Spatial correlation of crop yields

Peter McCullagh and David Clifford
University of Chicago

Minnesota, March 2006

www.stat.uchicago.edu/~pmcc/TR.html

Spatial correlation of crop yields

Empirical studies:

Mercer and Hall (1911) (cereals)

Batchelor and Reed (1918) fruit yields

Eden and Fisher (1929) cereals and root crops

Fairfield Smith (1938) cereals and other crops

Estimation of effects:

Eden and Fisher: blocking

Papadakis (1937): local control

Besag (1974),

Besag and Higdon (1999) MRF models,...

Bartlett (1978) MRF models

Zimmermann and Harville (1991) (Gaussian RFs)

Cullis and Gleeson (1991) (Product models)

Issues and distinctions

- Systematic or identifiable effects
  - (variety, treatment, soil type, shade,...)
- Anthropogenic effects:
  - ploughing, drilling, planting, harvesting, drainage,...
- Non-anthropogenic spatial correlation

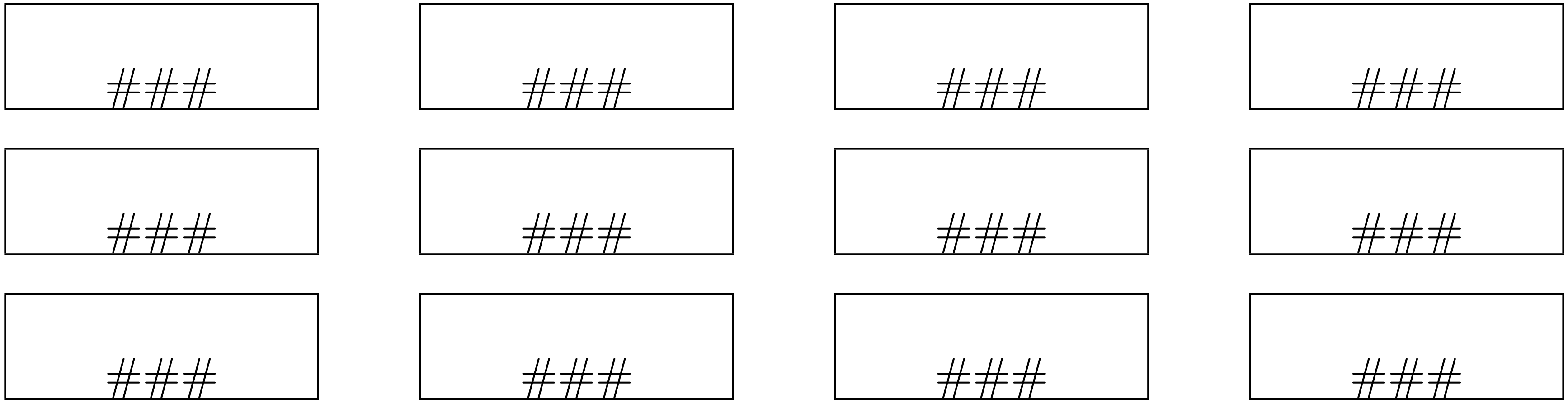
Non-anthropogenic spatial correlation

- anthropogenic effects usually associated with rows/drills
- non-anthropogenic effects stationary and isotropic???

Nature of the isotropic component

Source/cause of spatial correlation?

Typical field layout (cereals and root crops)



Guard strips:

usually present in variety trials

may be fallow or harvested and not recorded

may not be present in uniformity trials

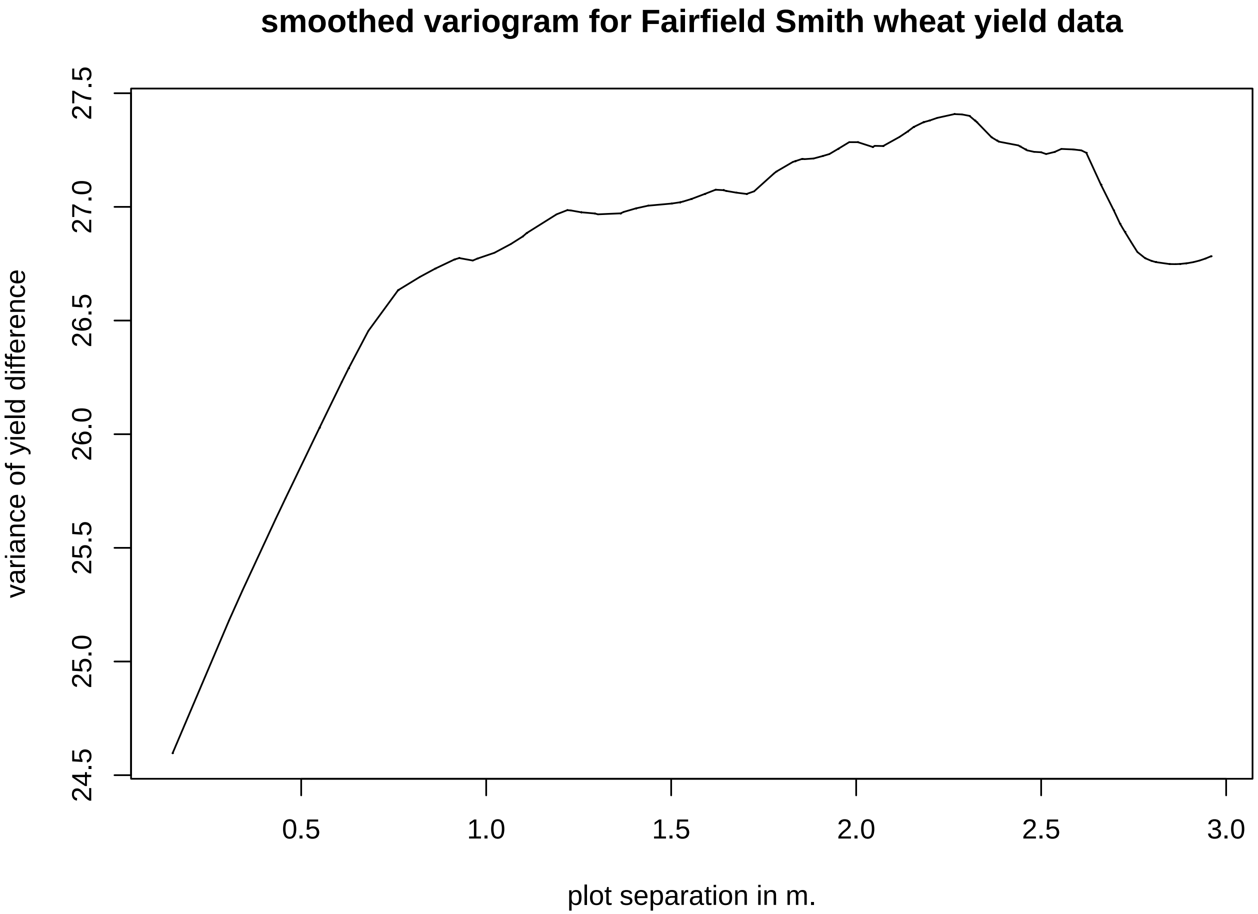
Plot size: highly variable,...

Number of plots 50-1500

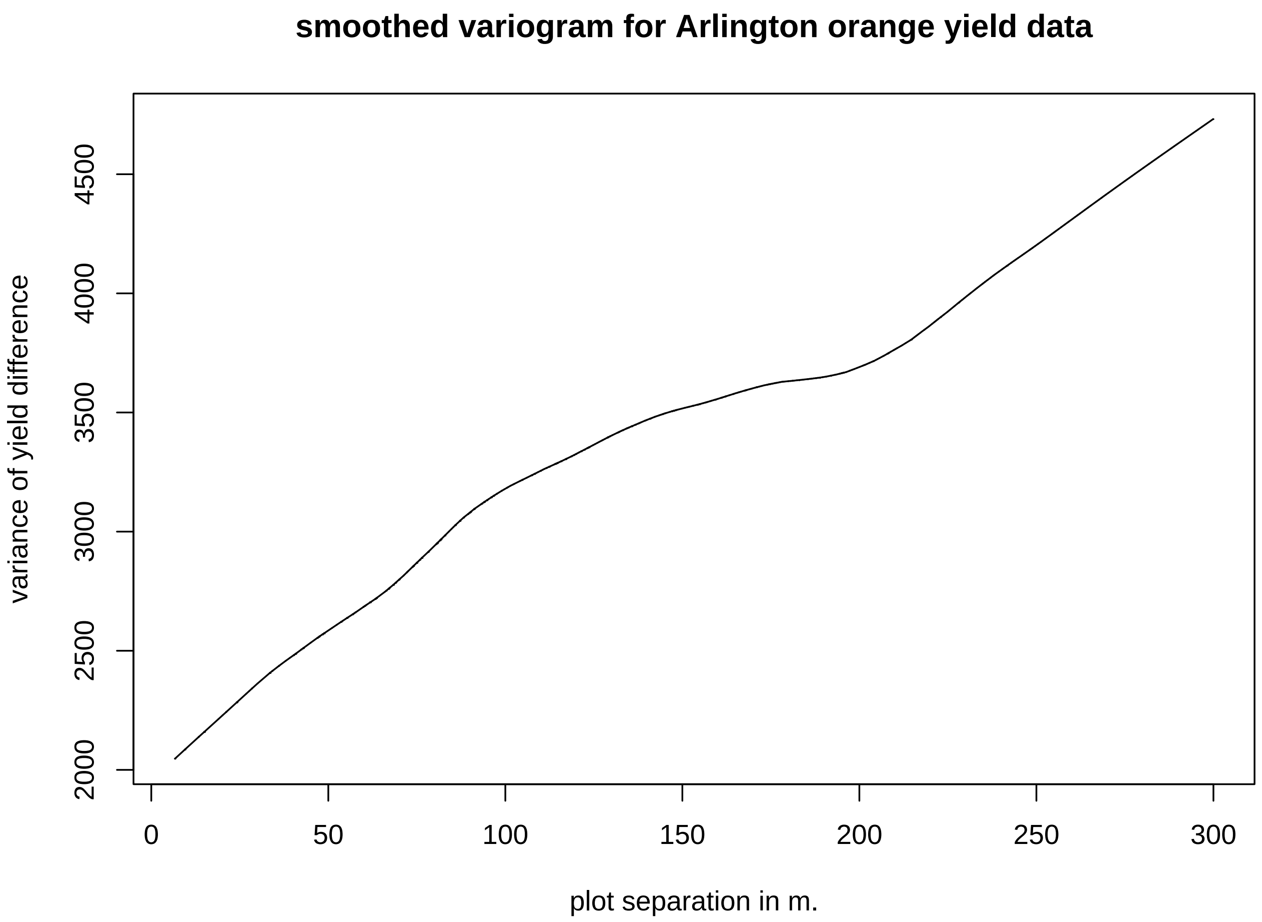
Observation is yield on each plot (possibly bivariate)

Need to know the geometry

Fairfield Smith wheat yield data (Australian)



Arlington orange grove data (1000 trees, California)



Mathematical symmetries: gedankenexperimenten (E. Mach)

Exchangeability in regression models

Two sets of units $\{u_1, \ldots, u_n\}$ and $\{u'_1, \ldots, u'_n\}$ such that

$x(u_j) = x(u'_j)$ have the same joint distribution

Stationarity of stochastic processes:

Temporal stationarity

Planar stationarity

Isotropy of stochastic processes

Temporal isotropy (reversibility)

Planar isotropy

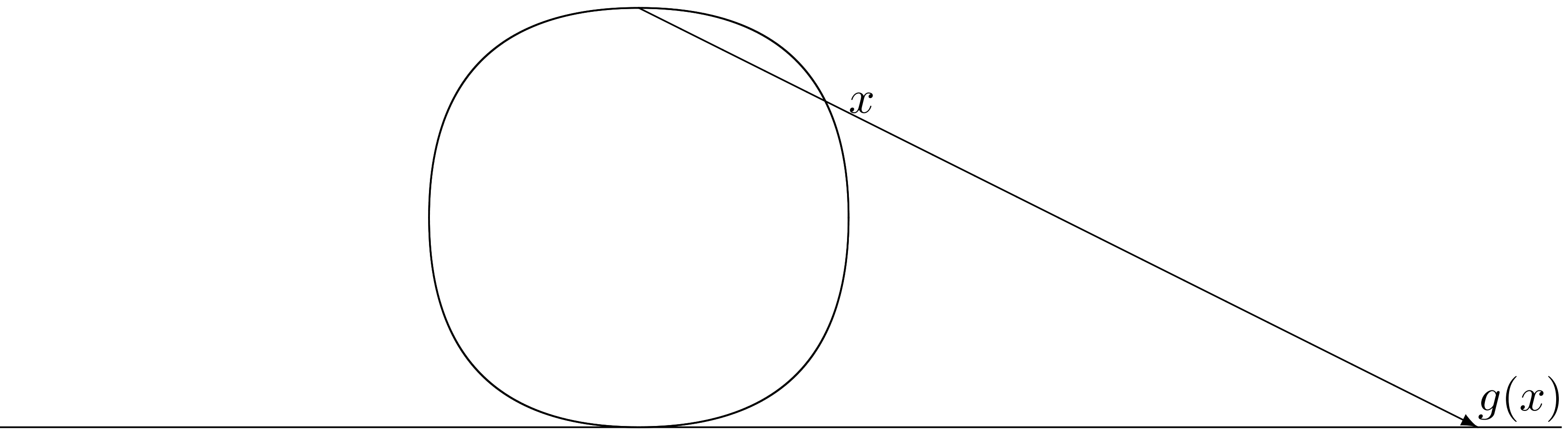
Spherical processes

Spherical stationarity

Mathematical symmetry versus physical symmetry

Spherical processes and planar processes

Stereographic projection



Preserves local Euclidean properties of space – meaning?

Planar process $Y \mapsto$ spherical process W

Planar contrasts $\rightarrow$ spherical contrasts – why?

Group of rigid motions, planar + spherical

Conformal process stationary, isotropic and ...

Conformal transformations as planar transformations

$$g(x) = \frac{ax + b}{cx + d}$$

(complex numbers: $ad - bc \neq 0$; $x = x_1 + ix_2$)

Complete list of conformally invariant Gaussian processes

(i) Random constant $K(x, x') = \sigma^2$

(ii) de Wijs : $K(x, x') = -\sigma^2 \log |x - x'|$ on non-atomic contrasts

(iii) Planar white noise with intensity

$$\lambda(x) = \left(\theta_0 - \bar{\theta}_1 x - \theta_1 \bar{x} + \theta_2 |x|^2\right)^2$$

is closed but not invariant

Thesis concerning crop yields

Non-anthropogenic variation is conformally invariant

stationarity, isotropy plus spherical rotations

Suggested model

$$K(x, x') = \sigma_0^2 \delta_{x-x'} - \sigma_1^2 \log|x - x'|$$

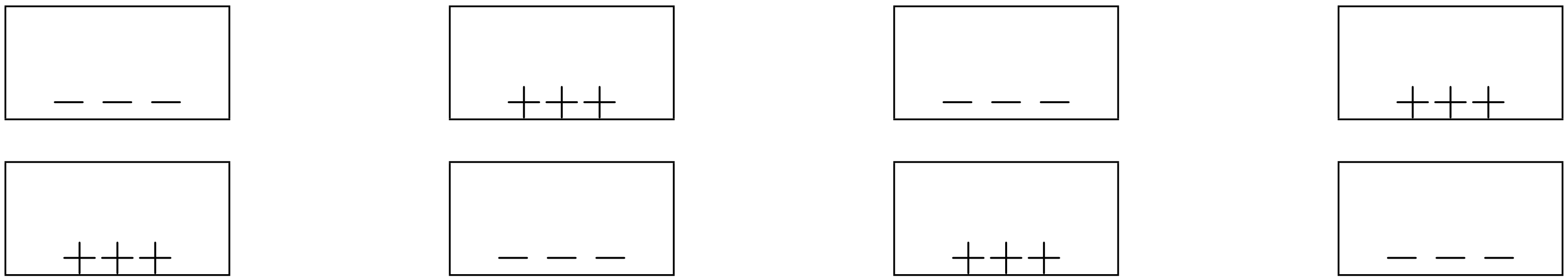
implies logarithmic rate of decrease of correlations
— or inverse square law for simple contrasts.

Are crop yields isotropic?

Do correlations decrease logarithmically?

What is the evidence?

The inverse square law for differences



Yield Differences $D_1 = Y_{11} - Y_{12}$, $D_2 = Y_{21} - Y_{22}, \ldots$

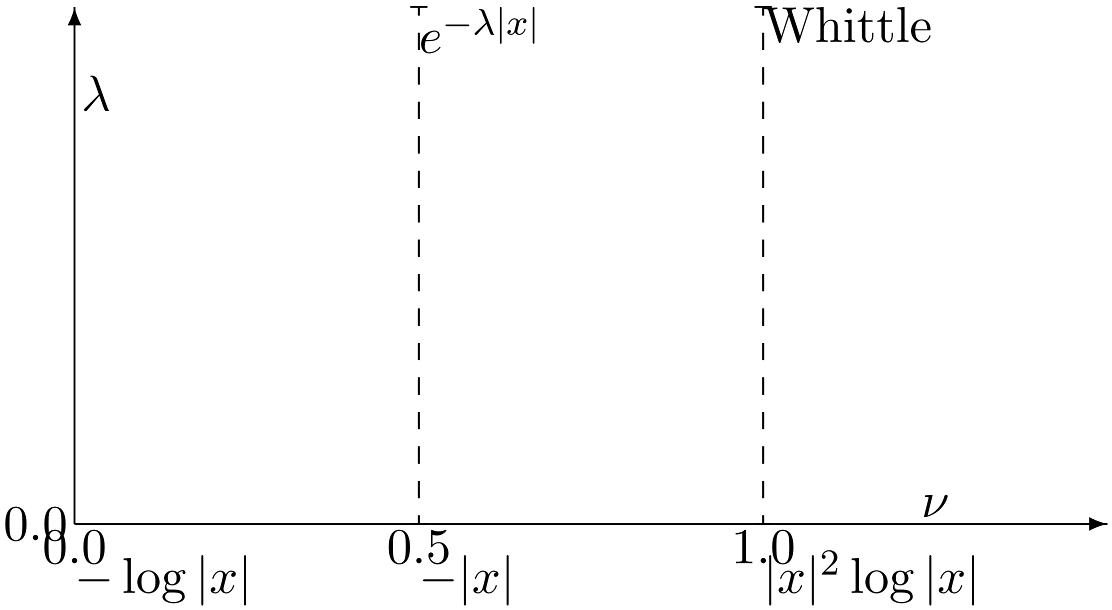
Variance $\text{cov}(D_1, D_r) \simeq k/r^2$ for large r

Coefficient depends on geometric configuration

The Matérn model and parameter space

$$\mathrm{cov}(Y(x), Y(x')) = (\lambda|x - x'|)^{\nu} K_{\nu}(\lambda|x - x'|)$$

Range $1/\lambda$; smoothness parameter $= \nu$.



Matern includes de Wijs, power family, Bessel_0, exponential, TPS, ...

Data from classical uniformity trials

Crop	Plot size	rows	cols	Spacing	Reference
Sawyer field data					
wheat 1925	66' x 66'	7	7		Rothamsted Report 1926
straw 1925	66' x 66'	7	7		Rothamsted Report 1926
swede root '26	66' x 66'	7	7		Rothamsted Report 1926
swede tops '26	66' x 66'	7	7		Rothamsted Report 1926
Fruit and nut yields:					
Arlington O I	22' x 22'	20	25		Batchelor \& Reed 1918
Arlington O II	22' x 22'	20	25		Batchelor \& Reed 1918
Antelope O	22' x 22'	15	33		
Valencia O	22' x 22'	12	30		
Eureka lemons	24' x 24'	14	26		
walnuts	25' x 25'	10	28	(25', 25')	
apples	16' x 16'	8	28	(0', 14')	
Cereals:					
Spring barley	5' x 14'	28,	7		Kempton \& Howes (1981)
wheat grain	11' x 8'	20	25	variable	Mercer \& Hall (1911)
wheat straw	11' x 8'	20	25	variable	Mercer \& Hall (1911)
Winter wheat	5' x 15'	52	7	(3', 70')	Besag \& Kempton 1986
F-S wheat	6'' x 12''	30	36		Fairfield Smith 1938
Federation W	1' x 15'	125	12		Wiebe (1935)
Long Hoos oats	18' x 60'	12	8	(3', 3')	Rothamsted report 1926
Long Hoos straw	18' x 60'	12	8	(3', 3')	Rothamsted report 1926
Root crops:					
Potatoes	18' x 60'	9	9	(5', 5')	Eden \& Fisher 1929
Mangold roots	7' x 30'	20	10	(7', 7')	Mercer \& Hall 1911
Mangold tops	7' x 30'	20	10	(7', 7')	Mercer \& Hall 1911
Brassicas:					
Brussels Sp	27' x 39'	6	8	(3', 3')	Rothamsted Report 1934

Residual Likelihood function (REML)

$$\mathrm{cov}(Y(x), Y(x')) = \sigma_0^2 \delta_{x-x'} + \sigma_1^2 K(x, x')$$

$$\mathrm{cov}(Y(A), Y(A')) = \sigma_0^2 |A| \delta_{A,A'} + \sigma_1^2 |A|^2 \,\mathrm{ave}_{A\times A'}\, K(x, x')$$

$$\Sigma = \sigma_0^2 |A| I_n + \sigma_1^2 |A|^2 V$$

$$Y \sim N(X\beta, \Sigma)$$

$$Q = I - X(X'X)^{-1}X, \qquad Q_\Sigma = I - X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}$$

Log likelihood based on $QY/\|QY\|$ is

$$-\tfrac{1}{2}(n-p)\log(Y'\Sigma^{-1}Q_\Sigma y) - \tfrac{1}{2}\log|\Sigma| - \tfrac{1}{2}\log|X'\Sigma^{-1}X|$$

$X = \mathrm{row} + \mathrm{col} + \mathrm{treat}$
to eliminate anthropogenic effects

Residual analysis

Residual contrasts $D_{ij}^{rs} = Y_{ij} - Y_{i+r,j} - Y_{i,j+s} + Y_{i+r,j+s}$

$T^{rs} = \sum_{ij} (D_{ij}^{rs})^2; \qquad \hat{T}^{rs} = E(T^{rs}; \hat{\theta})$

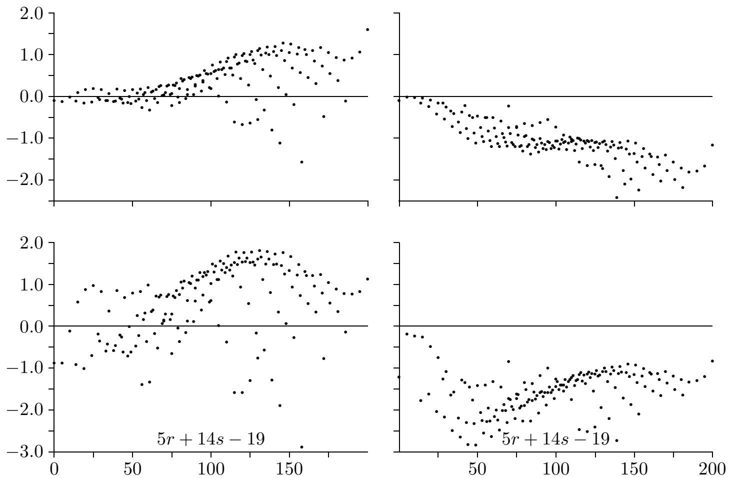


Fig. 2. Log ratios $\log(T^{rs}/\hat{T}^{rs})$ of empirical variances to model-based estimates, unstandardized in the top row and standardized in the bottom row, plotted against the distance $5r + 14s$. The left panels are for the conformal model, and the right panels for the alternative thin-plate spline model.

Conformal model: $K(x, x') = \sigma_0^2 \delta_{x-x'} - \gamma\sigma_0^2 \log|x - x'|)$

Log likelihood for γ for fruit and nut tree yields

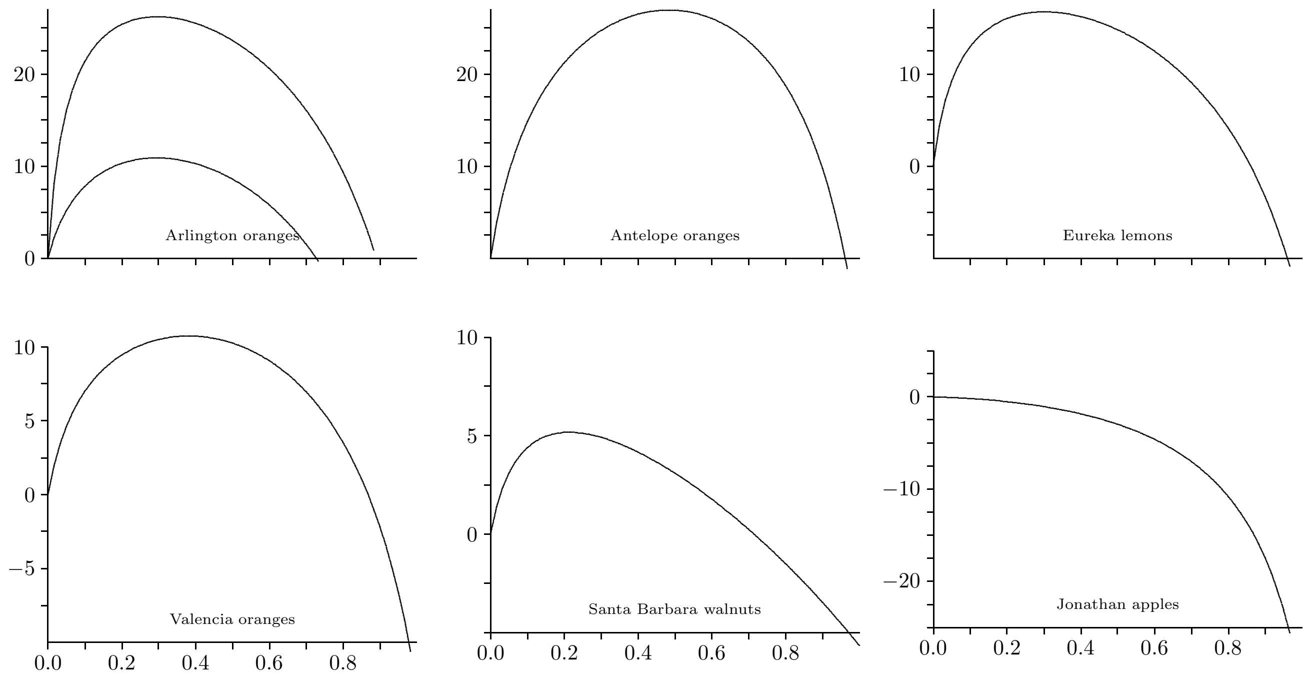


Fig. 3: Marginal log likelihood plots for γ in the conformal model applied to yields from six fruit and nut plantations.

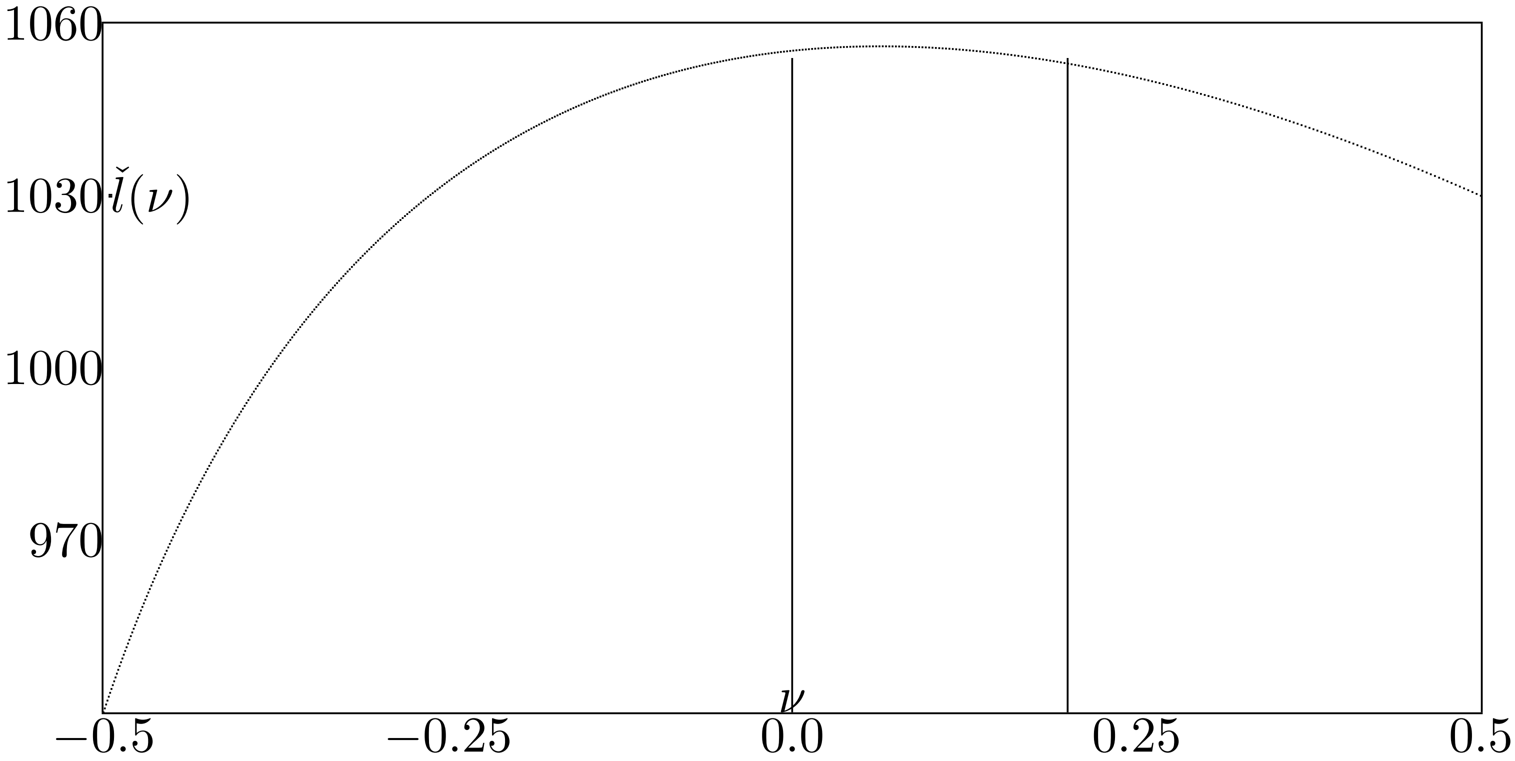
Log likelihood values (white noise = 0)	Confor	Matern	Diff	Power diff
Sawyer grain 1925:	11.41	12.88	1.47	0.97
Sawyer straw 1925:	10.50	11.70	1.20	0.67
Sawyer tops 1926:	5.26	6.40	1.14	1.08
Sawyer roots 1926:	8.24	8.50	0.26	0.26
Arlington I	26.19	26.21	0.02	0.02
Arlington II	10.89	12.44	1.55	1.43
Antelope oranges	26.91	27.57	0.66	0.29
Eureka lemons	16.71	17.58	0.87	0.87
Valencia oranges	10.75	11.06	0.31	0.31
Santa B. walnuts	5.17	6.15	0.98	0.92
Jonathan apples	0.00	0.34	0.34	0.04
Spring barley	84.46	85.80	1.34	1.34
M-H grain	52.22	53.07	0.85	1.21
M-H straw	39.07	41.04	1.97	1.46
Winter wheat	36.47	36.62	0.15	0.06
Federation wheat	589.90	592.74	2.84	2.84
F-S wheat	12.02	12.71	0.69	0.64
Long Hoos oats	3.35	4.15	0.80	0.30
Long Hoos straw	3.16	3.88	0.72	0.17

potatoes	7.59	7.98	0.39	0.04
mangold roots	15.84	17.16	1.32	0.05
mangold tops	7.10	8.66	1.56	0.19
alfalfa 1924	43.56	47.31	3.75	2.07
alfalfa 1925	22.34	26.09	3.75	2.09
Brussels Sp	1.92	2.32	0.40	0.06
			29.56	19.38 = total of 25

Nominal 5\% critical values 3.00 1.92

Nominal 1\% critical values 4.61 3.32

Meta analysis: Combination of information



Profile log likelihood for power index ν

All experiments combined

Conclusions:

Data are consistent with power model (self-similar process) with power 2ν in $(0, 0.5)$.

Anisotropy

Test by comparison with anisotropic model

Evidence of anisotropy in 1/3 cases

Arlington I &II (small but sig)
Antelope (small but sig)
Mangolds (moderately strong)
M-H straw (moderate: historical evidence of ridges)
F-S wheat (moderate-strong, small-scale)
Wiebe (strong, small-scale; plots 1' x 15')

Anisotropy always in drill orientation
suggests anthropogenic origin

Fairfield Smith's empirical power law

n contiguous plots of area x

$E(s^2) = c_n x^{1+b}$ with $0 < b < 1$

Conformal model

$m \times m$ array of plots of area x

$E(s^2) = \sigma_0^2 x + \sigma_1^2 x^2 m^2 \log(m)/(m^2 - 1)$

Comparison with F-S:

$m^2 x = T$ fixed and $x/T < 0.1$

$E(s^2/x) \propto 1 - \frac{1}{2}(x/T)\log(x/T) \times (T\sigma_1^2/\sigma_0^2)$

$E\log((s^2/x)) \propto -\frac{1}{2}(x/T)\log(x/T) \times (T\sigma_1^2/\sigma_0^2)$

Gives good qualitative agreement with F-S