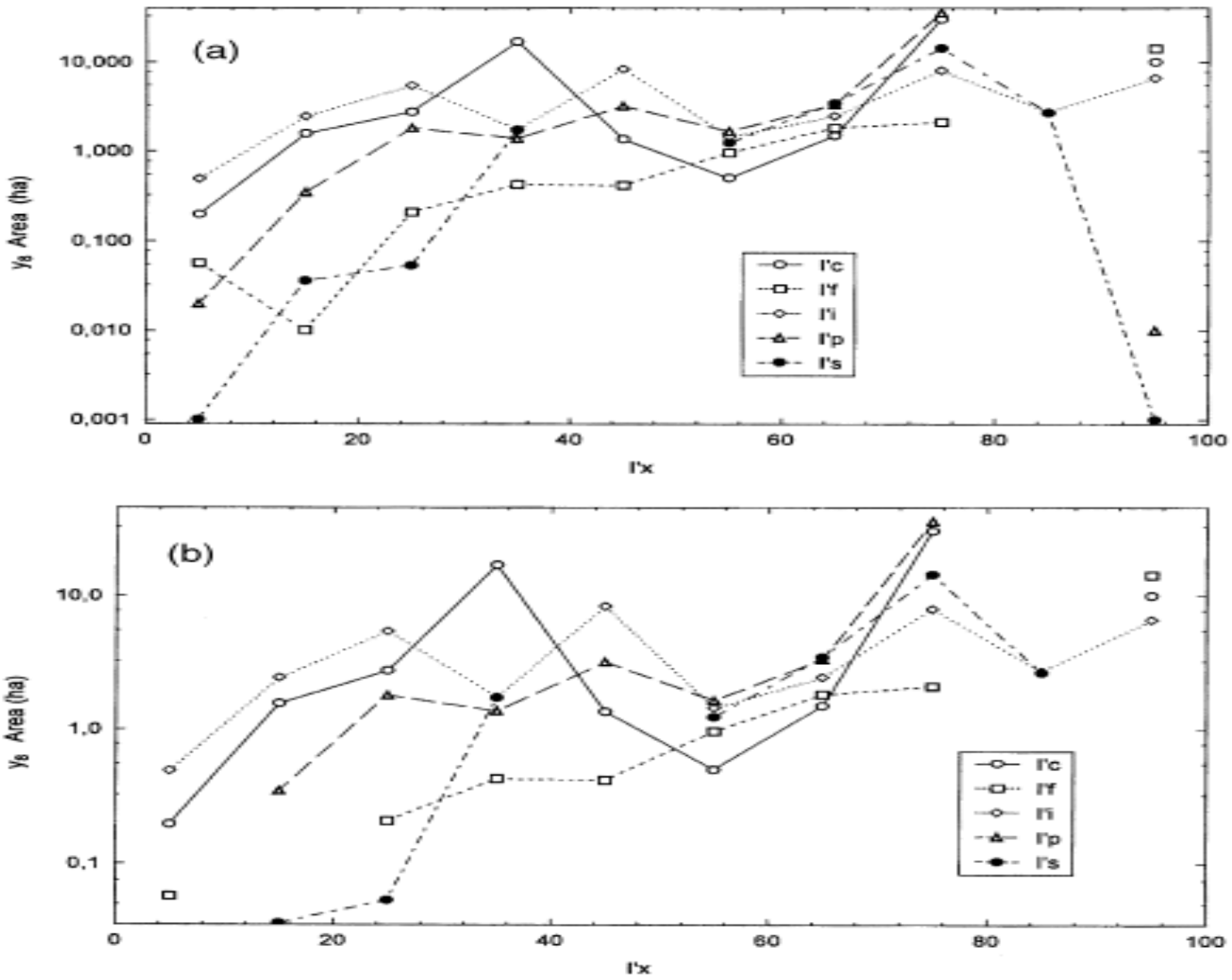


Figure 4. Average area y_8 (ha) burned daily: (a) A. H. Provence; (b) Veneto.

Burning Index (BI)

NFDRS:

Spread Component (SC) and Energy Release Component (ERC), each based on dozens of equations.

$$BI = [10.96 \times SC \times ERC]^{0.46}$$

- Uses daily weather variables, drought index, and vegetation info. Human interactions excluded.
- Predicts: flame length,
  … area/fire? # of fires? # of fires? Total burn area?

Some BI equations: (From Pyne et al., 1996:)

Rate of spread: $R = I_R \, \xi \, (1 + \phi_w + \phi_s) / (\rho_b \varepsilon \, Q_{ig})$. Oven-dry bulk density: $\rho_b = w_0/\delta$.

Reaction Intensity: $I_R = \Gamma' \, w_n \, h \, \eta_M \eta_s$. Effective heating number: $\varepsilon = \exp(-138/\sigma)$.

Optimum reaction velocity: $\Gamma' = \Gamma'_{max} \, (\beta / \beta_{op})^A \exp[A(1- \beta / \beta_{op})]$.

Maximum reaction velocity: $\Gamma'_{max} = \sigma^{1.5} \, (495 + 0.0594 \, \sigma^{1.5})^{-1}$.

Optimum packing ratios: $\beta_{op} = 3.348 \, \sigma^{-0.8189}$. $A = 133 \, \sigma^{-0.7913}$.

Moisture damping coef.: $\eta_M = 1 - 259 \, M_f/M_x + 5.11 \, (M_f/M_x)^2 - 3.52 \, (M_f/M_x)^3$.

Mineral damping coef.: $\eta_s = 0.174 \, S_e^{-0.19}$ (max = 1.0).

Propagating flux ratio: $\xi = (192 + 0.2595 \, \sigma)^{-1} \exp[(0.792 + 0.681 \, \sigma^{0.5})(\beta + 0.1)]$.

Wind factors: $\sigma_w = C U^B \, (\beta/\beta_{op})^{-E}$. $C = 7.47 \exp(-0.133 \, \sigma^{0.55})$. $B = 0.02526 \, \sigma^{0.54}$.
$E = 0.715 \exp(-3.59 \times 10^{-4} \, \sigma)$.

Net fuel loading: $w_n = w_0 \, (1 - S_T)$. Heat of preignition: $Q_{ig} = 250 + 1116 \, M_f$.

Slope factor: $\phi_s = 5.275 \, \beta^{-0.3} \, (\tan \phi)^2$. Packing ratio: $\beta = \rho_b / \rho_p$.

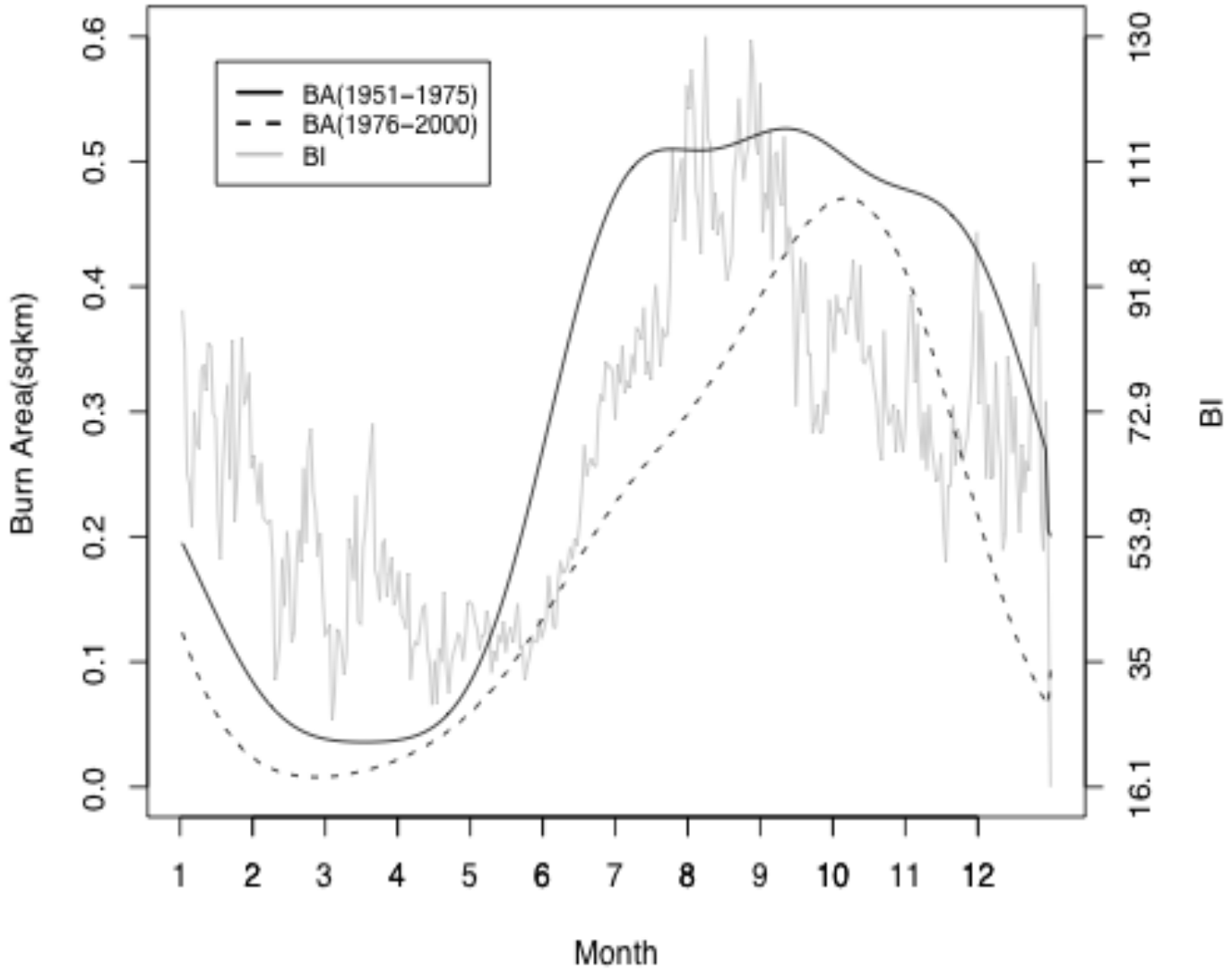
Good news for BI:

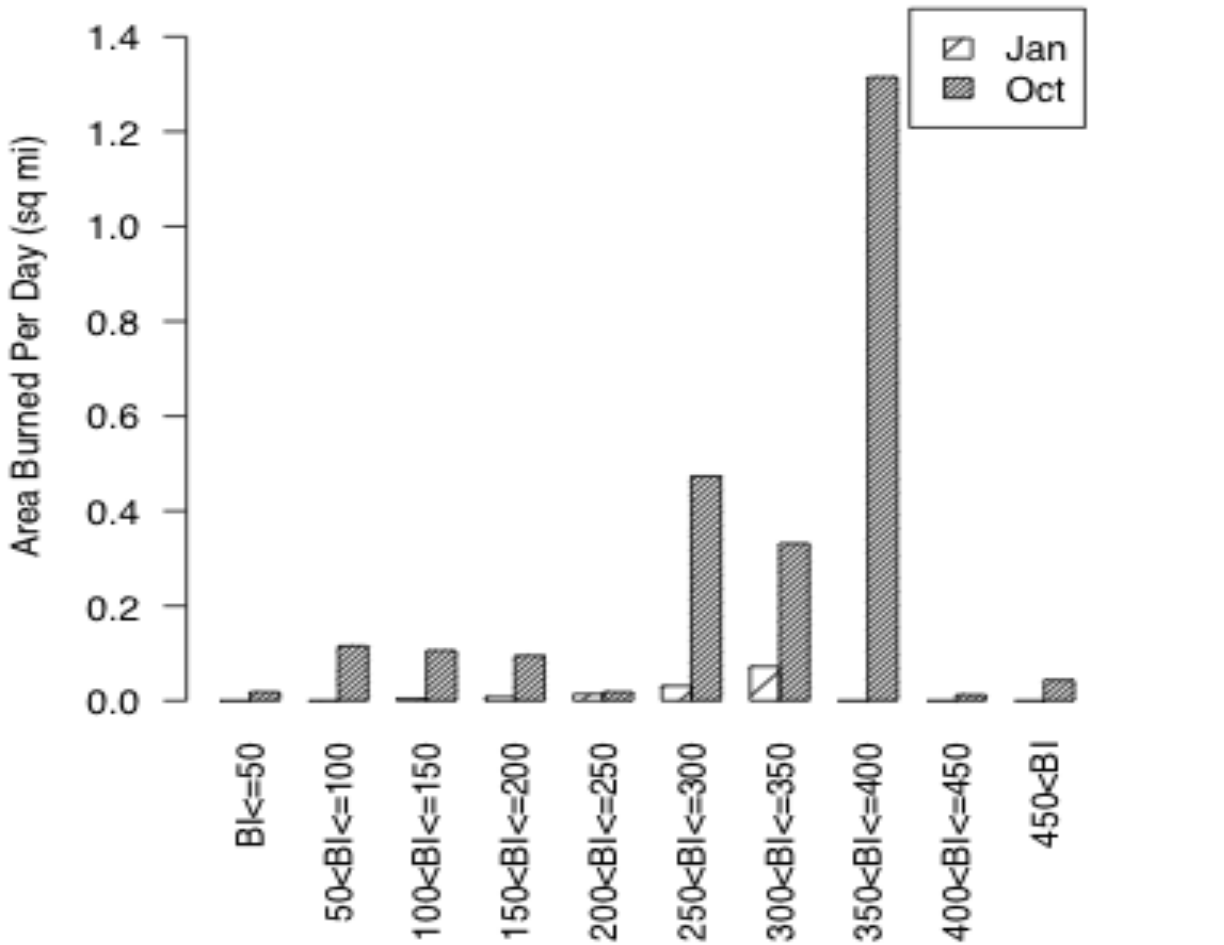
- BI is positively associated with wildfire occurrence.
- Positive correlations with number of fires, daily area burned, and area per fire.
- Properly emphasizes windspeed, relative humidity.

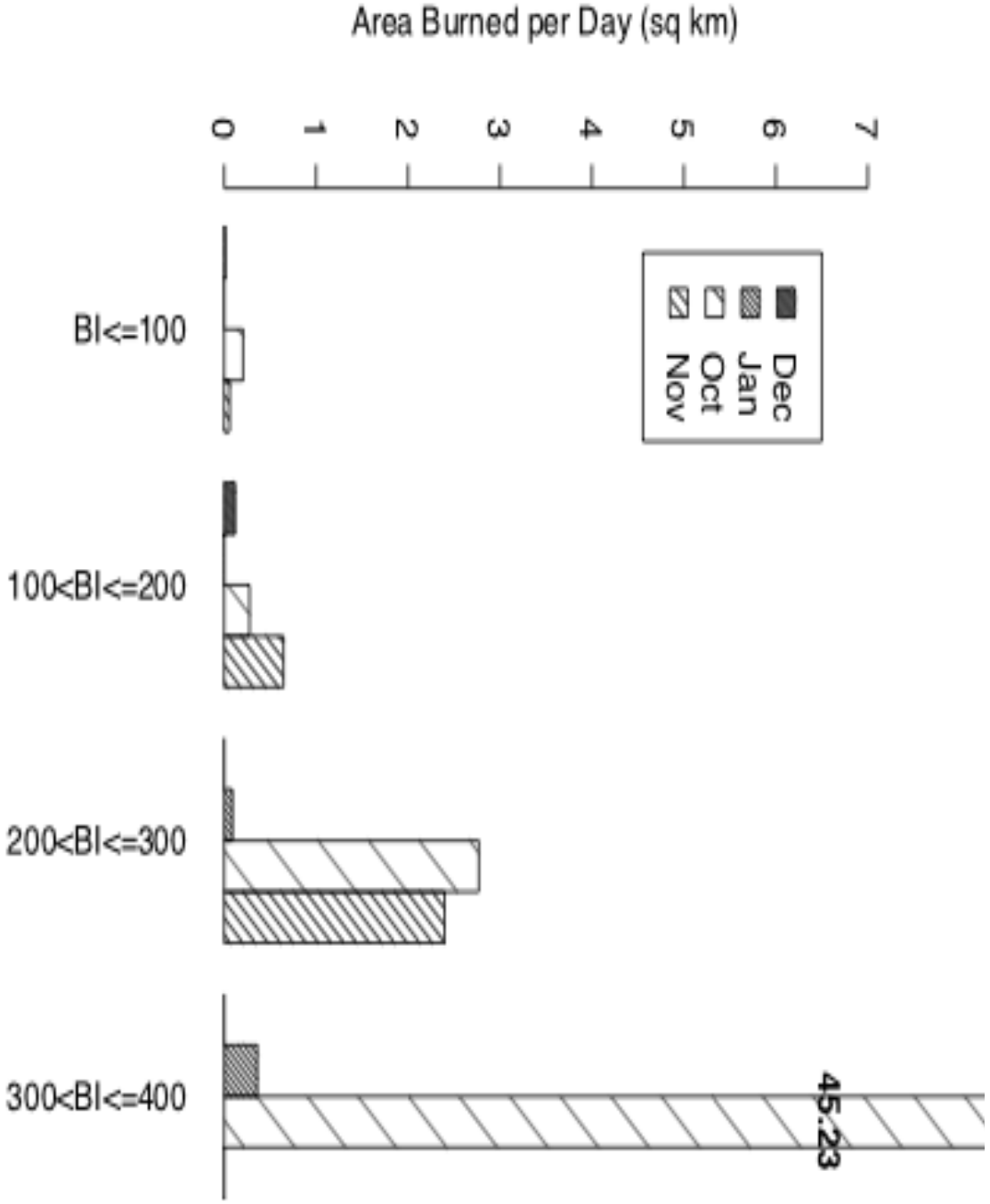
Some problems with BI

- Correlations are low.
  - Corr(BI, area burned) = 0.09
  - Corr(BI, # of fires) = 0.13
  - Corr(BI, area per fire) = 0.076
  - ! Corr(date, area burned) = 0.06
  - ! Corr(windspeed, area burned) = 0.159

- Too high in Winter (esp Dec and Jan)
  Too low in Fall (esp Sept and Oct)

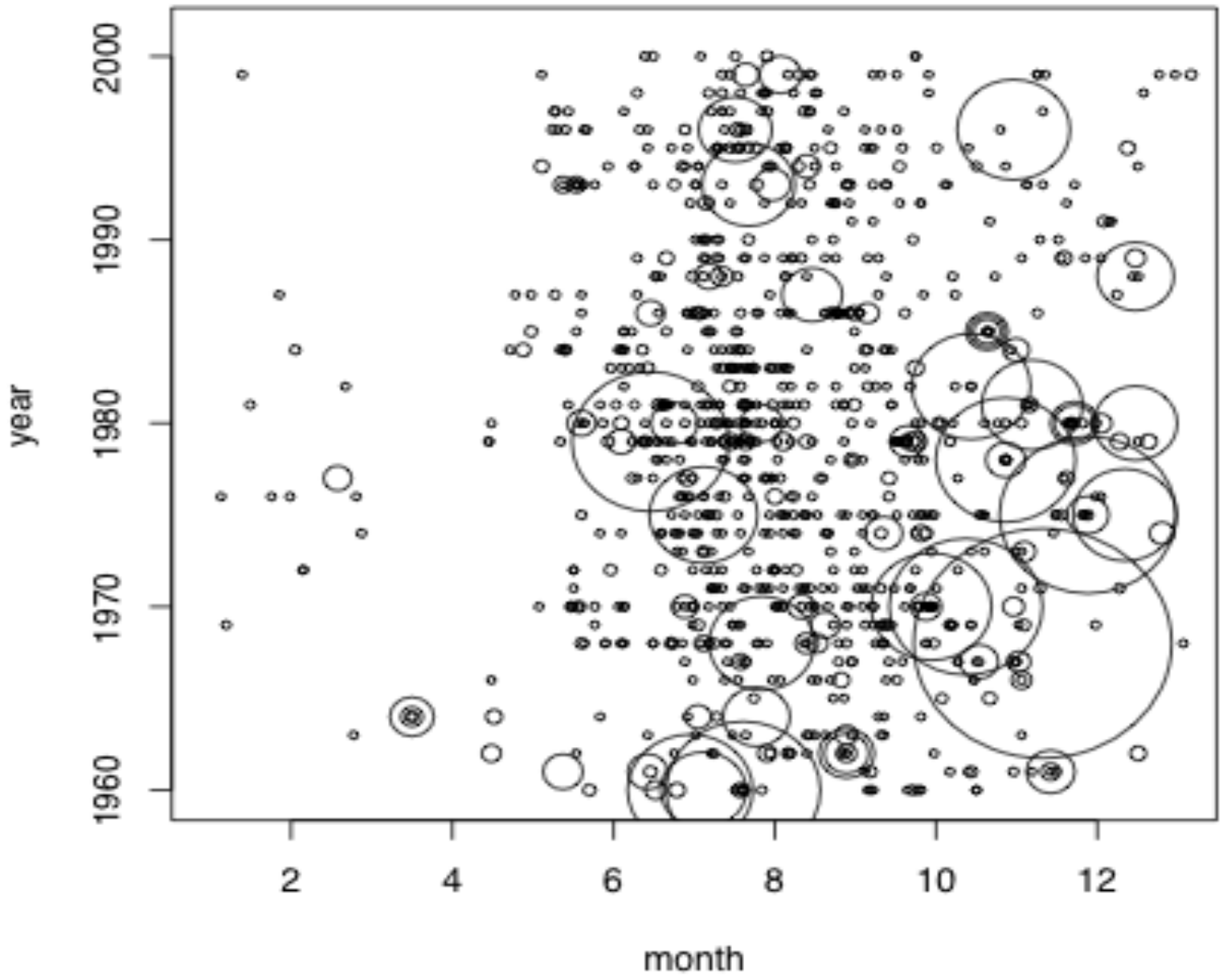

BA(1951–1975)
BA(1976–2000)
BI
Burn Area(sqkm)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
BI
16.1
35
53.9
72.9
91.8
111
130
1
2
3
4
5
6
7
8
9
10
11
12
Month

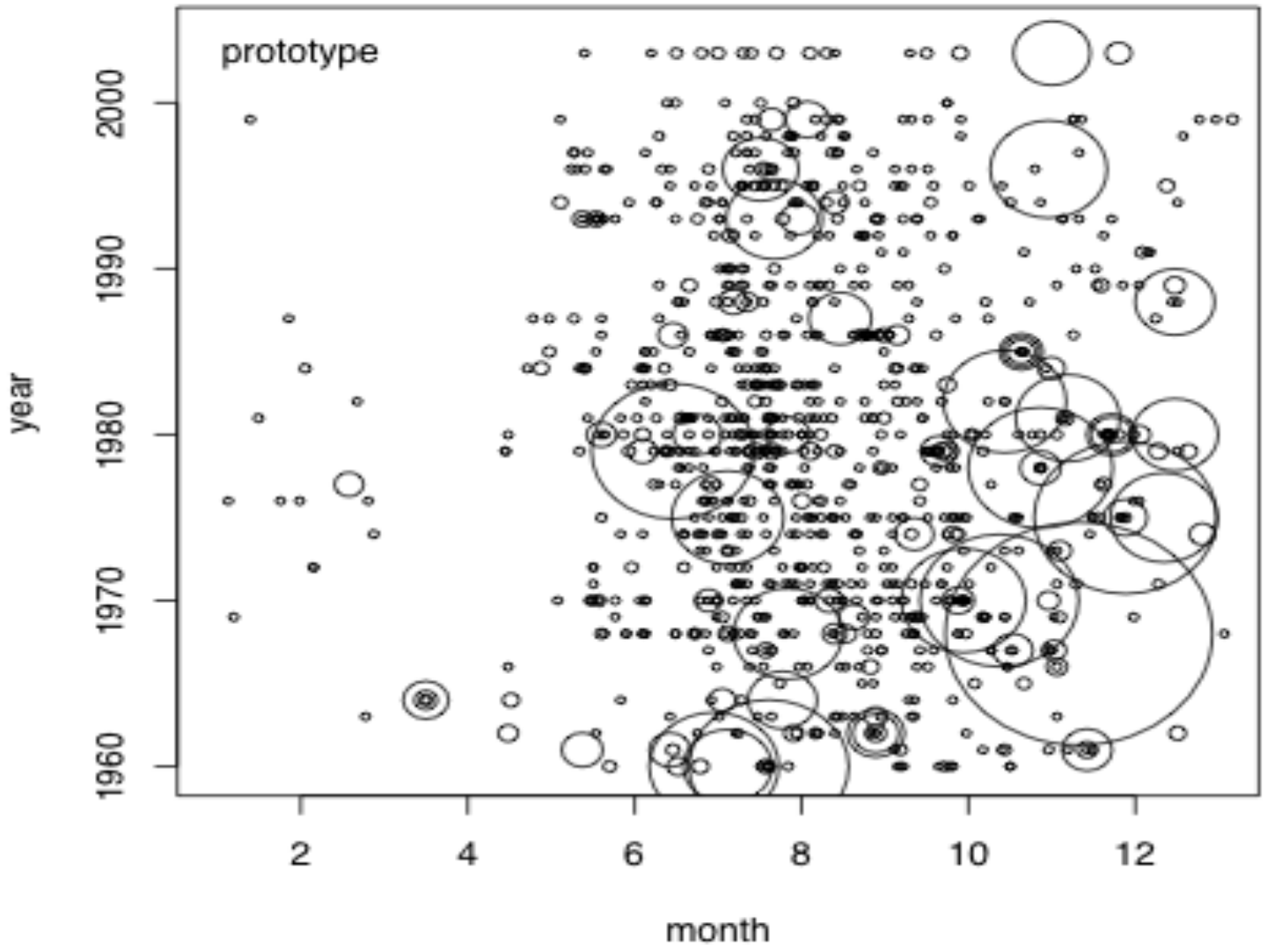

Comparison of Area Burned Per Day & BI (Year>1975)
Area Burned Per Day (sq mi)
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Jan
Oct
BI<=50
50<BI<=100
100<BI<=150
150<BI<=200
200<BI<=250
250<BI<=300
300<BI<=350
350<BI<=400
400<BI<=450
450<BI


Area Burned per Day (sq km)
0
1
2
3
4
5
6
7
BI<=100
100<BI<=200
200<BI<=300
300<BI<=400
Dec
Jan
Oct
Nov
45.23

Separable Estimation for Point Processes

- Consider $\lambda(t, x_1, \ldots, x_k; \theta)$. [For fires, x_1=location, x_2 = area.]


year
1960
1970
1980
1990
2000
2
4
6
8
10
12
month


prototype
year
1960
1970
1980
1990
2000
2
4
6
8
10
12
month

Separable Estimation for Point Processes

- Consider $\lambda(t, x_1, \ldots, x_k; \theta)$. [For fires, x_1=location, x_2 = area.]

- Say λ is *multiplicative in mark* x_j if
  $\lambda(t, x_1, \ldots, x_k; \theta) = \theta_0 \, \lambda_j(t, x_j; \theta_j) \, \lambda_{-j}(t, x_{-j}; \theta_{-j})$,
  where $x_{-j} = (x_1,\ldots,x_{j-1}, x_{j+1},\ldots,x_k)$, same for θ_{-j} and λ_{-j}

- If λ is multiplicative in x_j and if one of these holds, then $\tilde{\theta}_j$, the partial MLE, is consistent.

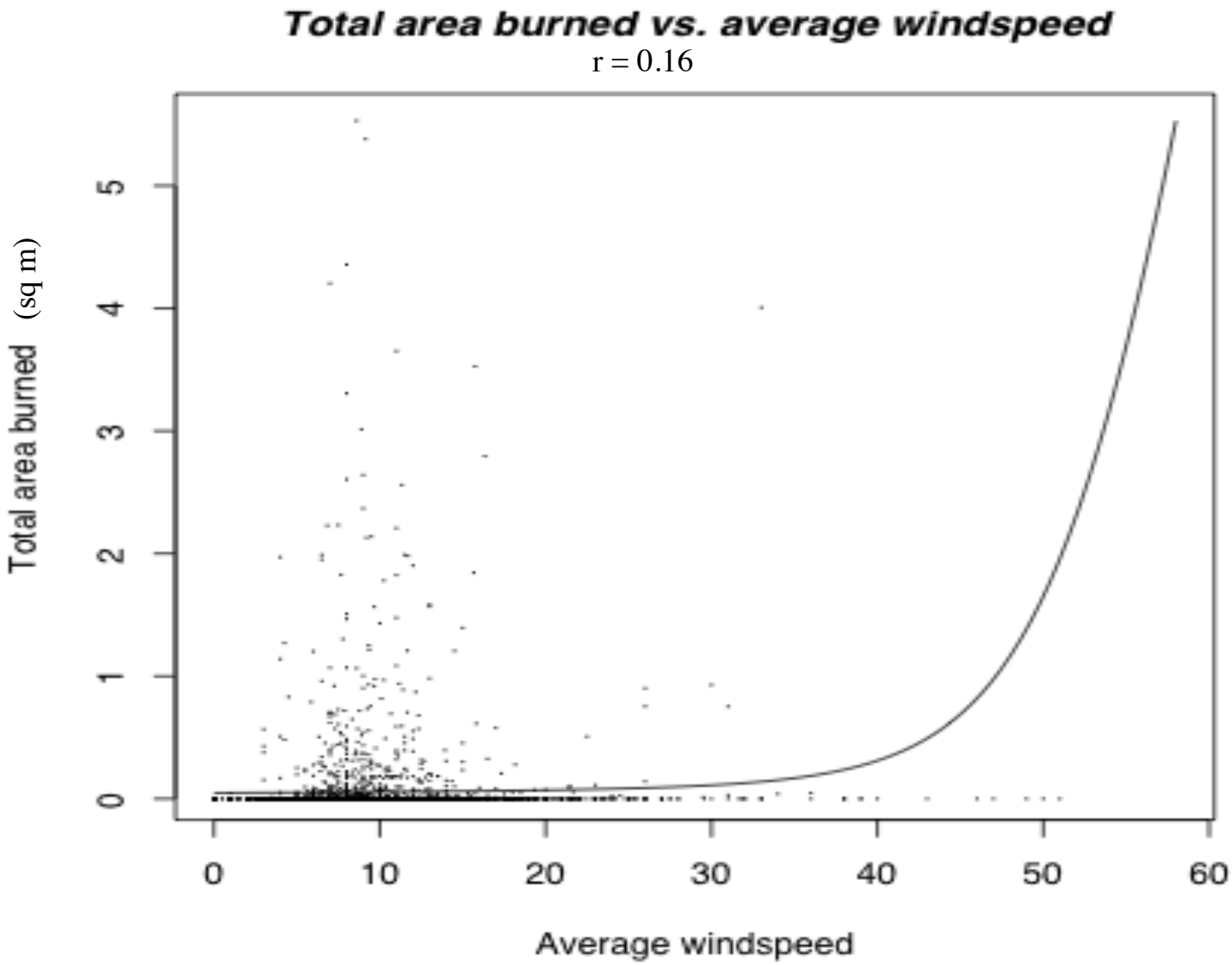
  - $\int \lambda_{-j}(t, x_{-j}; \theta_{-j}) \, d\mu_{-j} = \gamma$, for all θ_{-j}.

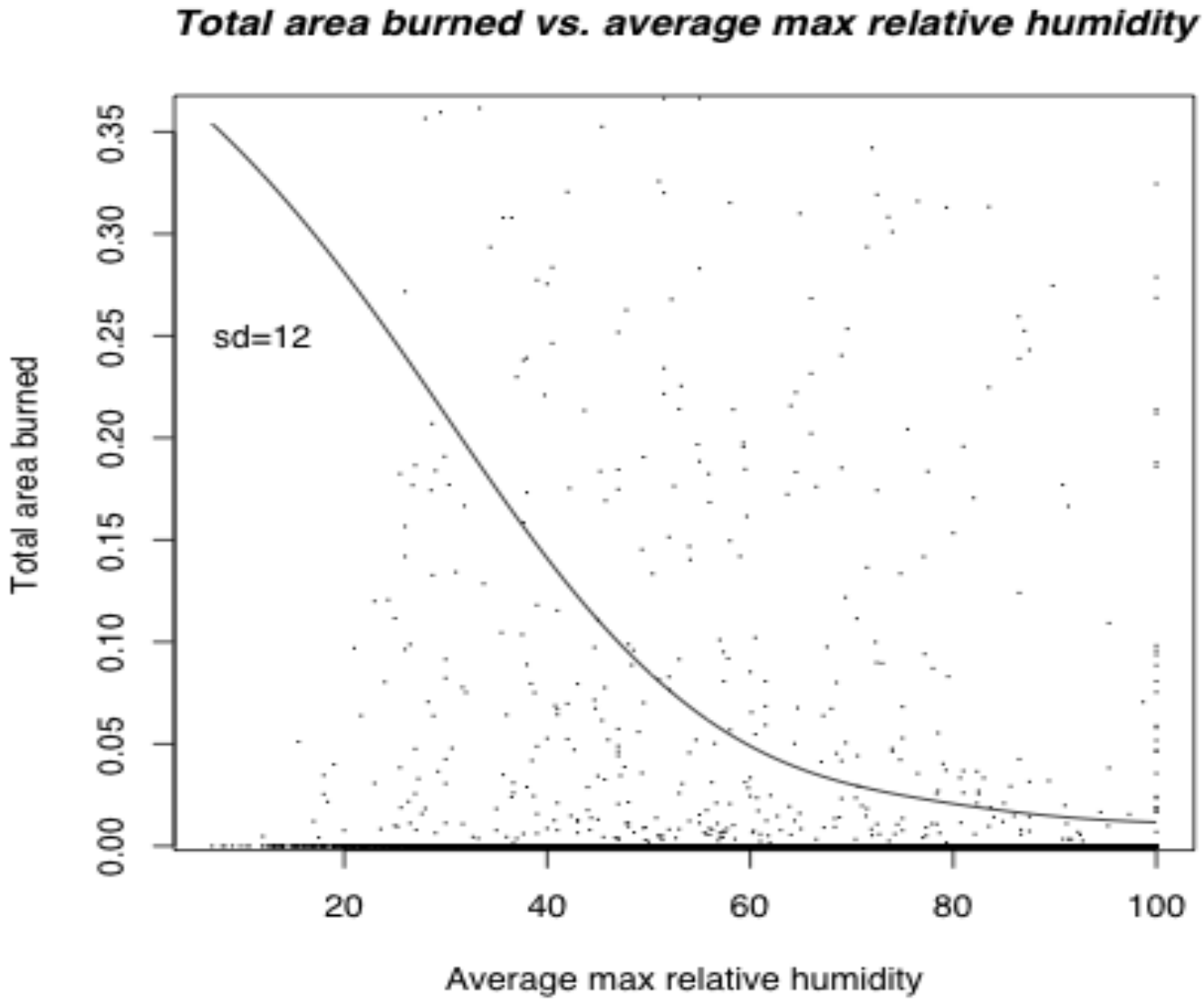
  - $\int \lambda_j(t, x_j; \theta_j) \, d\mu_j = \gamma$, for all θ_j.

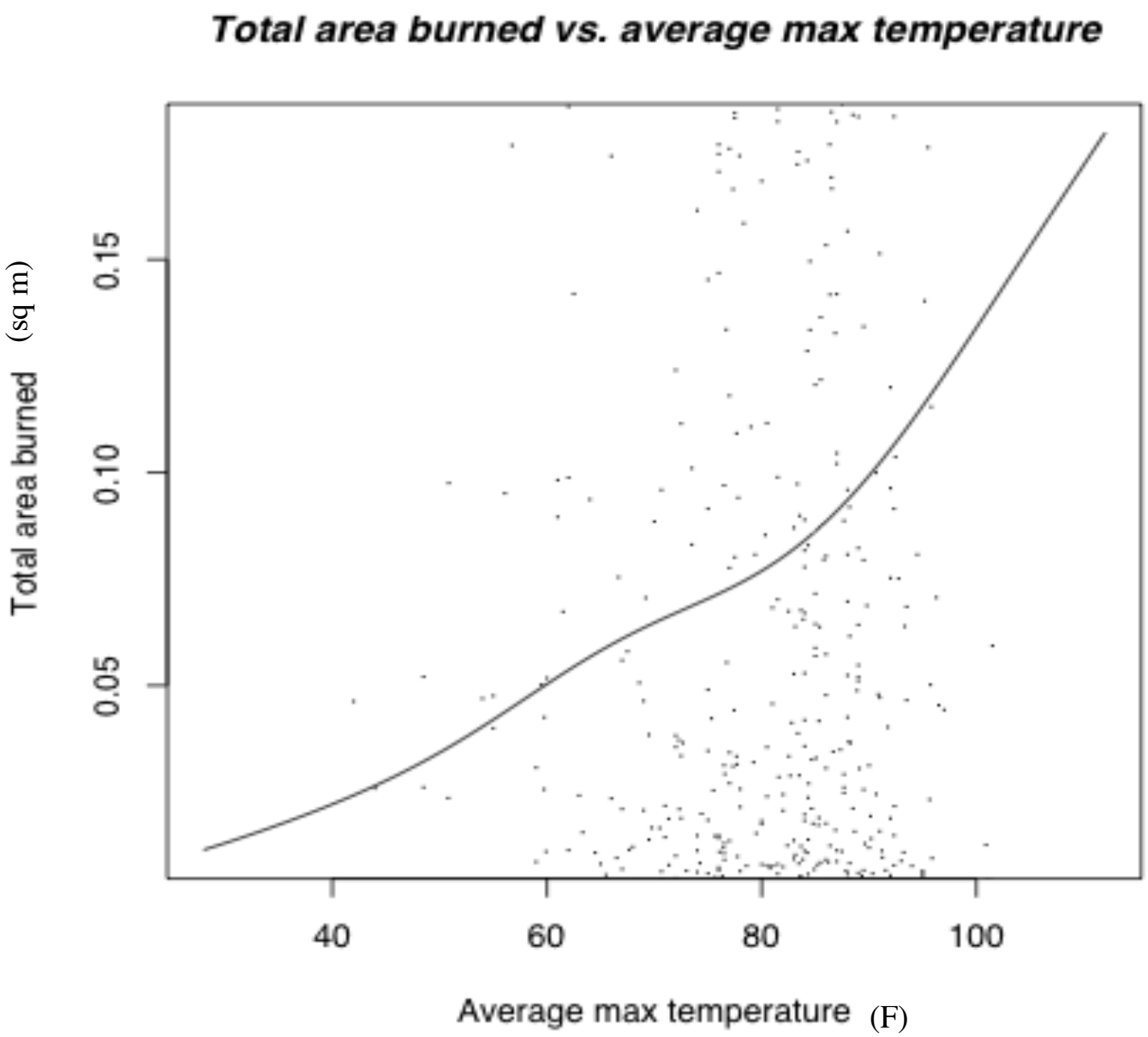
  - $\int \hat{\lambda}_j(t, x; \theta) \, d\tilde{\mu} = \int \lambda_j(t, x_j; \theta_j) \, d\mu_j = \gamma$, for all θ.

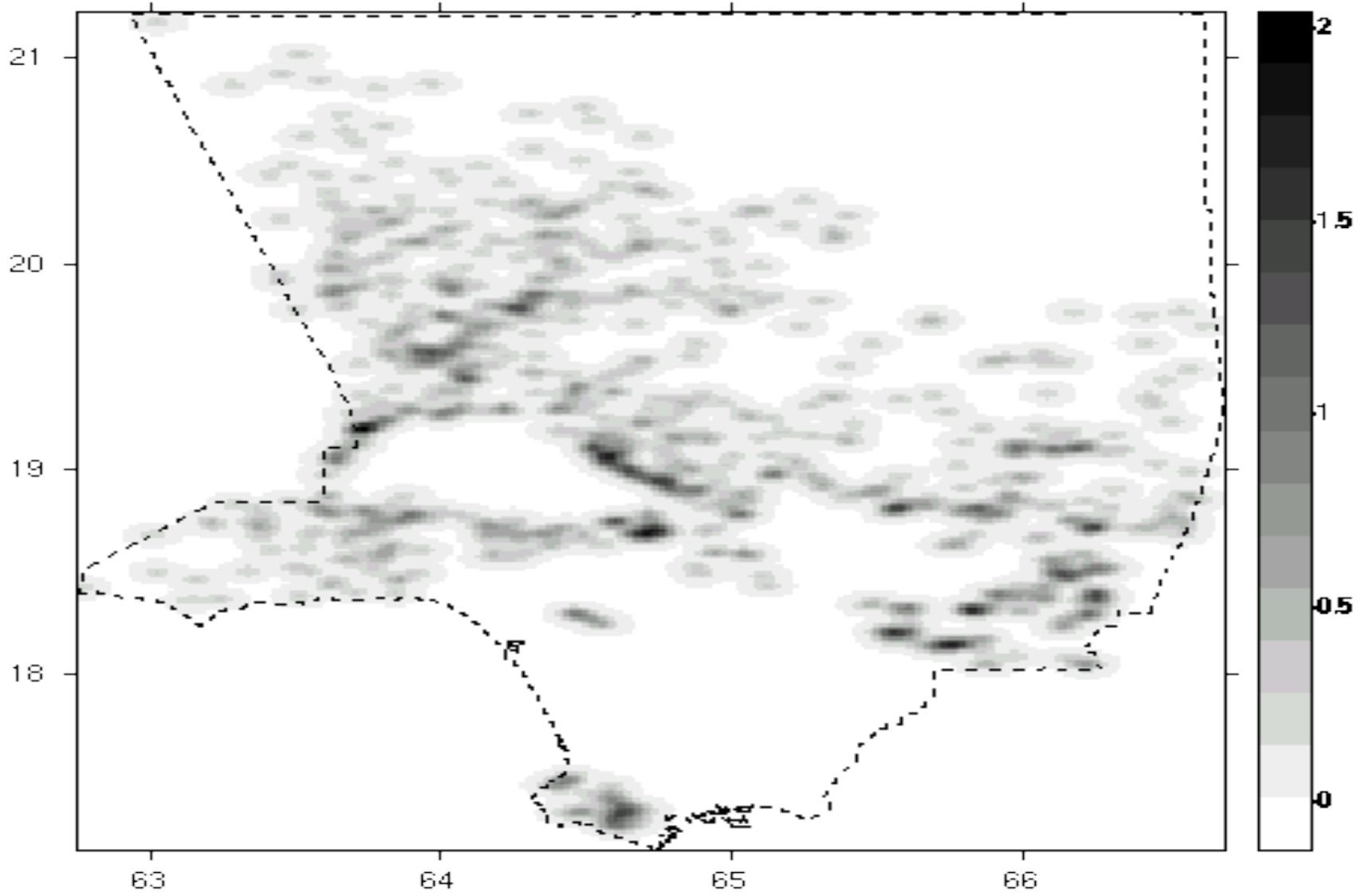
Impact

- Model building.
- Model evaluation / dimension reduction.
- Excluded variables.

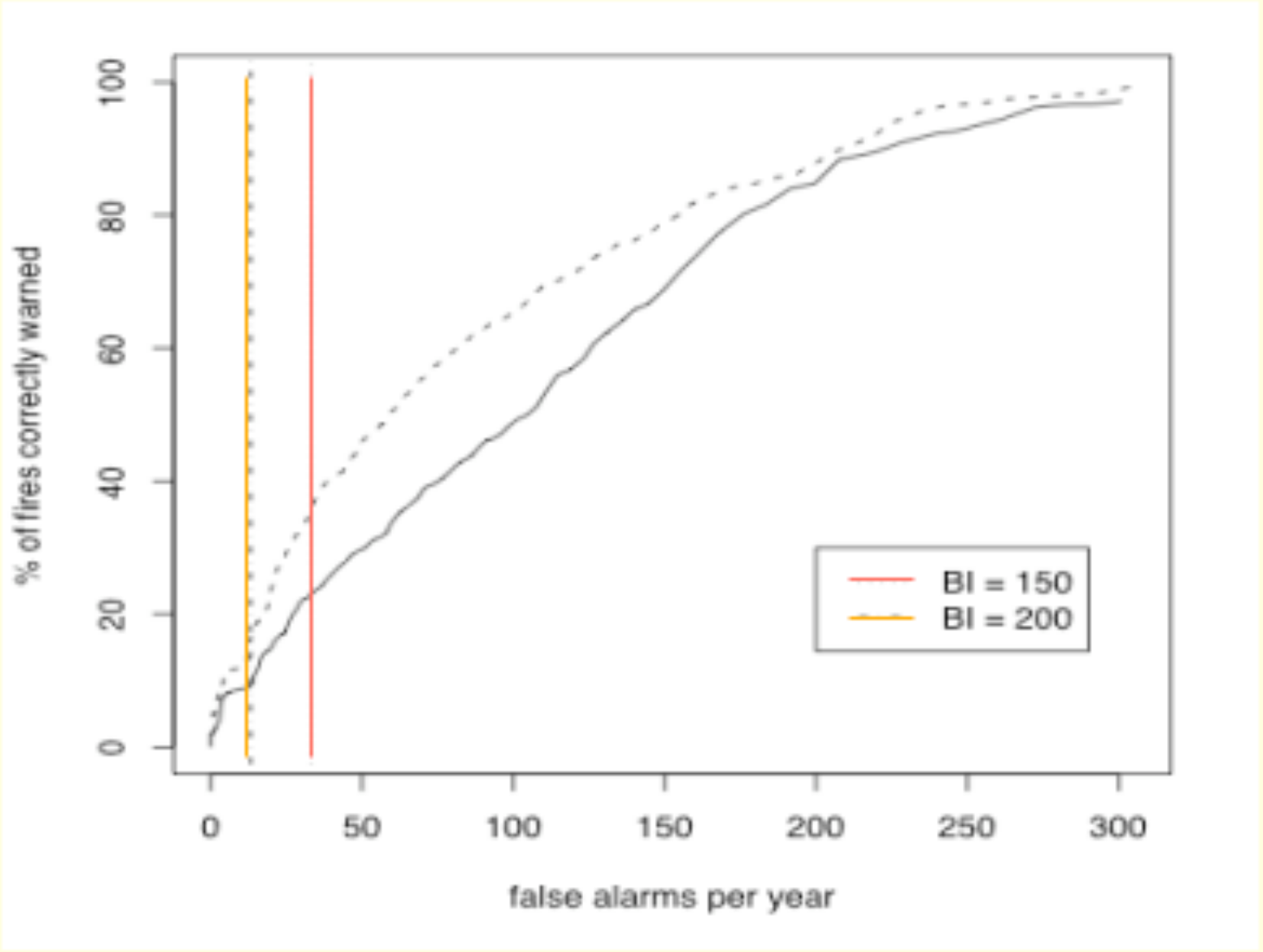

Total area burned vs. average windspeed
r = 0.16
Total area burned (sq m)
0
1
2
3
4
5
0
10
20
30
40
50
60
Average windspeed


Total area burned vs. average max relative humidity
Total area burned
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
sd=12
20
40
60
80
100
Average max relative humidity


Total area burned vs. average max temperature
Total area burned (sq m)
0.05
0.10
0.15
40
60
80
100
Average max temperature (F)


21
20
19
18
63
64
65
66
2
1.5
1
0.5
0

Error diagrams (Molchan 1990, Xu and Schoenberg 2011)



Model Construction

- Wildfire incidence seems roughly multiplicative.
  (only marginally significant in separability test)
- Windspeed. RH, Temp, Precip all matter.
- Tapered Pareto size distribution f, smooth spatial background μ.

[*] $\lambda(t,\mathbf{x},a) =$

$f(a)\,\mu(\mathbf{x})\,\beta_1 \exp(\beta_2 RH + \beta_3 WS)\,(\beta_4 + \beta_5 Temp)(\max\{\beta_6 - \beta_7 Prec, \beta_8\})$

Relative AICs (Model - Constant, so lower is better):

Constant	RH	BI	Model [*]
0	-262.9	-302.7	-601.1

Comparison of Predictive Efficacy

	False alarms per year	% of fires correctly alarmed
BI 150:	32	22.3
Model [*]:	32	34.1
BI 200:	13	8.2
Model [*]:	13	15.1

Forecasting Earthquakes.

The Hawkes process (Hawkes 1971) is a useful form for modeling clustered processes like earthquakes, where

$$\lambda(t,x,y,m) = \mu(x,y,m) + \sum_{i:ti<t} g(t-t_i, x-x_i, y-y_i, m_i).$$

An example is the Epidemic-Type Aftershock Sequence (ETAS) model of Ogata (1988, 1998), which has been used for earthquakes as well as invasive species (Balderama et al. 2012) and crime (Mohler et al. 2011). With ETAS, $\mu(x,y,m) = \mu(x,y)\, f(m)$, and e.g.

$$f(m) \propto \exp\{-\beta(m - m_0)\},$$

$$g(t,x,y,m) = \frac{K_0 \exp\{\alpha(m - m_0)\}}{(t+c)^p (x^2 + y^2 + d)^q},$$

parametric point process estimation is easy?

a. Point processes can generally be very well estimated by maximum likelihood estimation (MLE).

MLEs are generally consistent, asymptotically normal, and efficient

Fisher (1925), Cramer (1946), Ogata (1978), Le Cam (1990).

b. Maximizing a function is fast and easy.

400 yrs of work on this since Newton (1669). Every computer package has a strong quasi-Newton minimization routine, such as *optim()* in *R*.

For instance, suppose the function you want to minimize is

$$f(a,b) = (a-4.19)^4 + (b-7.23)^4.$$

```
f = function(p) (p[1]-4.19)^4 + (p[2]-7.23)^4
p = c(1,1) ## initial guess at (a,b)
est = optim(p,f)$par
```

c. The likelihood function is fairly simple.

In practice, it's convenient to compute log(likelihood), and minimize that.

$$\text{log(likelihood)} = \sum \log\{\lambda(t_i,x_i,y_i)\} - \iiint \lambda(t,x,y)\, dt\, dx\, dy.$$

ETAS (Ogata '98), $\lambda(t,x,y) = \mu\, \rho(x,y) + \sum g(t-t_j, x-x_j, y-y_j; M_j)$,

where $g(t,x,y,M) = K\,(t+c)^{-p}\, e^{a(M-M0)}\, (x^2 + y^2 + d)^{-q}$

or $K\,(t+c)^{-p}\, \{(x^2 + y^2)e^{-a(M-M0)} + d\}^{-q}$.

Parametric point process estimation is easy?

ETAS (Ogata '98), $\lambda(t,x,y) = \mu\, \rho(x,y) + \sum g(t-t_j, x-x_j, y-y_j; M_j)$
where $g(t,x,y,M) = K\,(t+c)^{-p}\, e^{a(M-M0)}\, (x^2 + y^2 + d)^{-q}$
or $K\,(t+c)^{-p}\, \{(x^2 + y^2)e^{-a(M-M0)} + d\}^{-q}$.

To make these spatial and temporal parts of g densities, I suggest writing g as

$g(t,x,y,M) = \{K\,(p-1)\, c^{p-1}\, (q-1)\, d^{q-1} / \pi\}\, (t+c)^{-p}\, e^{a(M-M0)}\, (x^2 + y^2 + d)^{-q}$

or

$g(t,x,y,M) = \{K\,(p-1)\, c^{p-1}\, (q-1)\, d^{q-1} / \pi\}\, (t+c)^{-p}\, \{(x^2 + y^2)e^{-a(M-M0)} + d\}^{-q}$.

Two reasons for this:

1) It is easy to see how to replace g by another density.
2) For each earthquake (t_j,x_j,y_j,M_j),
   $\iiint g(t-t_j, x-x_j, y-y_j; M_j)\, dt\, dx\, dy = K\, e^{a(M-M0)}$.

Actually, parametric point process estimation is hard!

Main obstacles:

A. Small problems with optimization routines.
Extreme flatness, local maxima, choosing a starting value.

B. The integral term in the log likelihood.
$\log(\text{likelihood}) = \sum \log\{\lambda(t_i,x_i,y_i)\} - \iiint \lambda(t,x,y)\, dt\, dx\, dy.$

The sum is easy, but the integral term is extremely difficult to compute.
By far the hardest part of estimation (Ogata 1998, Harte 2010).

Ogata '98: divide space around each pt into quadrants and integrate over them.
PtProcess, Harte '10: uses optim() or nlm(). User must calculate the $\int$ somehow.

Numerical approximation is slow and is a poor approx. for some values of Θ.

1,000 computations x 1,000 events x 1,000 function calls = 1 billion computations.

Problem B. contributes to reluctance to repeat estimation and check on A.

My suggestions.

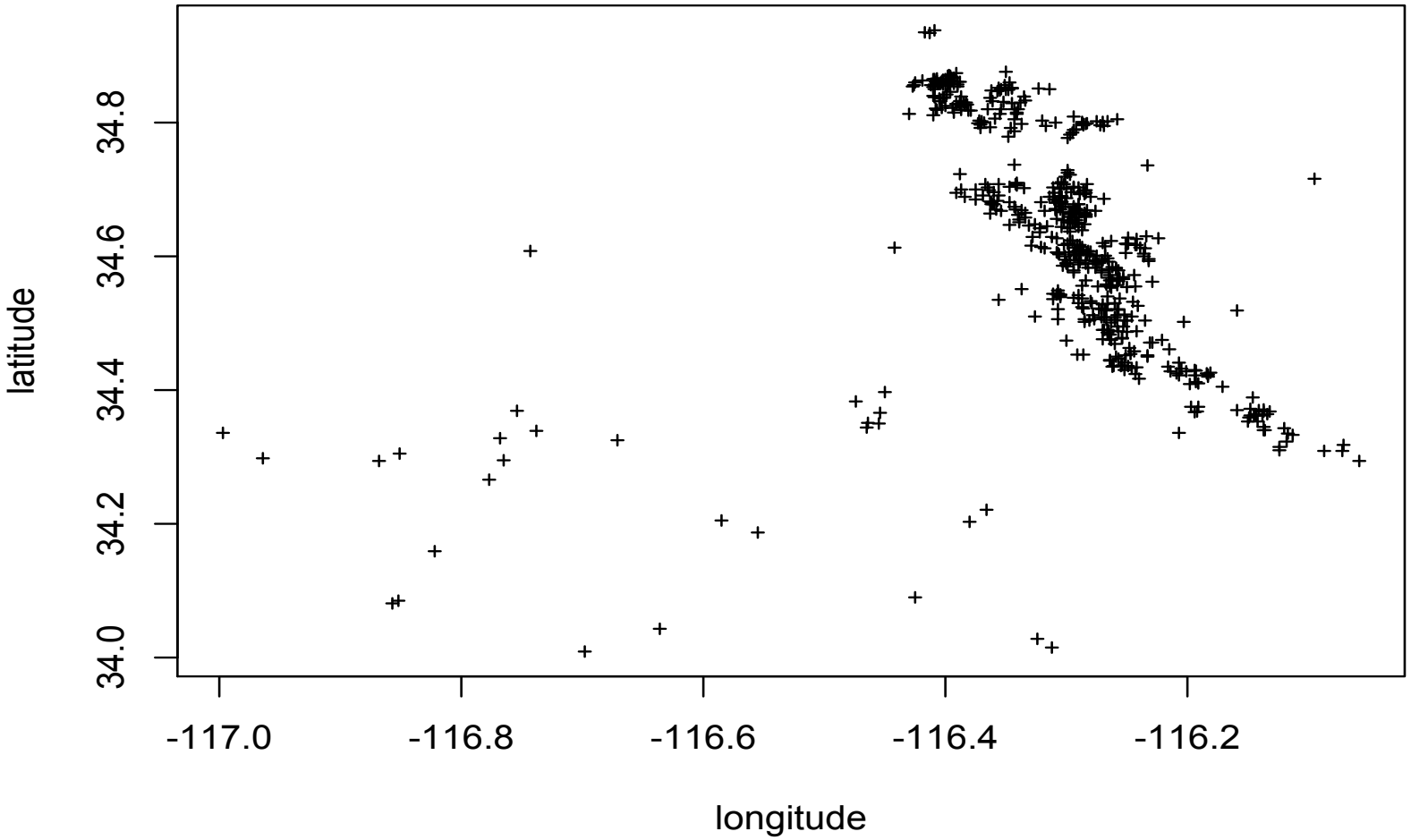
1. Write the triggering function as a density in time and space.
2. Approximate the integral term as $K \sum e^{a(Mj-M0)}$.
3. Given data from time 0 to time T,

   estimate ETAS progressively,

   using data til time T/100, then 2T/100, ..., and assess convergence.

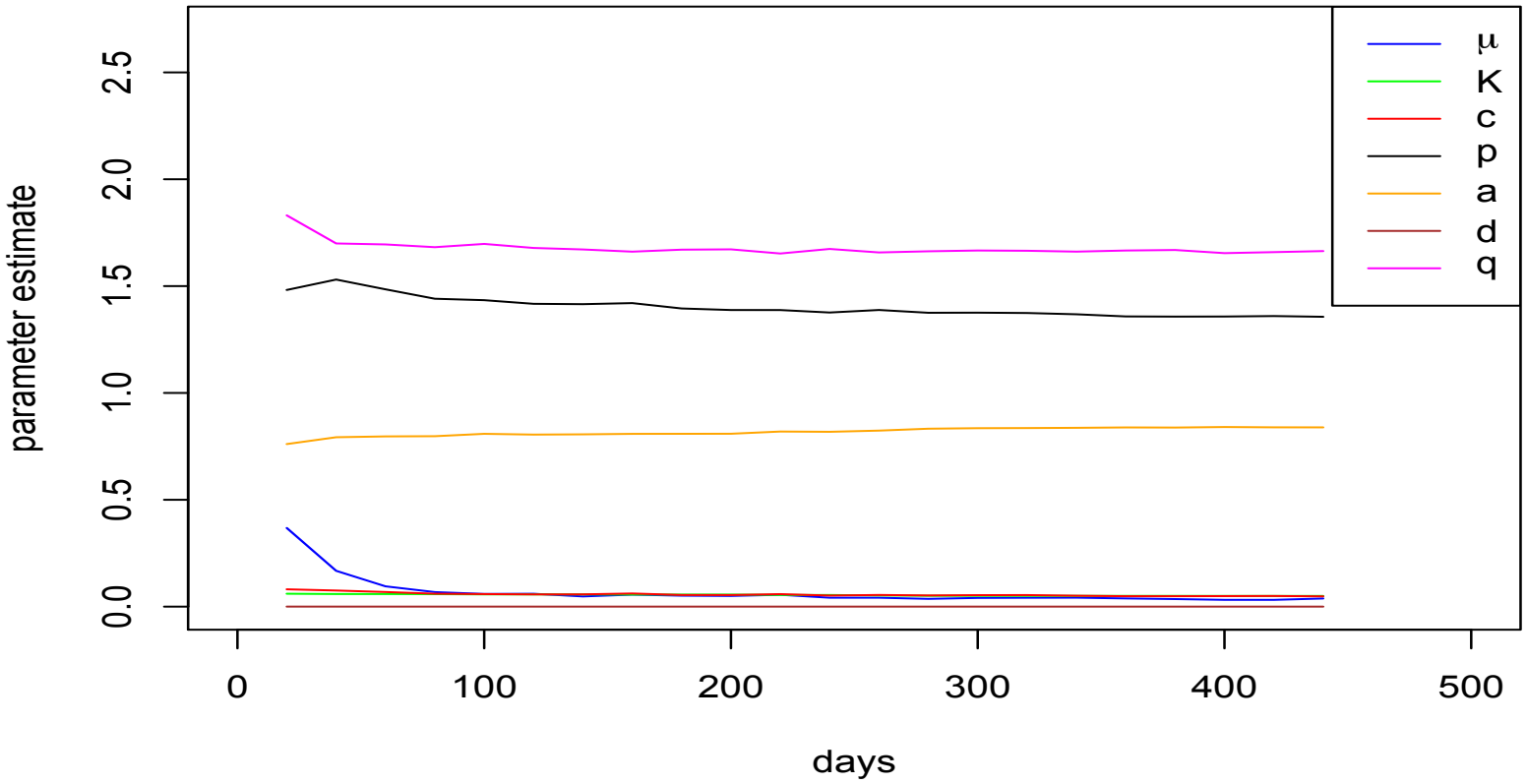
Example.

CA earthquakes from 14 months after Hector Mine, M ≥ 3, from SCSN / SCEDC, as discussed in Ogata, Y., Jones, L. M. and Toda, S. (2003).

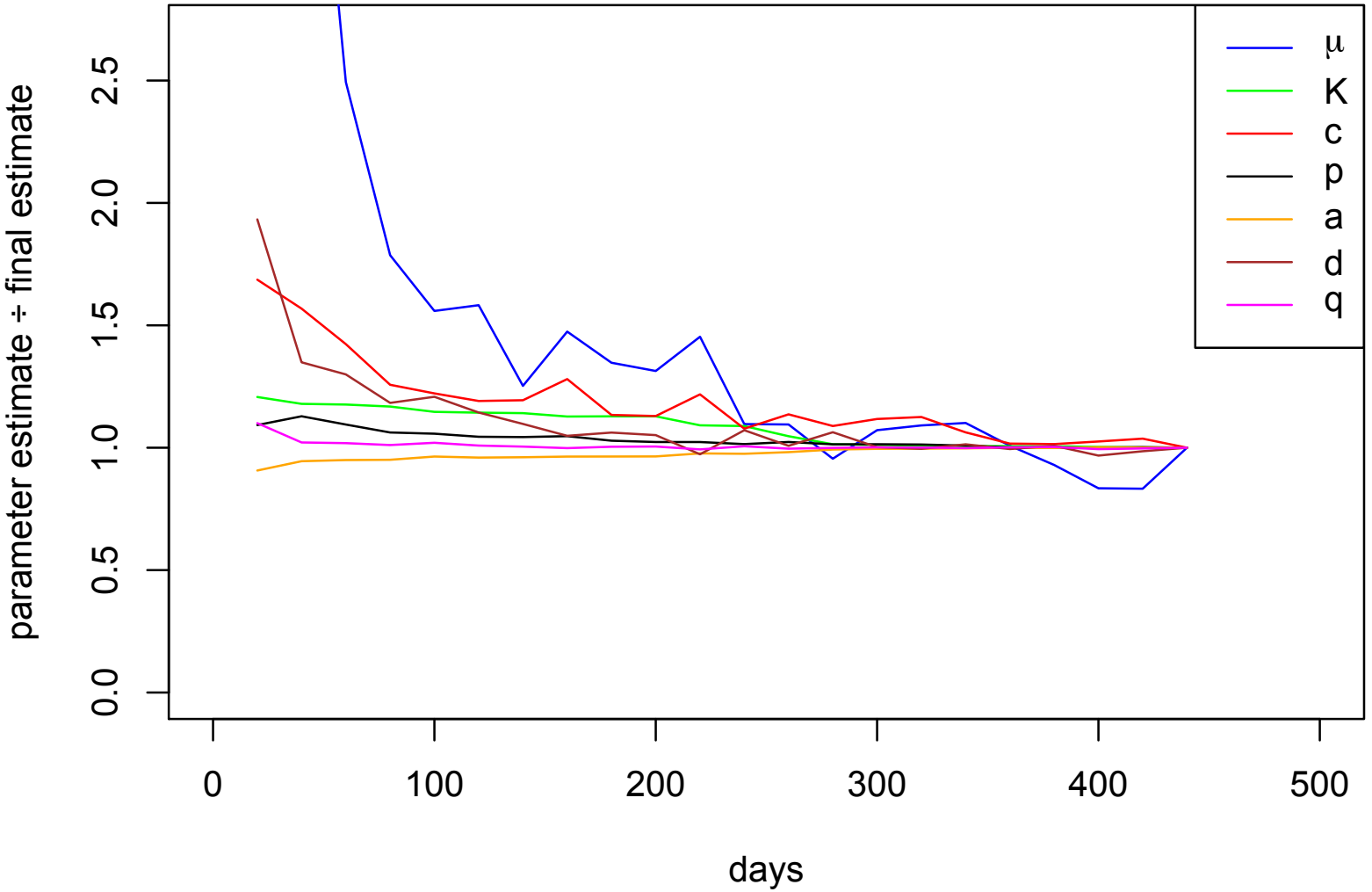


Hector Mine data. Convergence of ETAS parameters is evident after ~ 200 days.

Note that this *repeated* estimation over different time windows is easy if one uses the integral trick. This seems to lead to more stability in the estimates, since a little local maximum is less likely to appear repeatedly in all these estimations.



Hector Mine data. ETAS parameter estimates as ratios of final estimates.

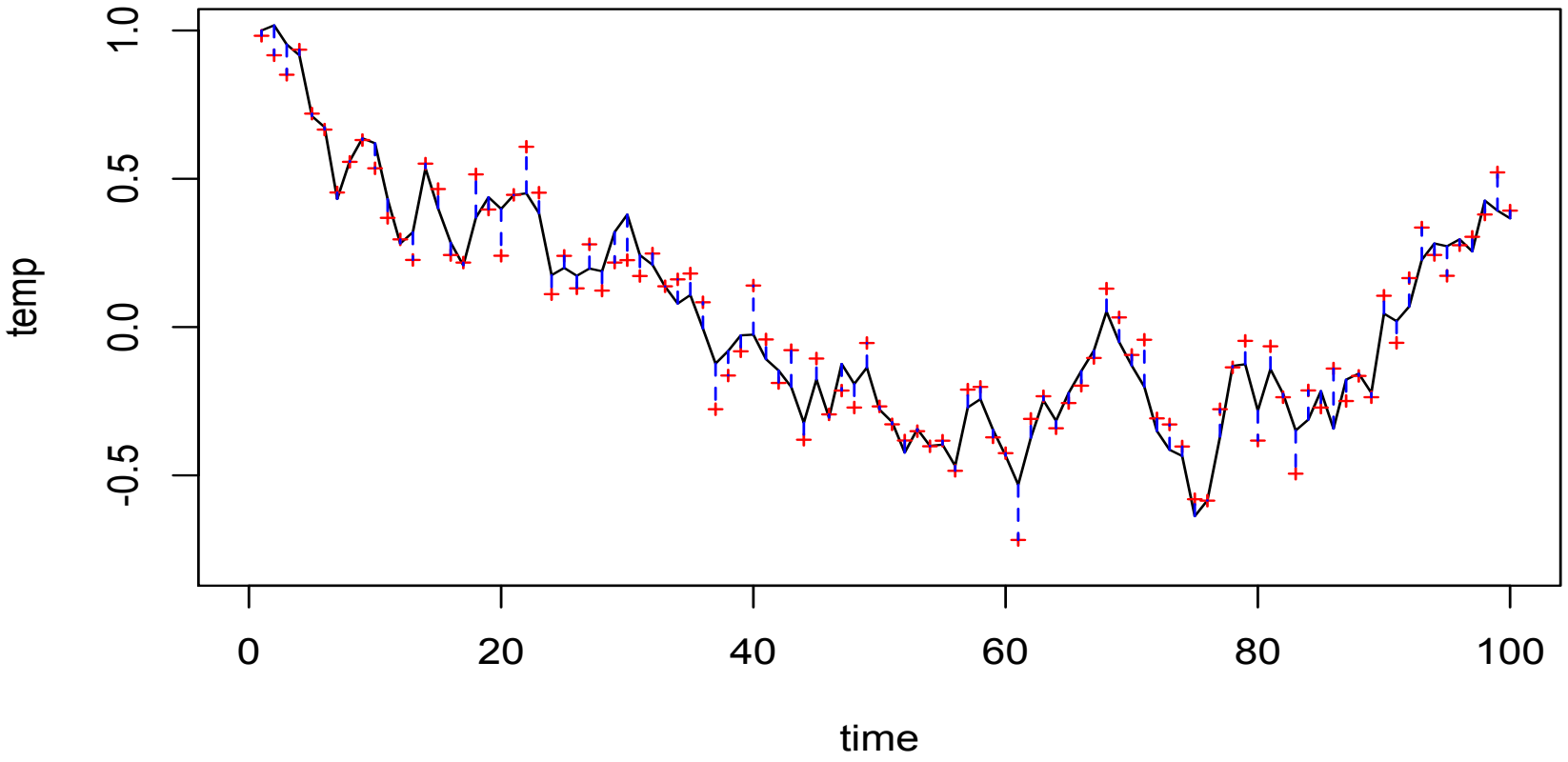


Difficulty of point process model evaluation.

With most types of data, such as regression style data, time series data, or observations on a spatial-temporal grid, can just look at residuals

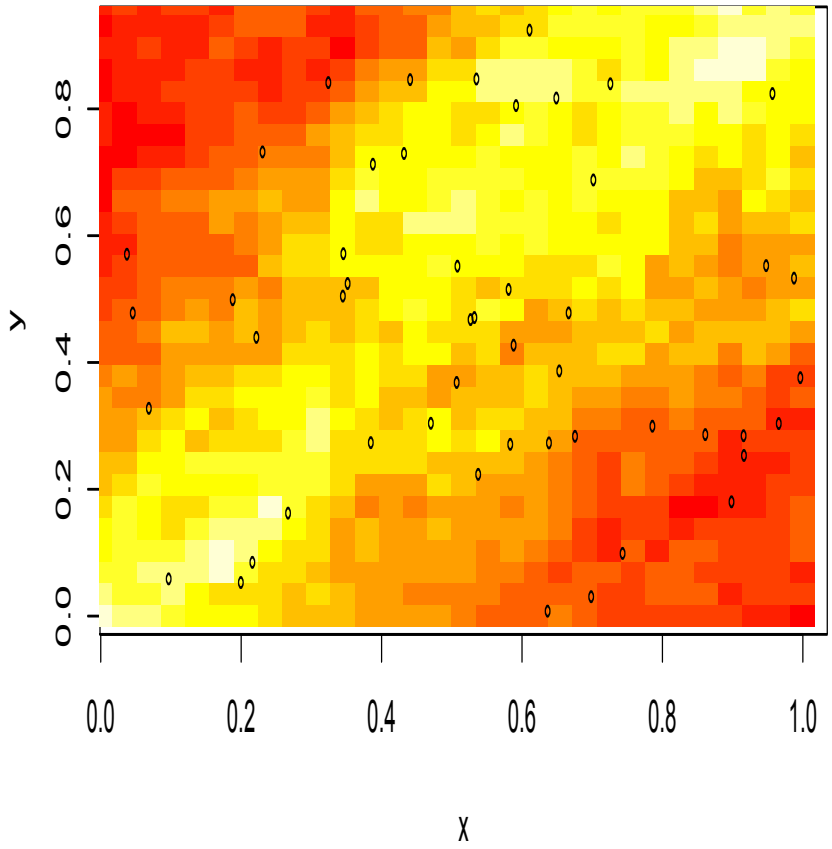
observed – predicted

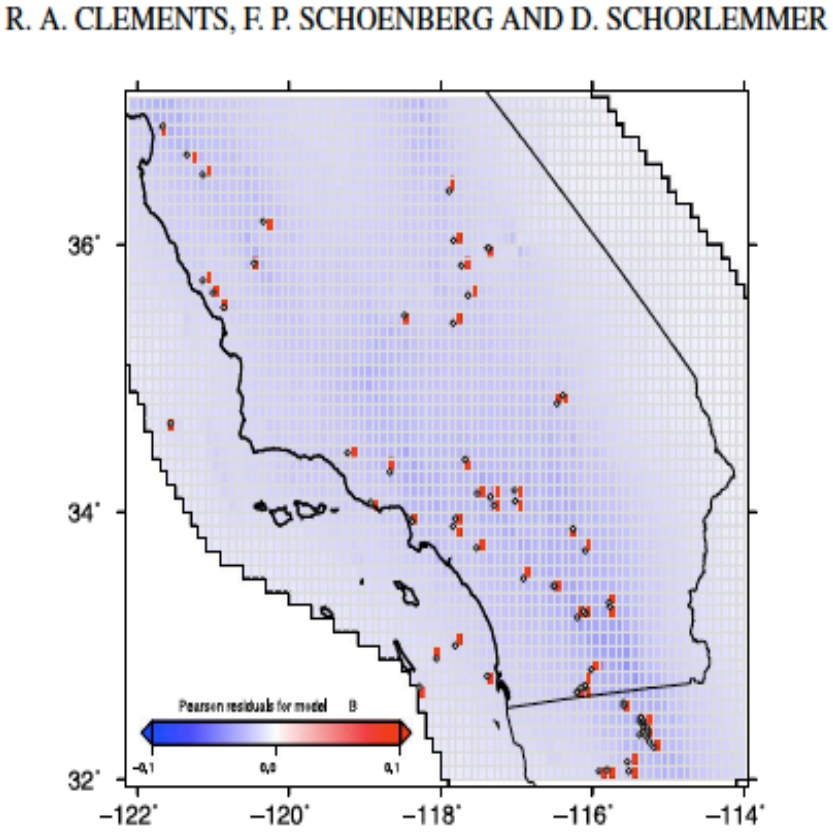
for each observation. Closer to 0 = better.



With point processes, this doesn't work.

With point processes, the observations are a collection of points indicating where and when the phenomenon occurred, and typically the model output is an intensity function
$\lambda(u)$ = anticipated rate of points around spatial-temporal location u.
Can compare number observed and number predicted over each bin, but there are problems.
large pixels yield low power.
with small pixels, residuals are mostly 0 or 1. Highly skewed.
For many models, the residual plot simply looks like a plot of the points themselves.





2. RELM and CSEP.

The Regional Earthquake Likelihood Models (RELM) project [Field (2007)] led to the Collaboratory for the Study of Earthquake Predictability (CSEP) [Jordan (2006)].

RELM tested earthquake forecast models for California.

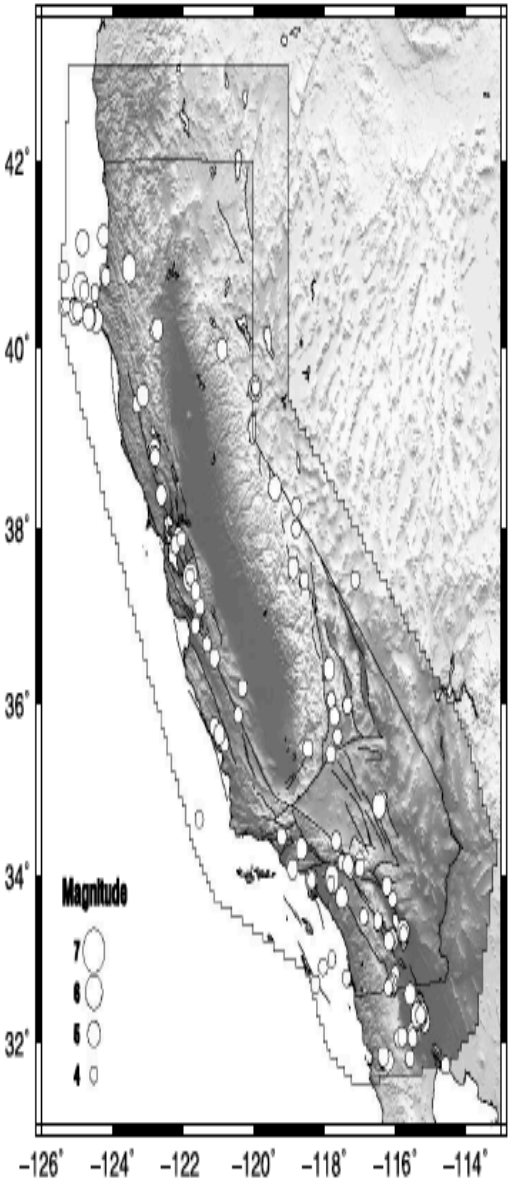
Rigorous, prospective 5 year testing in a dedicated testing center, 1/1/06-1/1/11. [Schorlemmer and Gerstenberger (2007)].

CSEP expanded to regional earthquake forecasts around the world, including California, Japan, New Zealand, Italy, the Northwest Pacific, the Southwest Pacific and the entire globe. Testing centers in Japan, Switzerland, New Zealand and the United States.

CSEP models are five-year, one-day, or recently three-month forecasts. Forecast an expected number of events in each space time magnitude bin. For 1 day models, bins are 0.1◦ lon by 0.1◦ lat by 0.1M from M3.95 to 8.95.

For M8.95-10, one bin of size 0.1◦ by 0.1◦ by 1.05M.

The U.S. testing center is located at the So. California Earthquake Center (SCEC) and hosts forecast experiments for California, the Northwest and Southwest Pacific, and the global experiments.



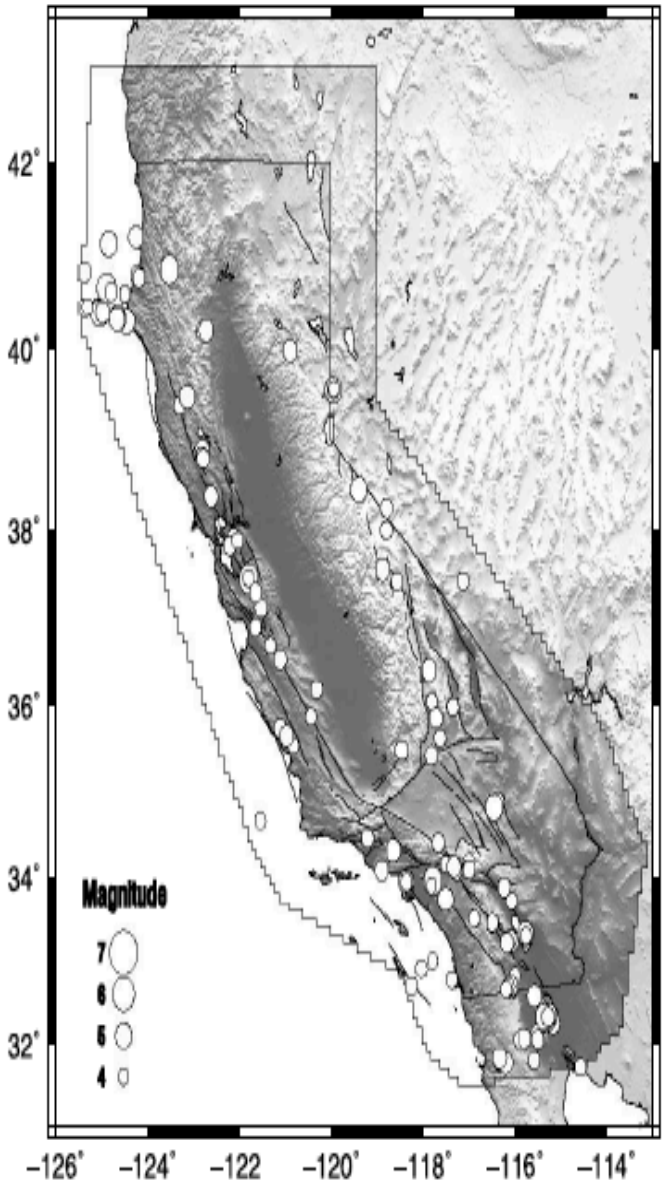
Some models in CSEP.

A. Helmstetter, Kagan and Jackson (2007)
B. Kagan, Jackson and Rong (2007).
C. Shen, Jackson and Kagan (2007).
Epidemic-Type Aftershock Sequence (ETAS) model [Zhuang, Ogata and Vere-Jones (2004), Ogata and Zhuang (2006)].
Short-Term Earthquake Probabilities (STEP) model [Gerstenberger et al. (2005)].

All based exclusively on previous seismicity except C, which uses geodetic and geological info.

Earthquake catalogs were obtained from the Advanced National Seismic System (ANSS).

142 shallow earthquakes with a M≥3.95 occurred in RELM's spatial temporal window.
Note that each RELM model does not necessarily produce a forecasted seismicity rate for every pixel in the space–time region.

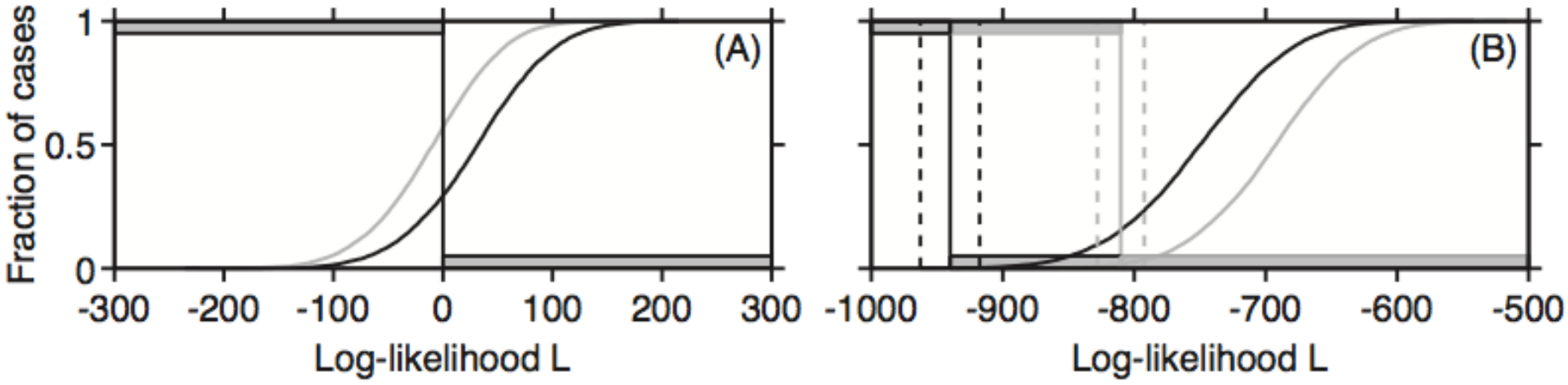


Numerical summaries.

In RELM, consensus was reached that all models would be tested using a certain suite of numerical tests [Jackson and Kagan (1999), Schorlemmer et al. (2007)].

N-Test compares total number of earthquakes for model and observations.
L-Test compares likelihood.
R-Test compares likelihood ratio for two models.

L and N test can be used to see if discrepancies for 1 model are stat. significant.
Simulate, and see if the likelihood for the data is in the middle 95% range.



Schorlemmer et al. 2007

Some drawbacks of these tests were discovered (Schorlemmer et al. 2010).

The N-test and L-test have low power and are typically unable to discern significant lack of fit unless the overall rate of the model fits very poorly.

Further, even when the tests do reject a model, they do not typically indicate where or when the model fits poorly, or how it could be improved.

Model	ℓ_{obs}	γ	N_{obs}	δ
		Mainshock+Aftershock		
A. Helmstetter	−22881.46	0.000	142	**0.000**
B. Kagan	−10765.43	0.008	81	**0.001**
C. Shen	−10265.20	0.002	86	**0.043**
		Daily		
ETAS	−387.69	1.00	85	**0.00**
STEP	−50.43	0.00	83	**0.99**

Clements et al. 2011

Functional summaries.

K-function (Ripley 1979) measures the avg. number of other events within distance *r* of an event, divided by the overall mean rate.
$K(r) \sim \pi r^2$ for a stat. Poisson process,
so centered version is $L(r) = \sqrt{[K(r)/\pi]} - r$.
Useful to compare clustering in data with stat. Poisson process.

Weighted K [Baddeley, Møller and Waagepetersen 2000, Veen and Schoenberg 2005] is useful for comparing degree of clustering in the model with the data. Each point is weighted inversely to $\lambda(t,x,y)$ = the conditional rate of points at (t,x,y), given previous seismicity.

$$K_W(r) = \frac{b}{\int_S \hat{\lambda}_0(\mathbf{x})\, d\mathbf{x}} \sum_i \hat{\lambda}_0(\mathbf{x}_i)^{-1} \sum_{j \neq i} \hat{\lambda}_0(\mathbf{x}_j)^{-1} 1_{\{|\mathbf{x}_j - \mathbf{x}_i| \le r\}}$$

$$K_W(r) \sim N\left(\pi r^2, \frac{2\pi r^2 A}{[\int_S \hat{\lambda}_0(\mathbf{x})\, d\mathbf{x}]^2}\right).$$

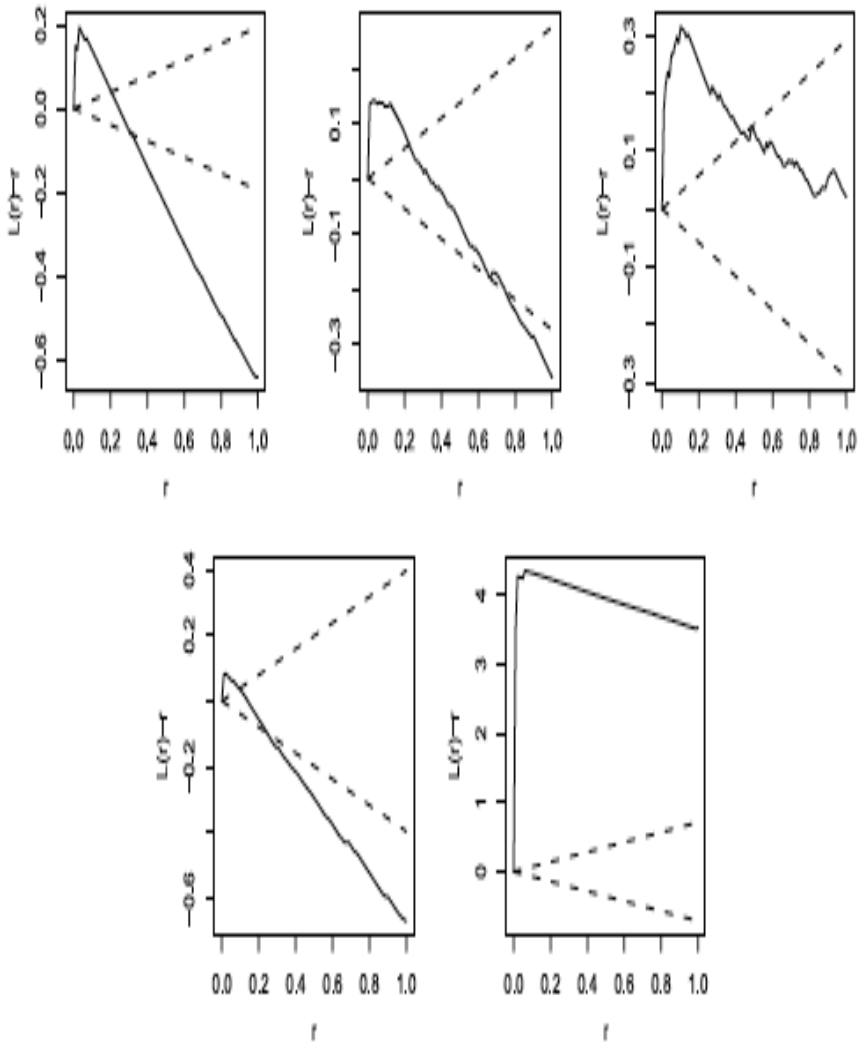


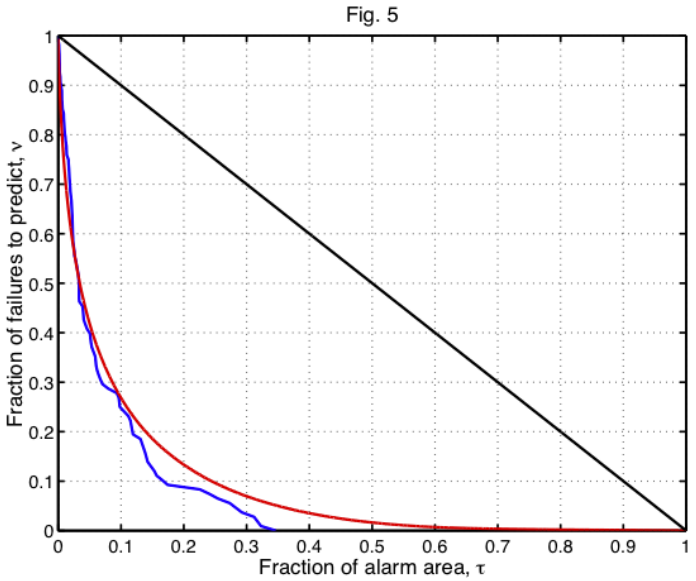
FIG. 5. *Estimated centered weighted L-function (solid curve) and 95% confidence bands (dashed curves). Top-left panel: (a) model A. Top-center panel: (b) model B. Top-right panel: (c) model C. Bottom-left panel: (d) ETAS. Bottom-right panel: (e) STEP.*

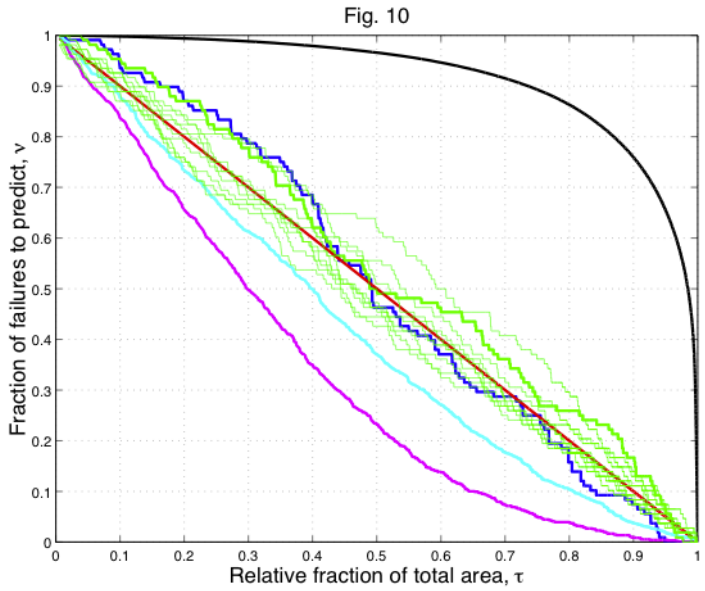
Adelfio and Schoenberg (2009) discuss similar weighted summaries.

Other functional summaries.

Molchan diagrams plot (normalized) number of alarms vs. (normalized) number of false negatives (failures to predict). (Molchan 1990; Molchan 1997; Zaliapin & Molchan 2004; Kagan 2009).

Similar to ROC curves (Swets 1973).





Problems:

-- Must focus near axes.

[consider relative to given model (Kagan 2009)]

-- Difficult to see *where* model fits poorly.

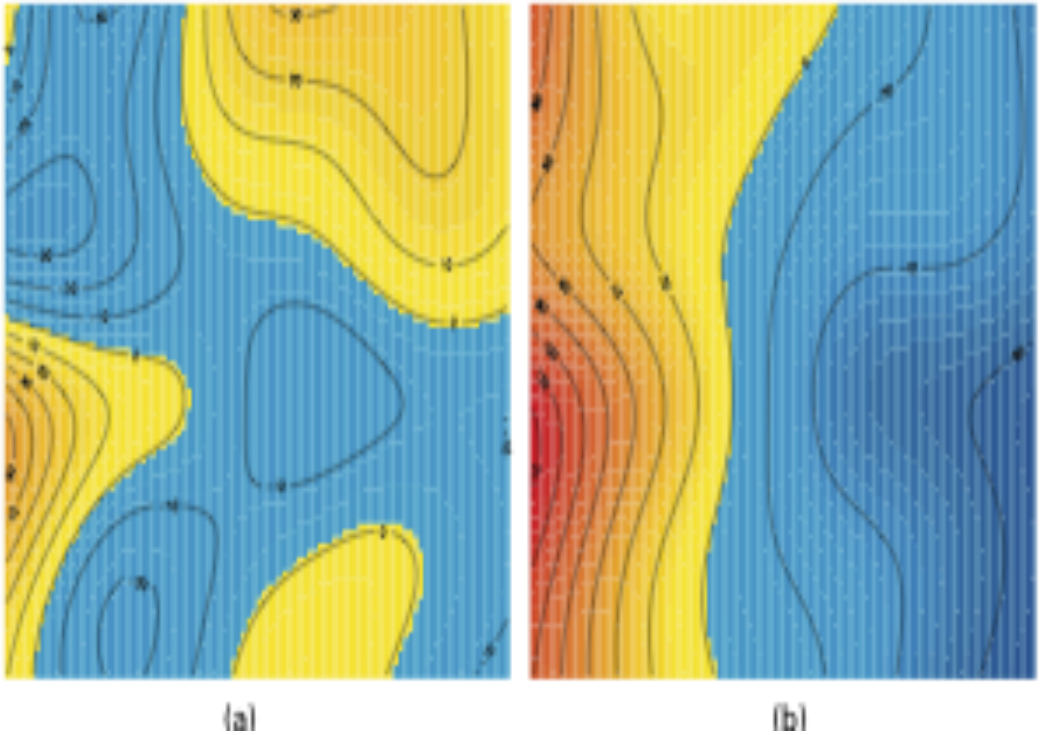
Pixel-based residuals.

Compare $N(A_i)$ with $\int_A \lambda(t, x)\, dt\, dx$, on pixels A_i.

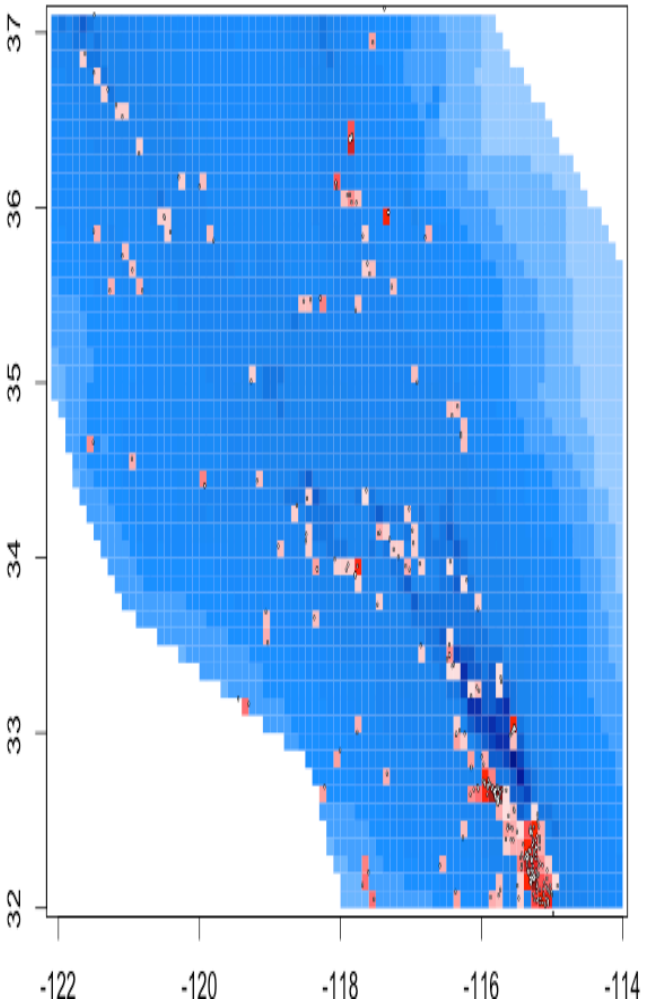
(Baddeley, Turner, Møller, Hazelton, 2005)

Problems:

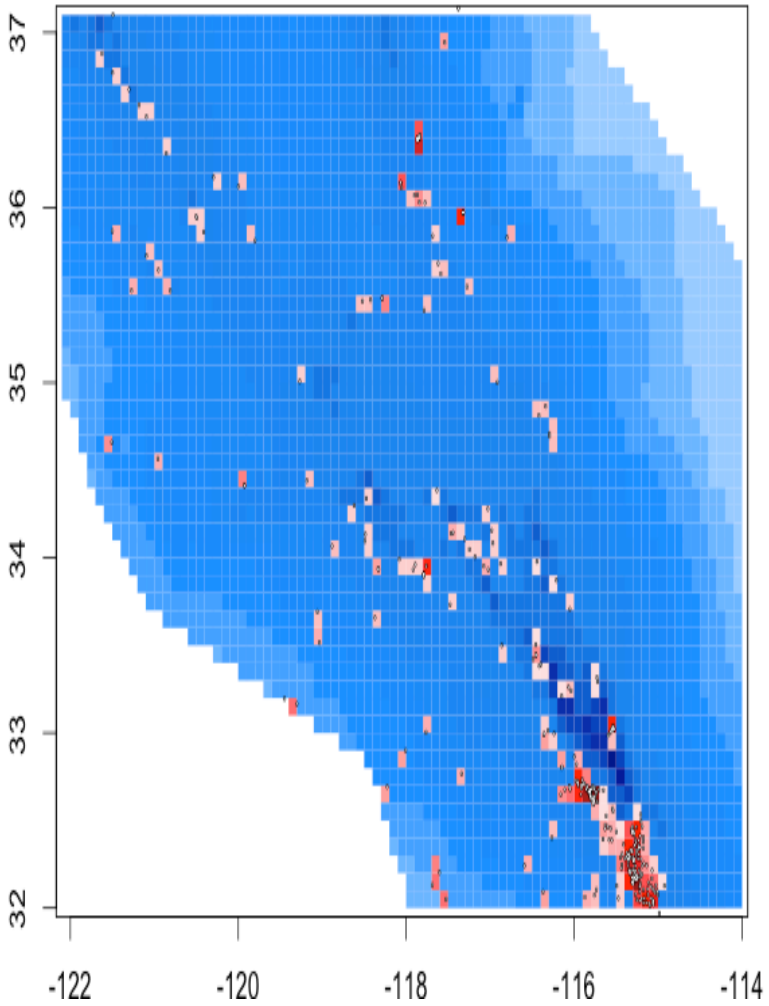
* If pixels are large, lose power.
* If pixels are small, residuals are mostly ~ 0,1.
  -- non-normality after standardization.
* Smoothing reveals only gross features.



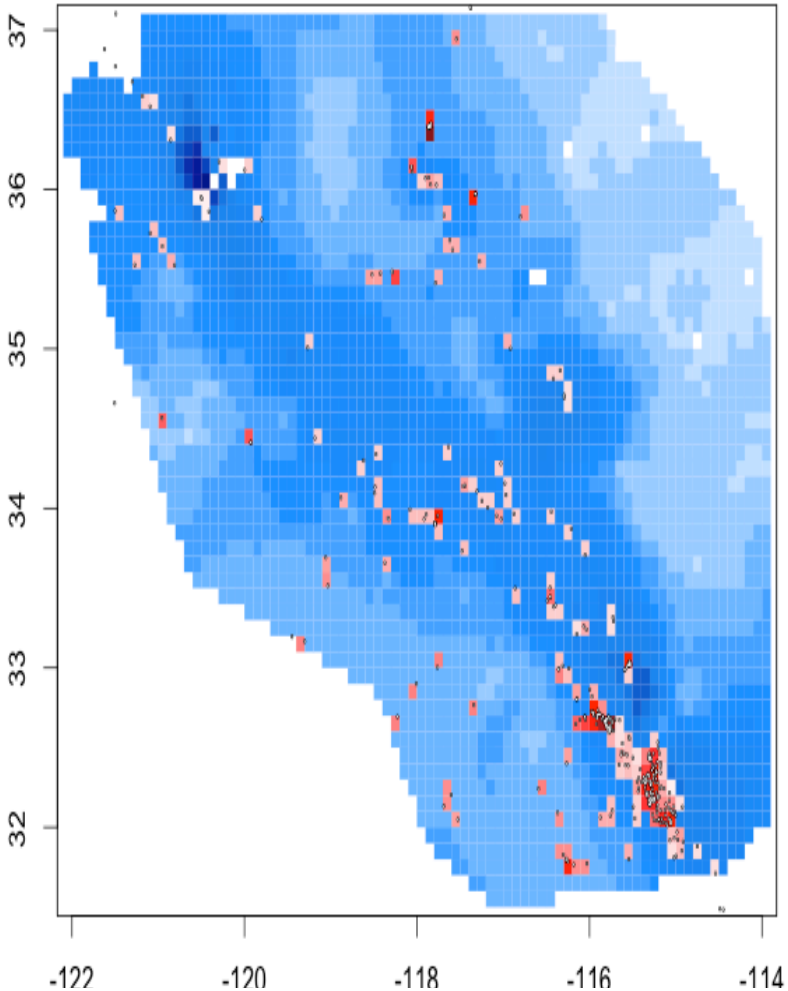
Pearson residuals for Model B:



Pearson residuals for Model B:



Pearson residuals for Model C:



How can you see how well the model fits?

a) Deviance residuals

b) Superthinned residuals

c) Voronoi residuals.

-- Given two competing models, can consider the difference between residuals, number of observed fires – number expected, over each pixel.

Problem: Hard to interpret. If difference = 3, is this because model A overestimated by 3? Or because model B underestimated by 3? Or because model A overestimated by 1 and model B underestimated by 2?

-- Better: consider difference between *log-likelihoods*, in each pixel. The result may be called ***deviance residuals*** (Clements et al. 2011), ~ resids from gen. linear models.

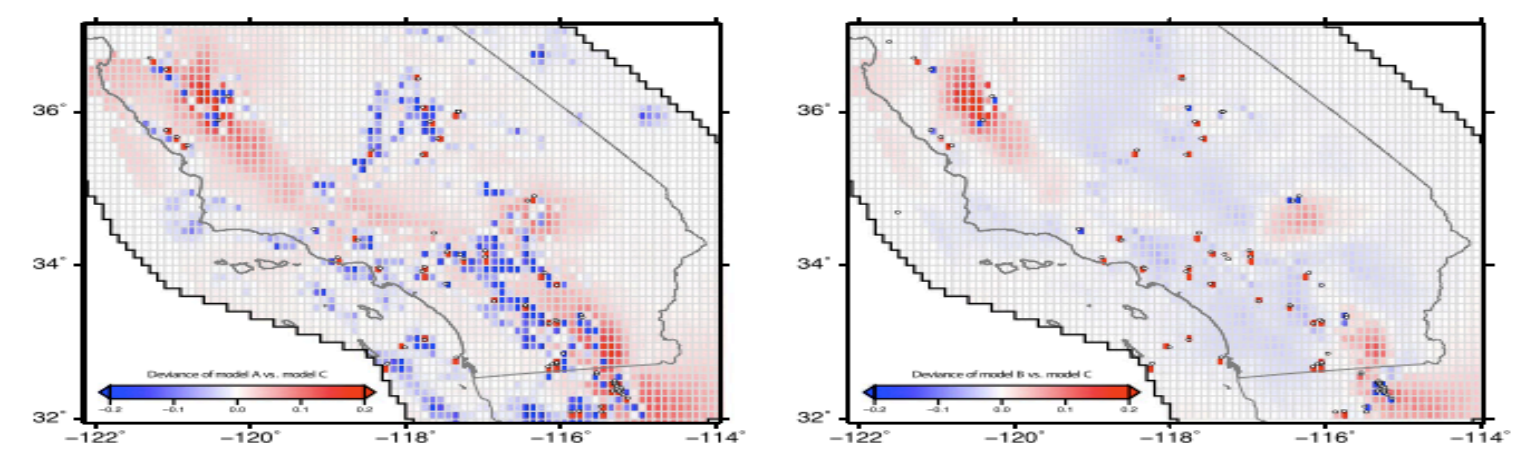


FIG. 4. *Left panel* (a): *deviance residuals for model A versus C. Sum of deviance residuals is* 86.427. *Right panel* (b): *deviance residuals for model B versus C. Sum of deviance residuals is* −7.468.

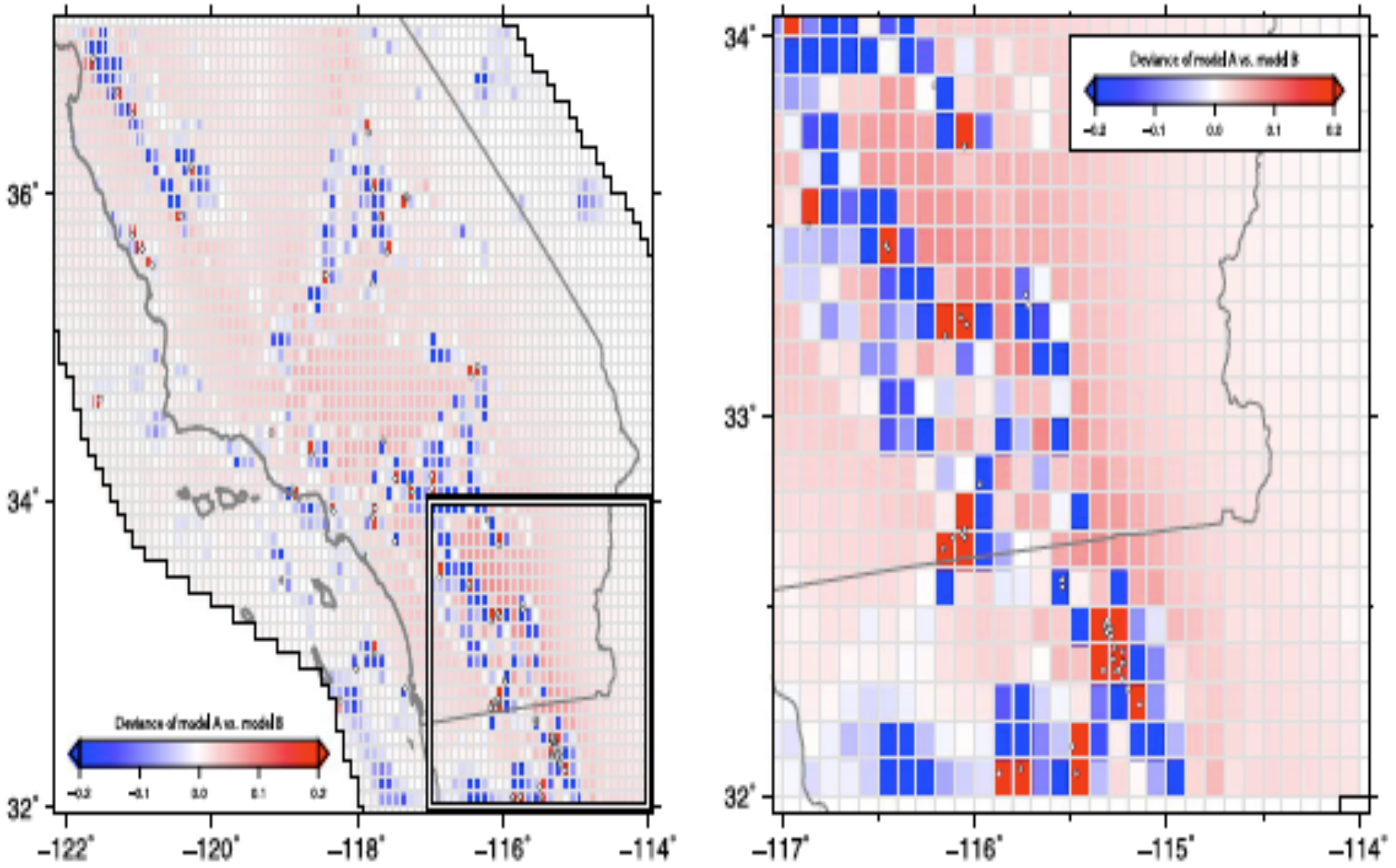
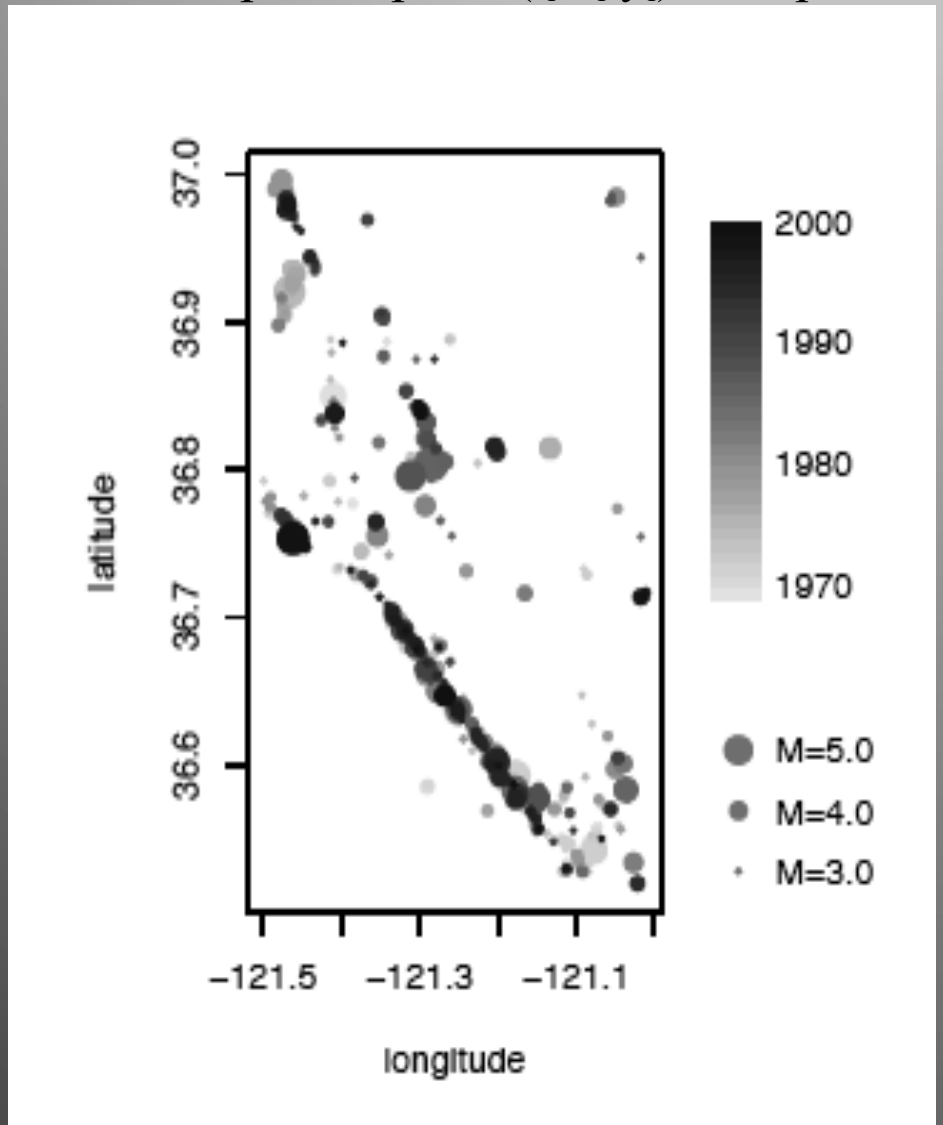


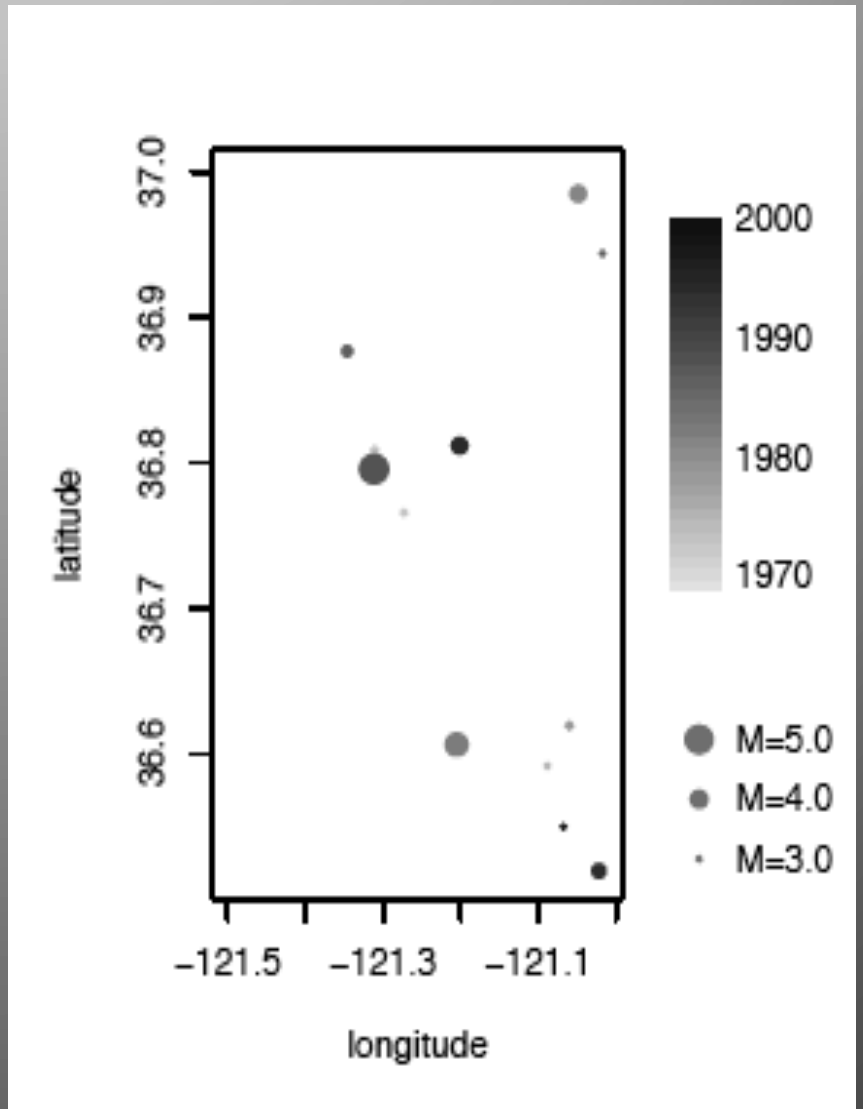
FIG. 3. *Left panel* (a): *deviance residuals for model A versus B. Sum of deviance residuals is* 84.393. *Right panel* (b): *close-up of deviance residuals for model A versus B near the Imperial fault.*

Superthinning.

Thinning: Suppose $\inf \lambda(t_i, x_i, y_i) = b$.

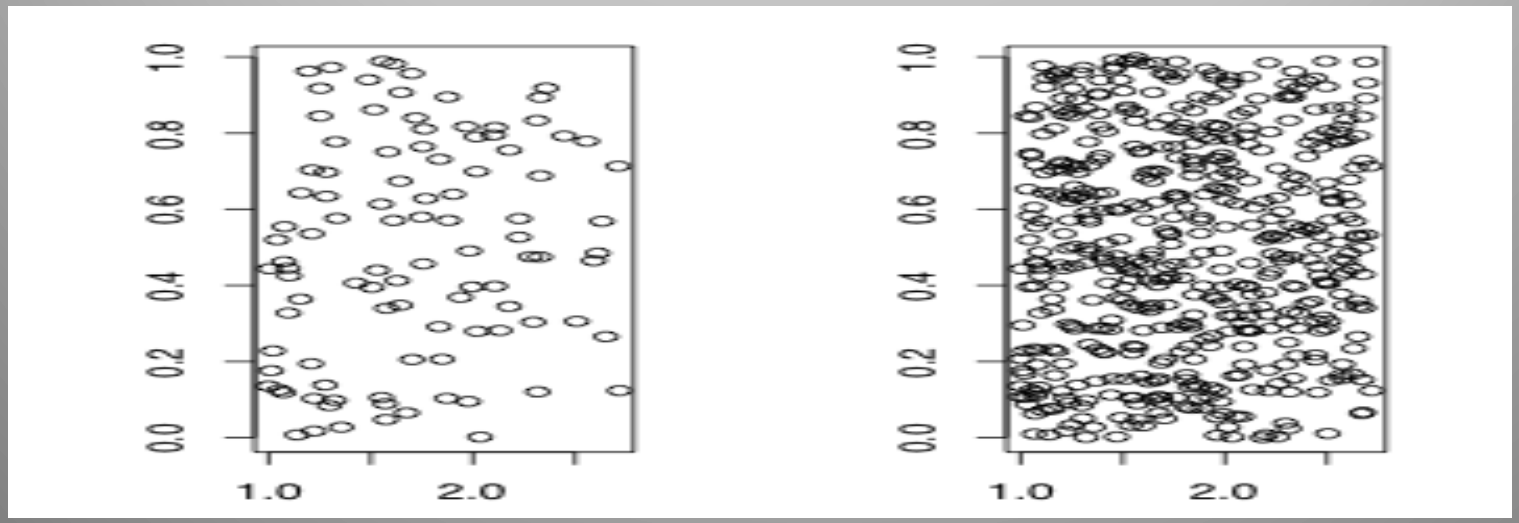
Keep each point (t_i,x_i,y_i) with probability $b / \lambda(t_i,x_i,y_i)$.





Superposition: Suppose sup $\lambda(t, x, y) = c$.

Superpose N with a simulated Poisson process of rate $c - \lambda(t, x, y)$.



Problems with thinning and superposition:

Thinning: Low power. If $b = \inf \lambda(t_i,x_i,y_i)$ is small, will end up with *very* few points.

Superposition: Low power if $c = \sup \lambda(t_i,x_i,y_i)$ is large: most of the residual points will be simulated.

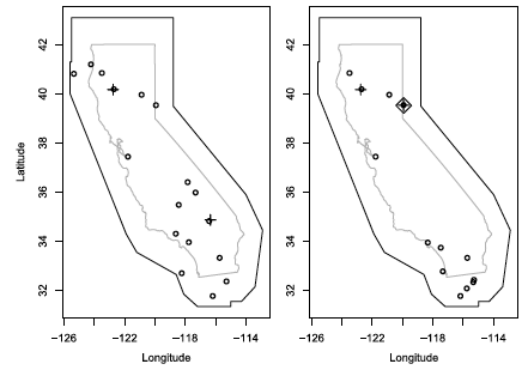
Superthinning: superpose where $\lambda(t,x,y) < c$, and thin where $\lambda(t_i,x_i,y_i) > c$.

Superthinning (Clements et al., 2012)

Choose some number c ~ mean($\hat{\lambda}$).

Superpose: where $\hat{\lambda}$ (t ,x,y) < c, add in points of a simulated Poisson process of rate c - ($t\hat{\lambda}$, y) .

Thin: where $\hat{\lambda}(t_i,x_i,y_i) > c$, keep each point (t_i,x_i,y_i) with prob. c / $\hat{\lambda}(t_i,x_i,y_i)$.



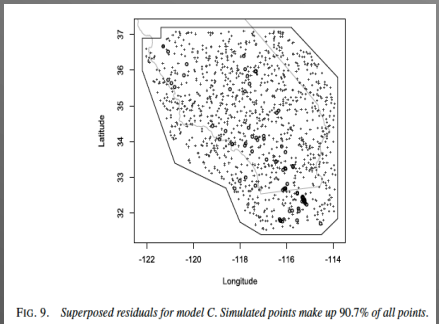


FIG. 9. *Superposed residuals for model C. Simulated points make up 90.7% of all points.*

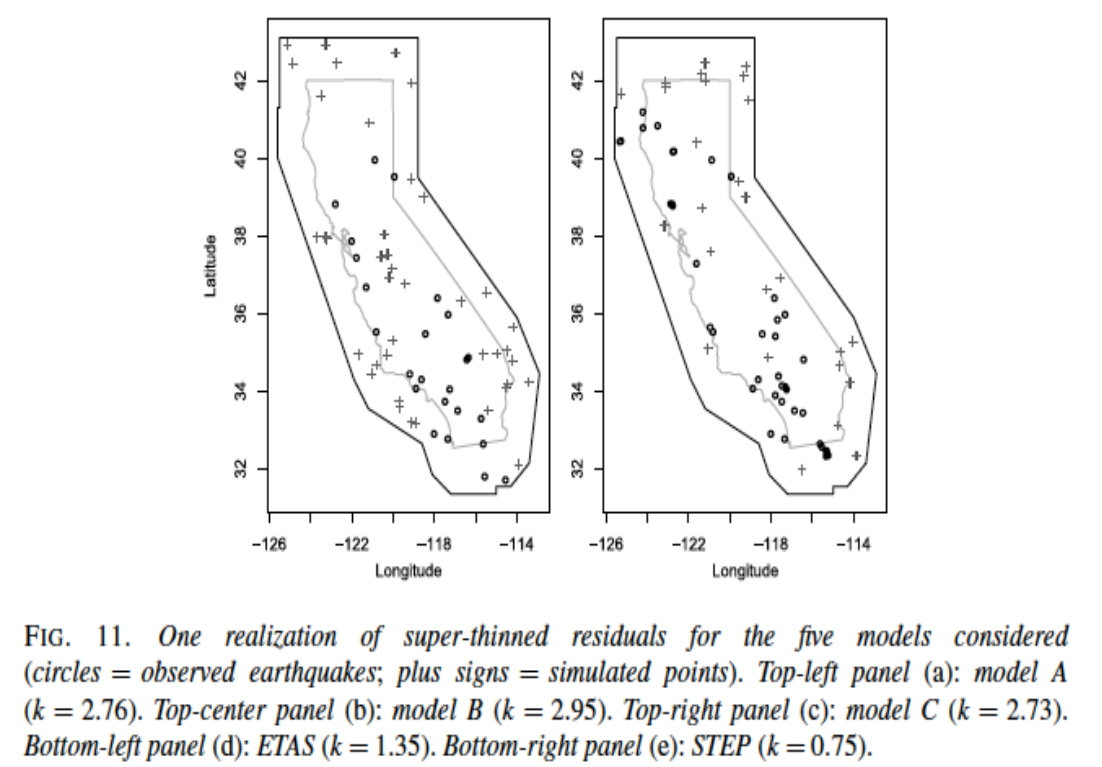
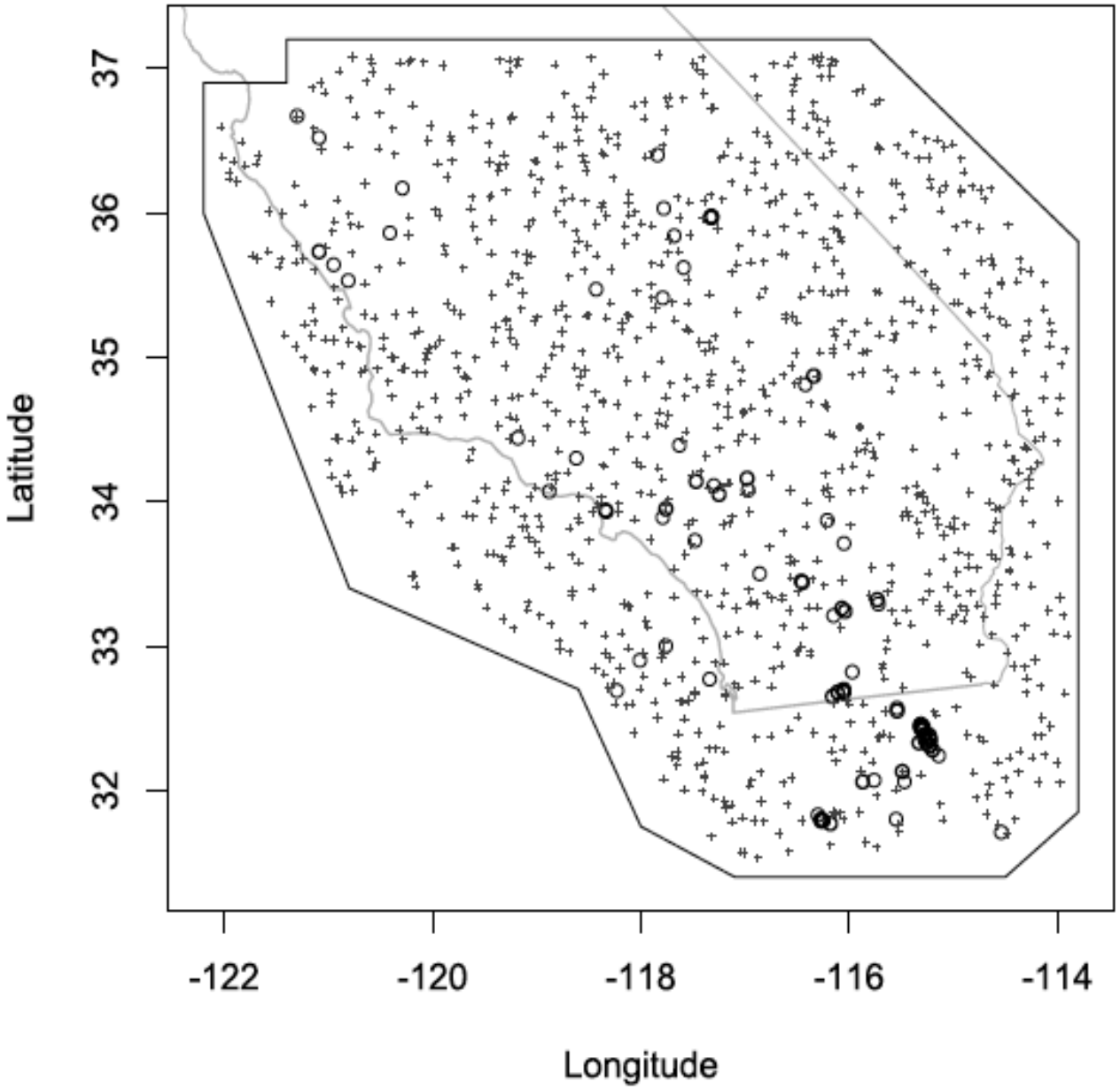


FIG. 11. *One realization of super-thinned residuals for the five models considered (circles = observed earthquakes; plus signs = simulated points). Top-left panel* (a): *model A* ($k = 2.76$). *Top-center panel* (b): *model B* ($k = 2.95$). *Top-right panel* (c): *model C* ($k = 2.73$). *Bottom-left panel* (d): *ETAS* ($k = 1.35$). *Bottom-right panel* (e): *STEP* ($k = 0.75$).



Superposed residuals for model C. Simulated points make up 90.7% *of all points.*

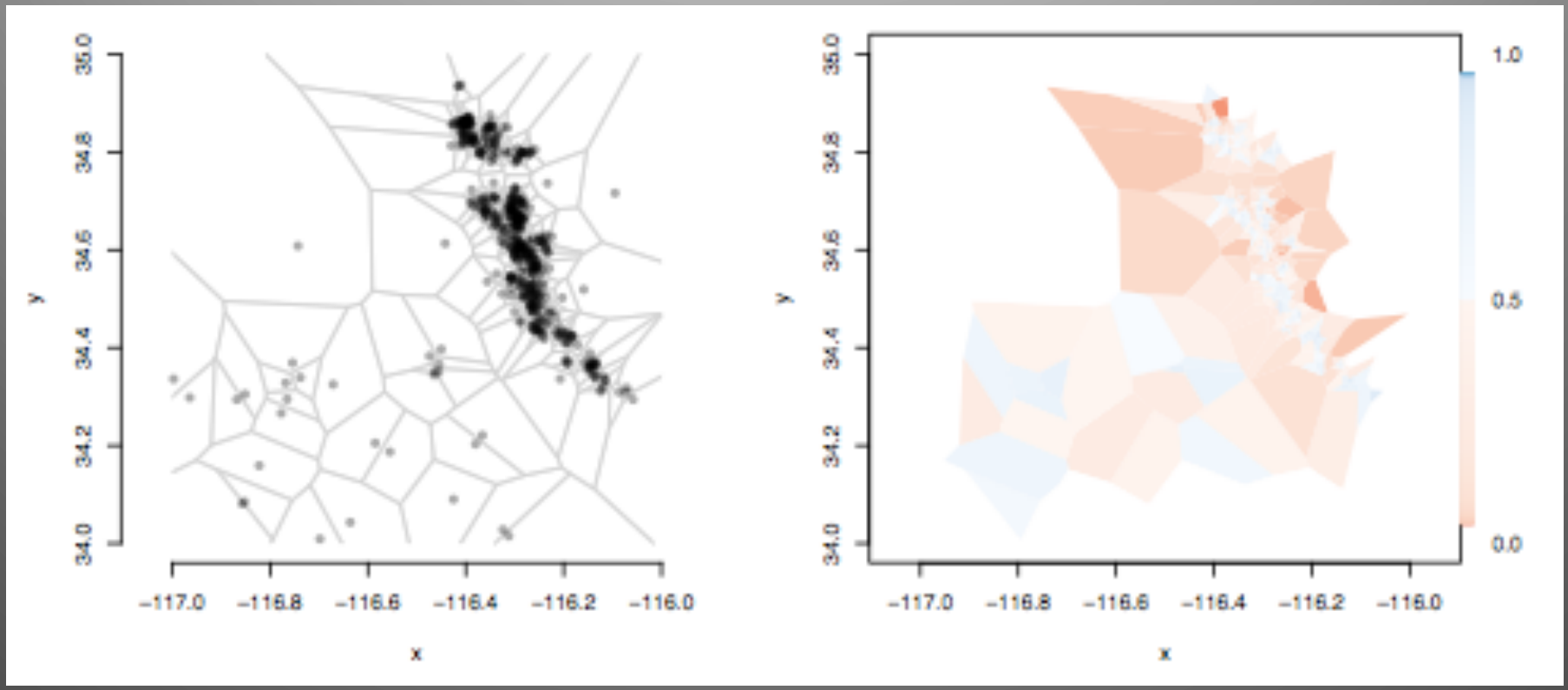
c) Voronoi residuals (Bray et al. 2013)

A Voronoi tessellation divides a space into cells Ci where Ci contains all locations closer to event i than any other observed event.

Within each cell, calculate residuals

$$r \sim 1 - X\,; \quad X \sim \Gamma(3.569, 3.569)$$ (Tanemura 2003)

$$\begin{aligned} \hat{r}_i &:= 1 - \int_{C_i} \hat{\lambda}\, \mathrm{d}\mu \\ &= 1 - |C_i|\bar{\lambda}, \end{aligned}$$

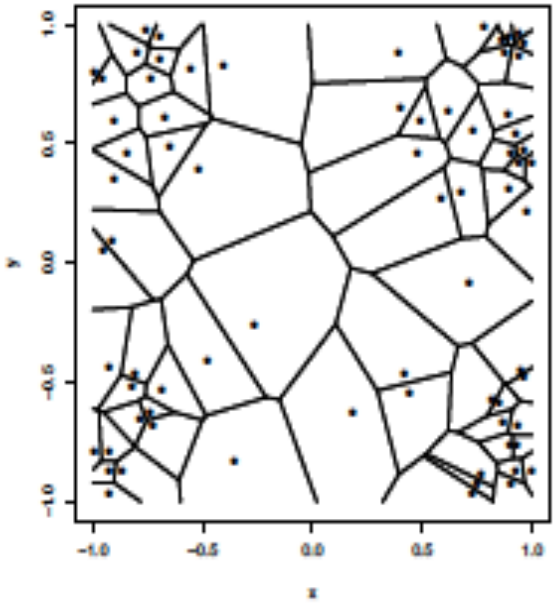


spatially adaptive and nonparametric.

Voronoi Residuals

Voronoi Tessellation: (Okabe, 2000) a partitioning of S into n convex polygons (tiles)

$$D_i = \{x \in X : ||x - x_i|| \leq ||x - x_j||, \forall j \neq i\}$$

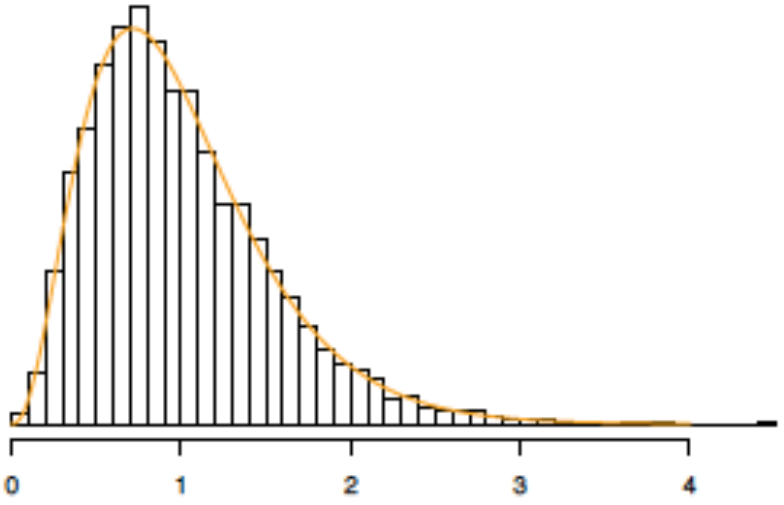


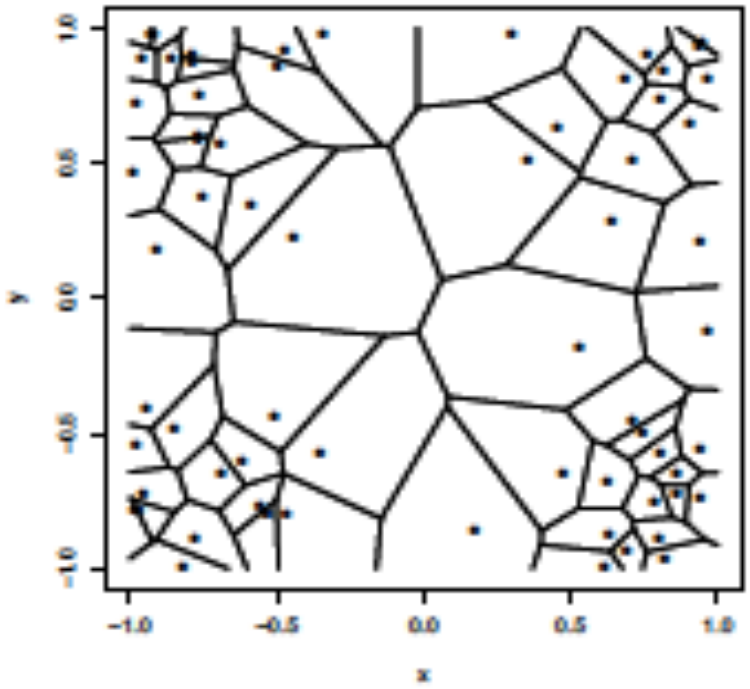
Voronoi Residual:

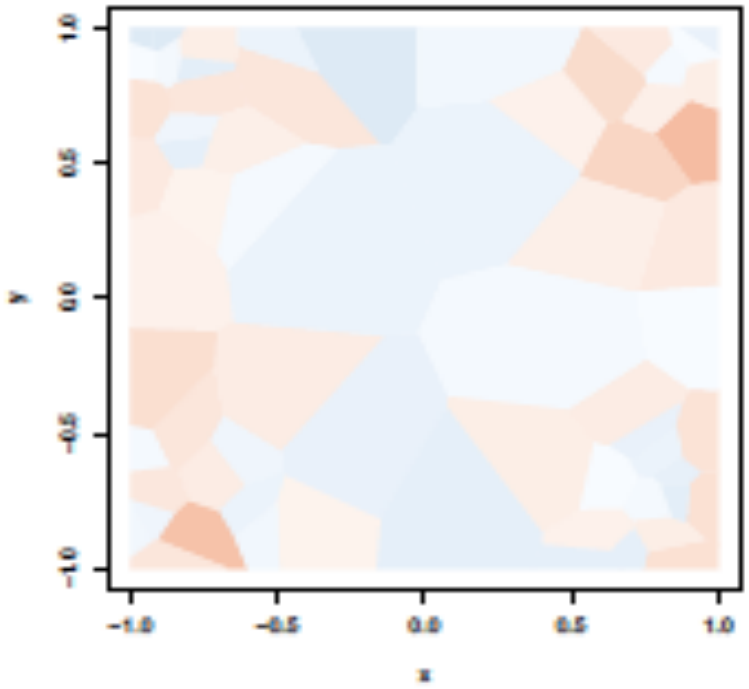
$$\hat{r}_i = \frac{N(D_i) - \int_{D_i} \hat{\lambda}(x_i)\,\mathrm{d}x_i}{SE\left(\int_{D_i} \hat{\lambda}(x_i)\,\mathrm{d}x_i\right)} = \frac{1 - |D_i|\bar{\lambda}}{SE\left(|D_i|\bar{\lambda}\right)}$$

Properties of Voronoi Residuals

- Non-parametric
- Spatially adaptive
- $|D|\bar{\lambda} \sim Gamma(3.5, 3.5)$ for homogeneous (Tanemura 2003)
- Approximately Gamma for inhomogeneous (Barr and Schoenberg 2010, Barr and Diez in progress)

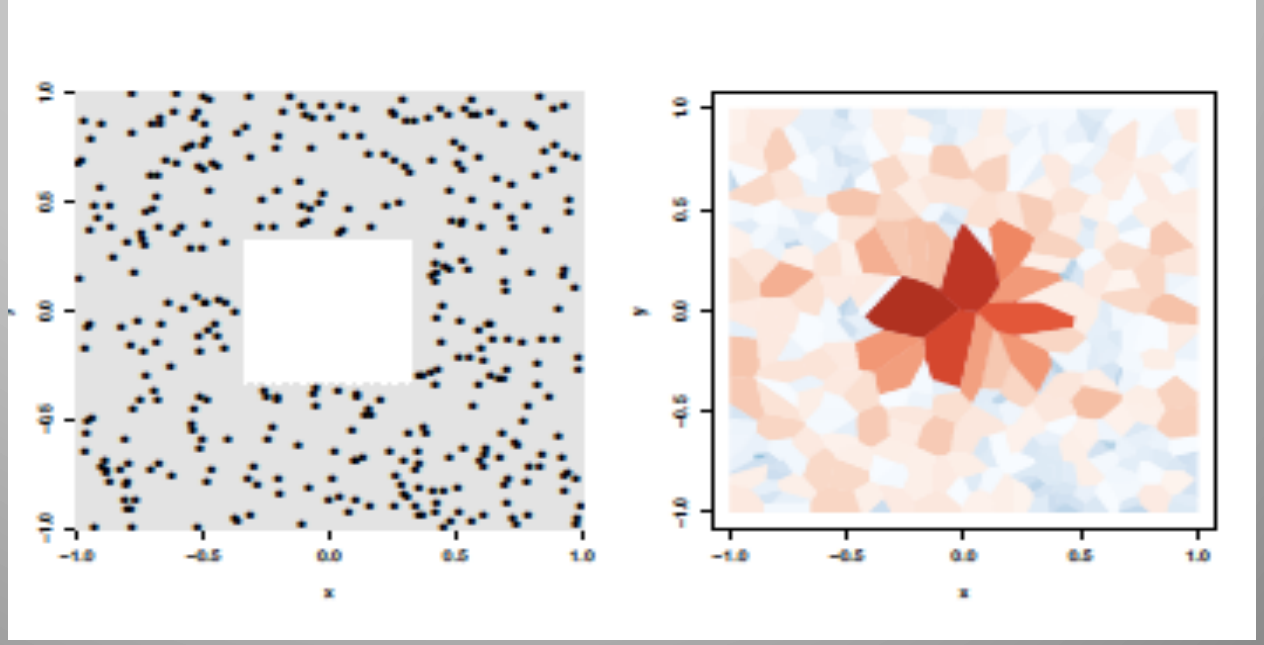




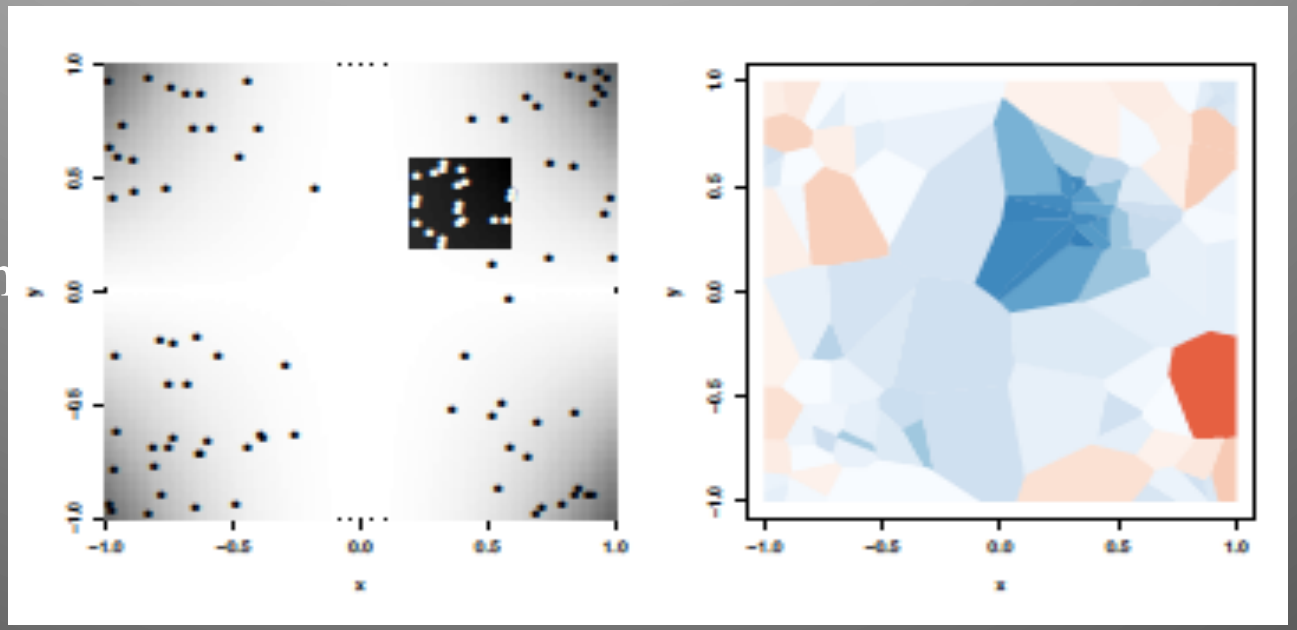


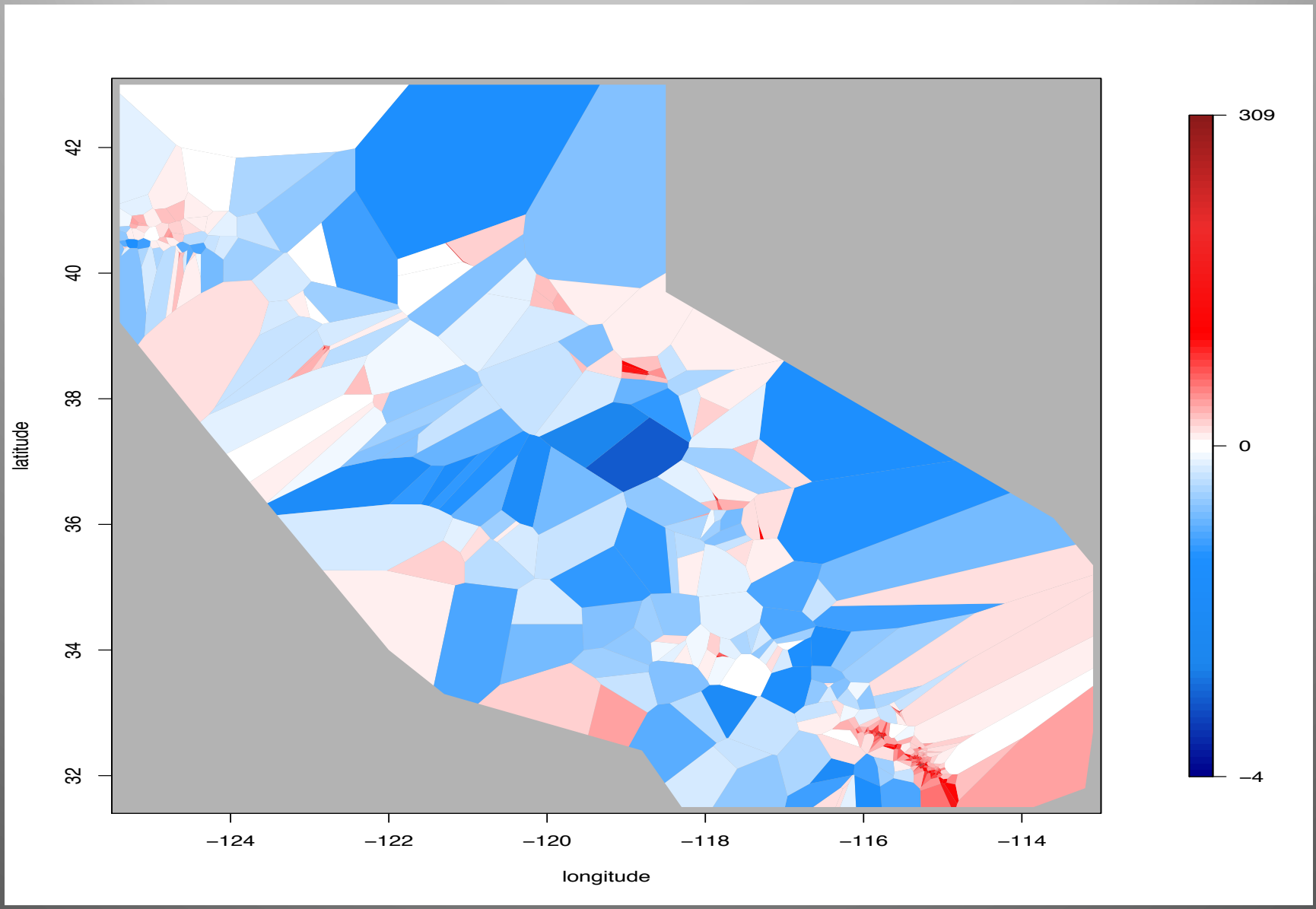
- generating model = fitted model = $100x^2|y|$
- less-skewed residuals with narrower range

overprediction



underprediction

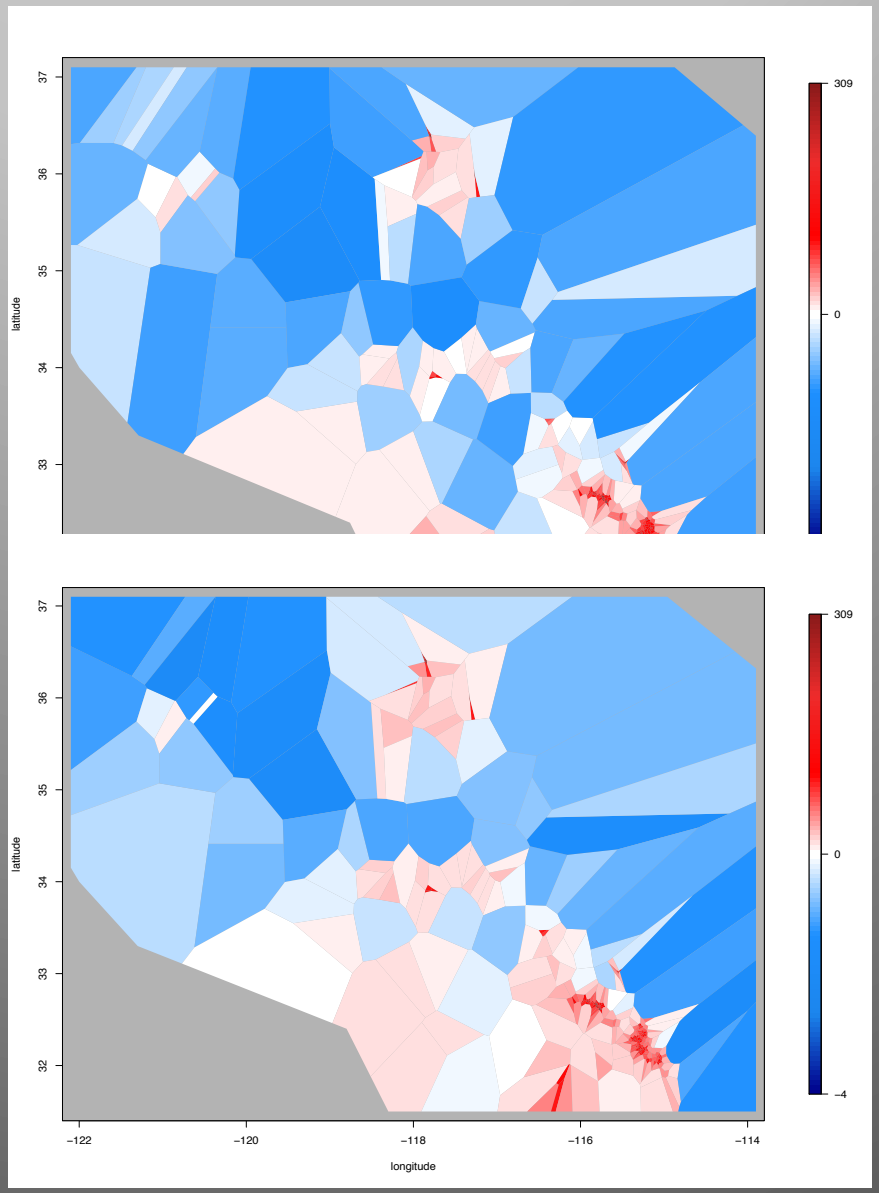



latitude
42
40
38
36
34
32
−124
−122
−120
−118
−116
−114
longitude
309
0
−4

Model A

Model B

Model C



Conclusions.

* Likelihood methods could facilitate evaluation of probabilistic spread models.

* Deviance residuals are useful for comparing models on grid cells.

* Superthinned residuals do not rely on a grid and can be useful to highlight where a model overpredicts or underpredicts.

* Voronoi residuals use an automatically adaptive grid and are powerful for both comparison and to see where a particular model over or underpredicts.

Forecasting global temperature data.

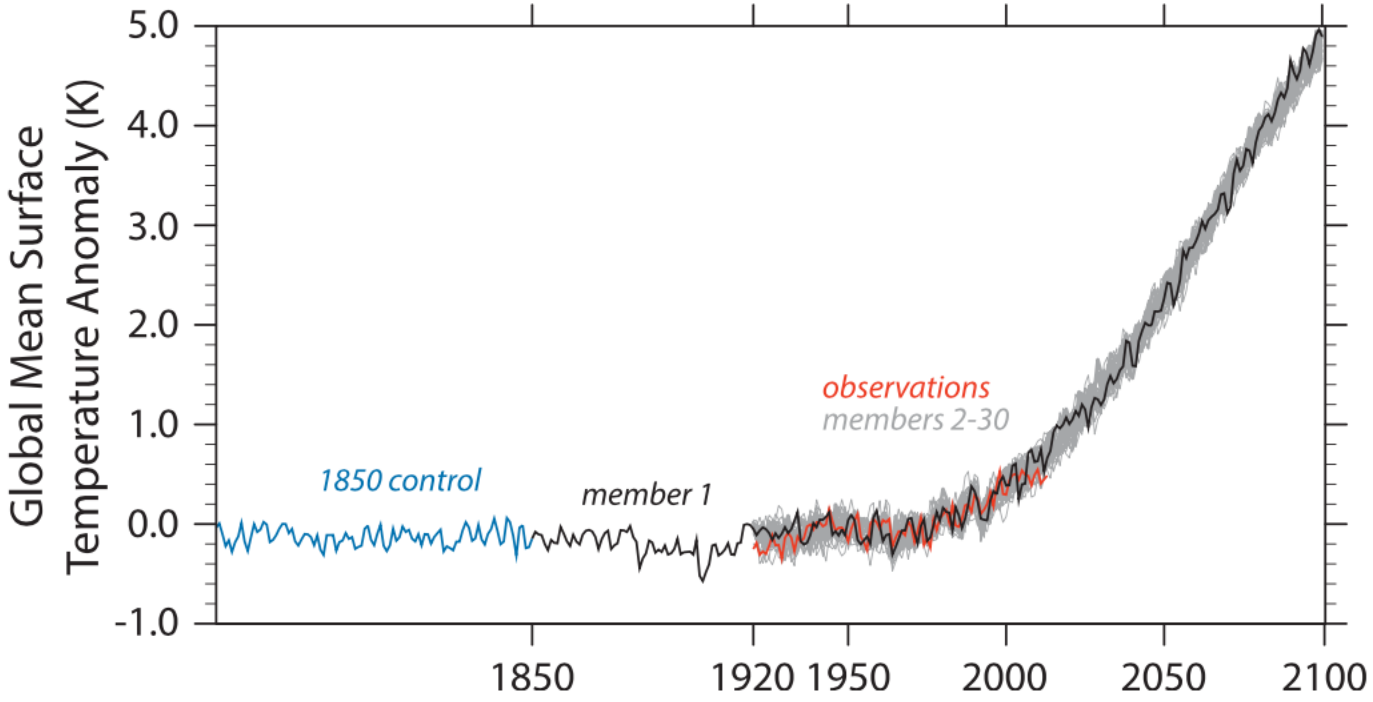


FIG. 2. Global surface temperature anomaly (1961–90 base period) for the 1850 control, individual ensemble members, and observations (HadCRUT4; Morice et al. 2012).