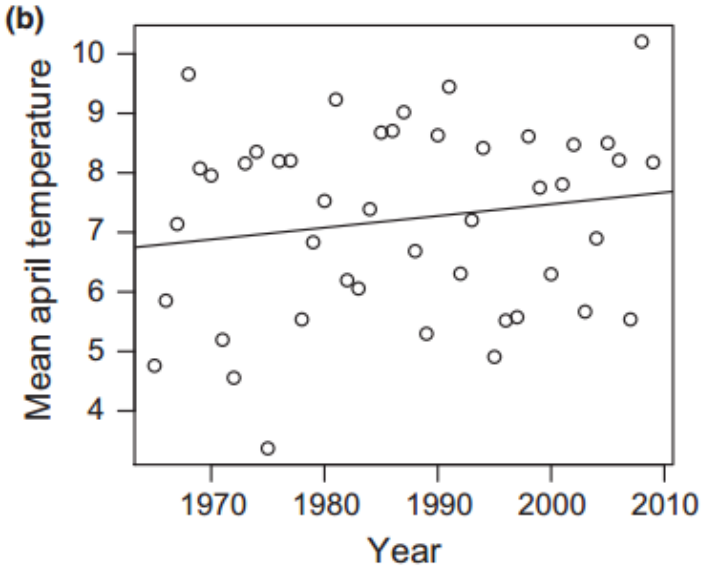
The data and the model: an introduction

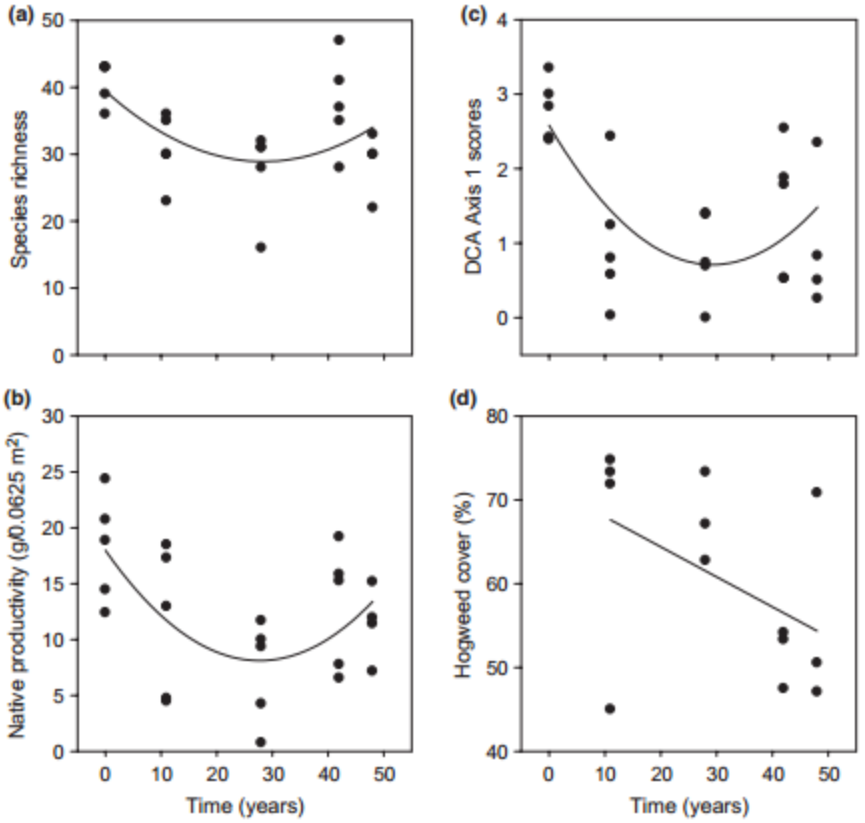
Petr Keil

petr.keil@yale.edu
www.petrkeil.com

11/7/2013



Bartholomeus et al. 2013 Eco. Lett.



Dostal et al. 2013 Eco. Lett.

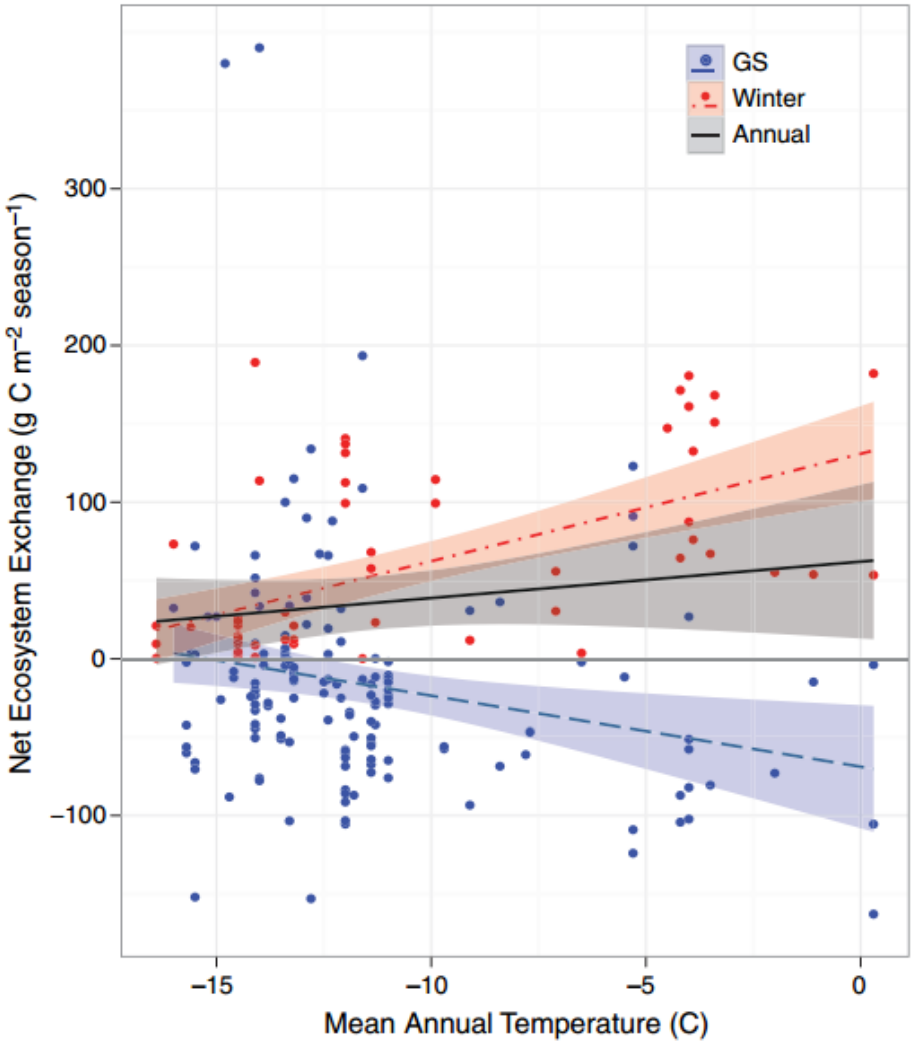
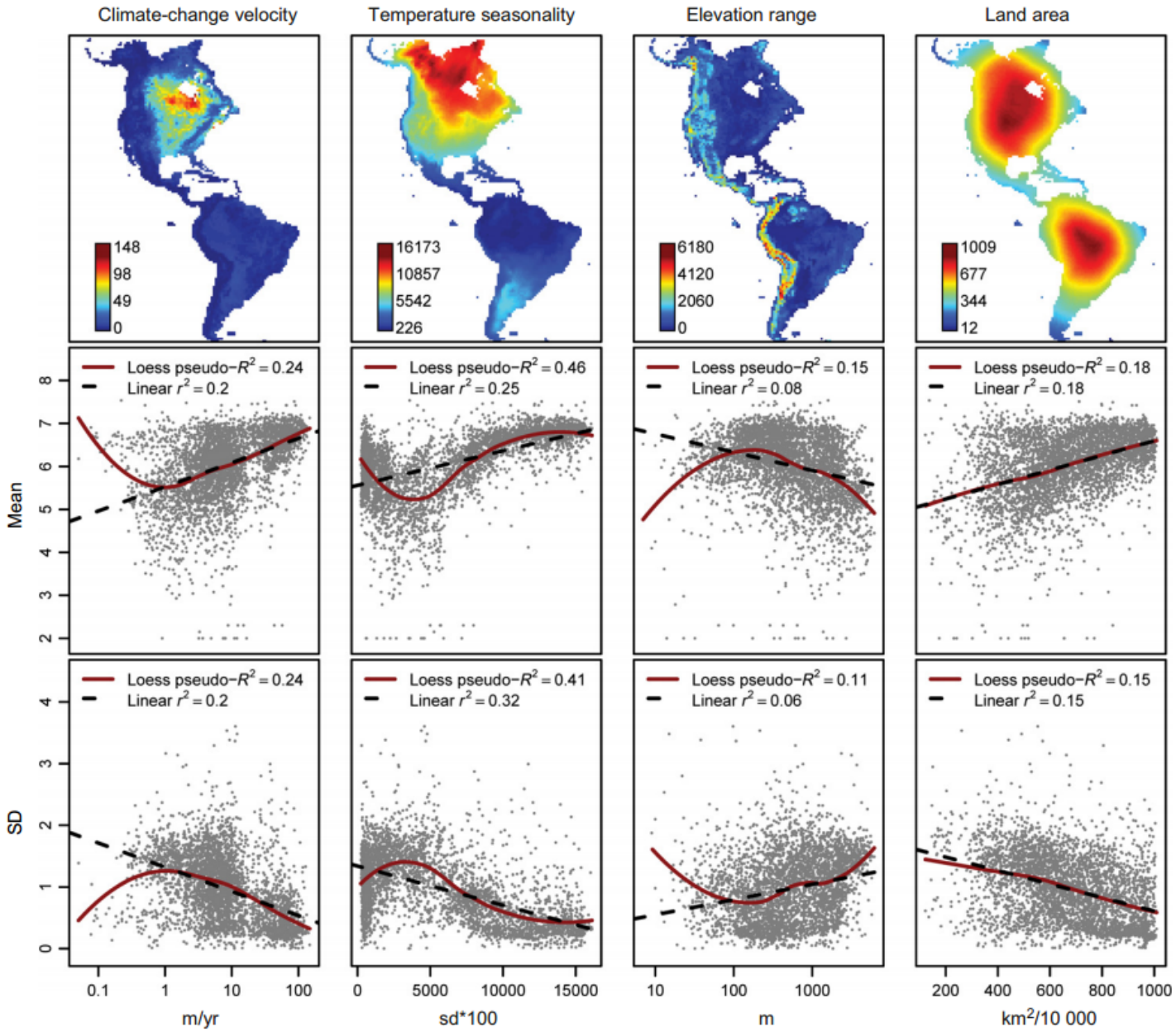


Figure 5 Estimates of linear relationships between mean annual temperature (°C) and net ecosystem exchange (g C m^{-2} $season^{-1}$) of CO_2 during the growing season (GS; blue), and winter (red) with 95% confidence intervals, and the annual estimate (black) of the temperature relationship produced by summing predicted values of seasonal trends and their variances. Positive values denote a C source.

Belshe, Schurr & Bolker et al. 2013 Eco. Lett.



Morueta-Holme et al. 2013 Eco. Lett.

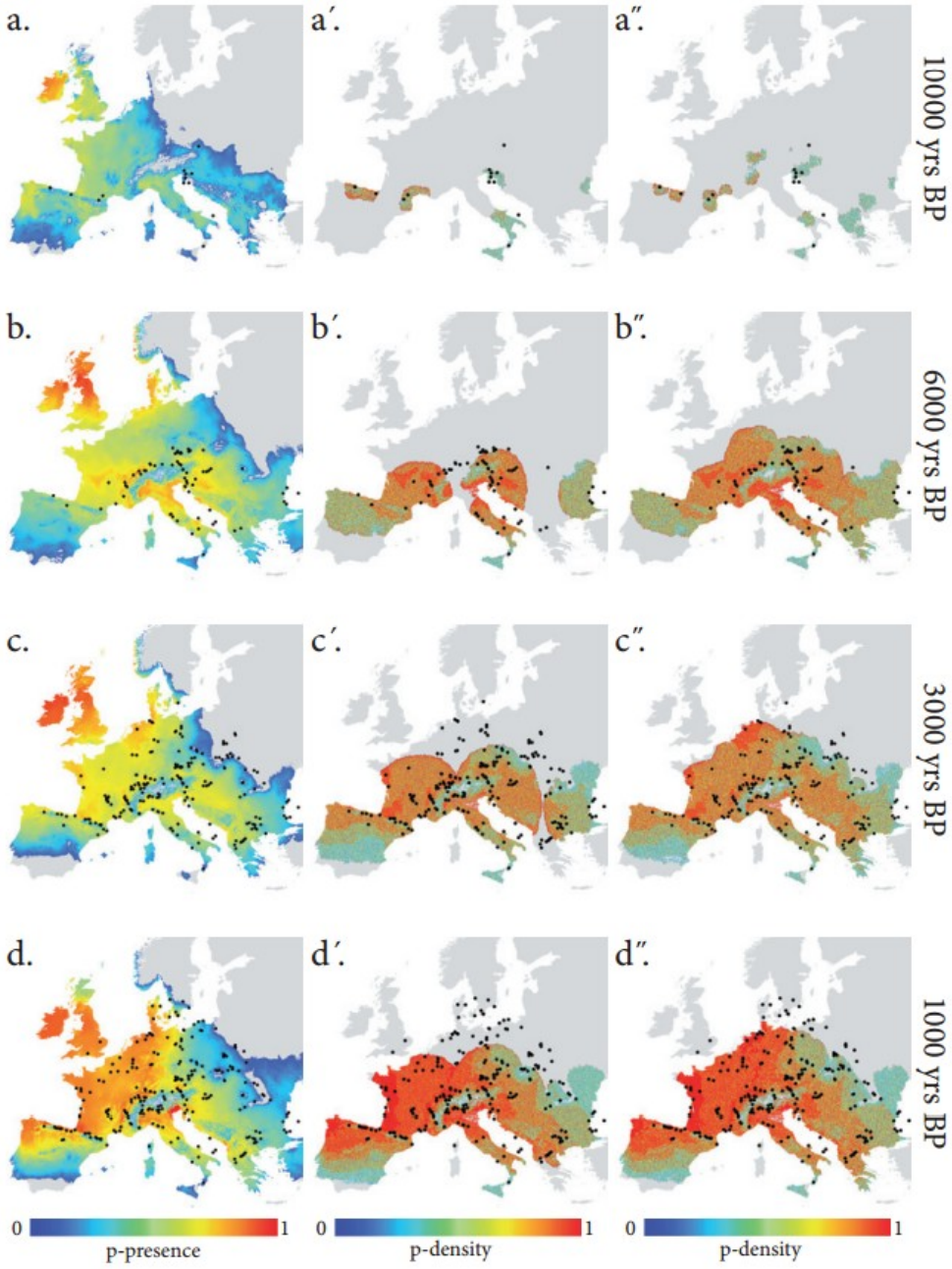
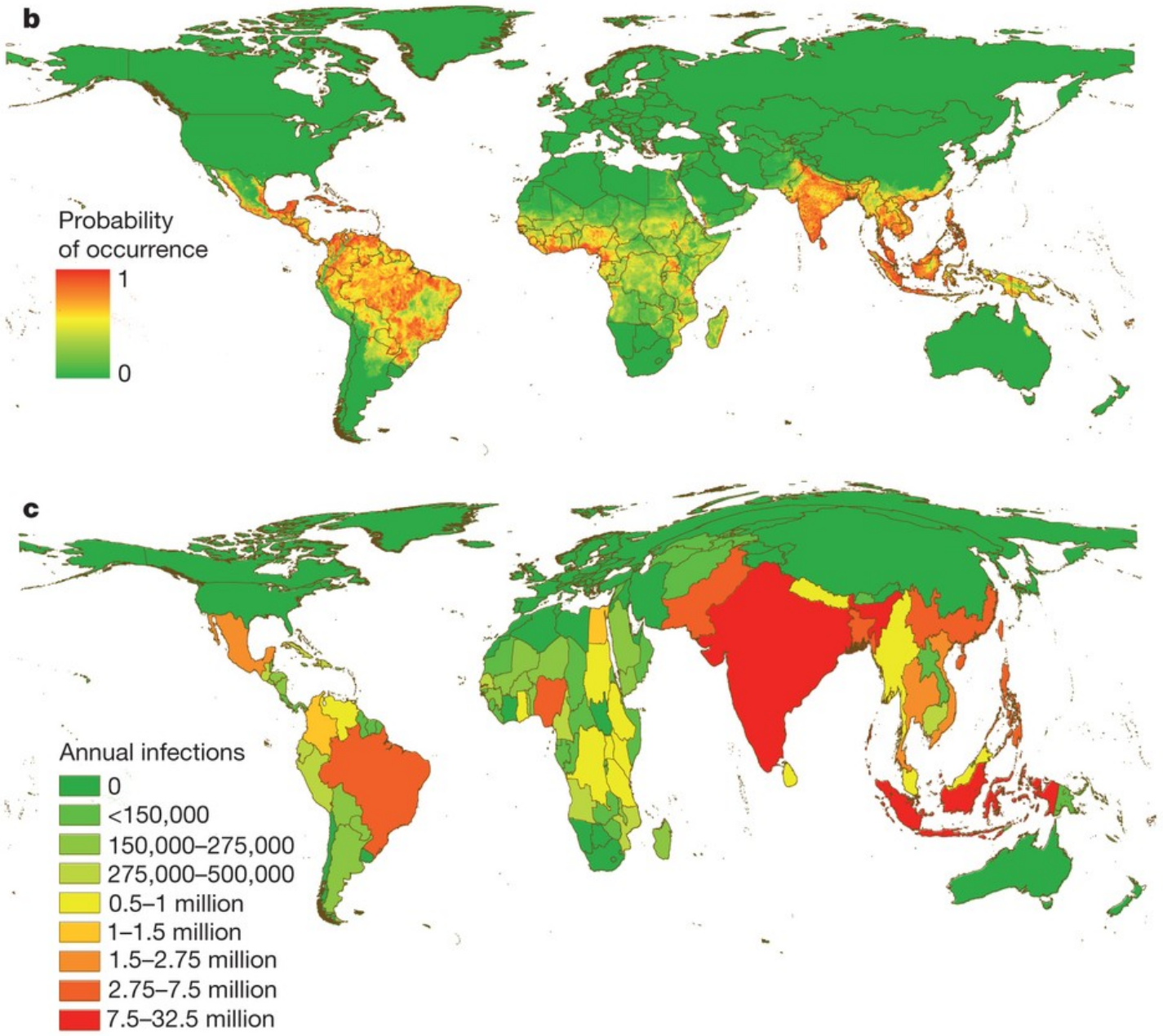


Figure 2 *Fagus sylvatica* distribution maps at 10 (a, a′ , a″), 6 (b, b′ , b″), 3 (c, c′ , c″) and 1 (d, d′ , d″) kyr BP. Potentially suitable habitats (a–d) derived from probabilities of presence (p-presence) simulated by PHENOFIT. Colonized habitats, starting from the first (a′–d′) and the second (a″–d″) refugia scenario, derived from probabilities of density (p-density, number of cohorts in each grid cell) simulated by PHENOFIT and the Gibbs-based model. Grey areas correspond to a simulated absence of beech. Black points correspond to palaeoarchives of beech [pollen > 2% and macrofossils, (Magri et al., 2006)] at the same dates (indicated on the right side of the figure).

Saltre et al. 2013 GEB



Bhatt et al. 2013 Nature

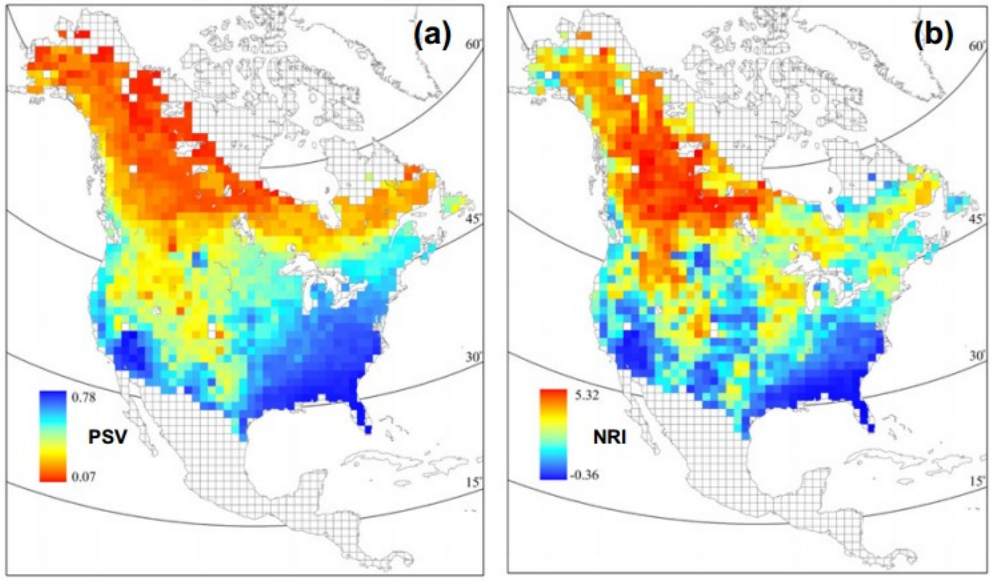
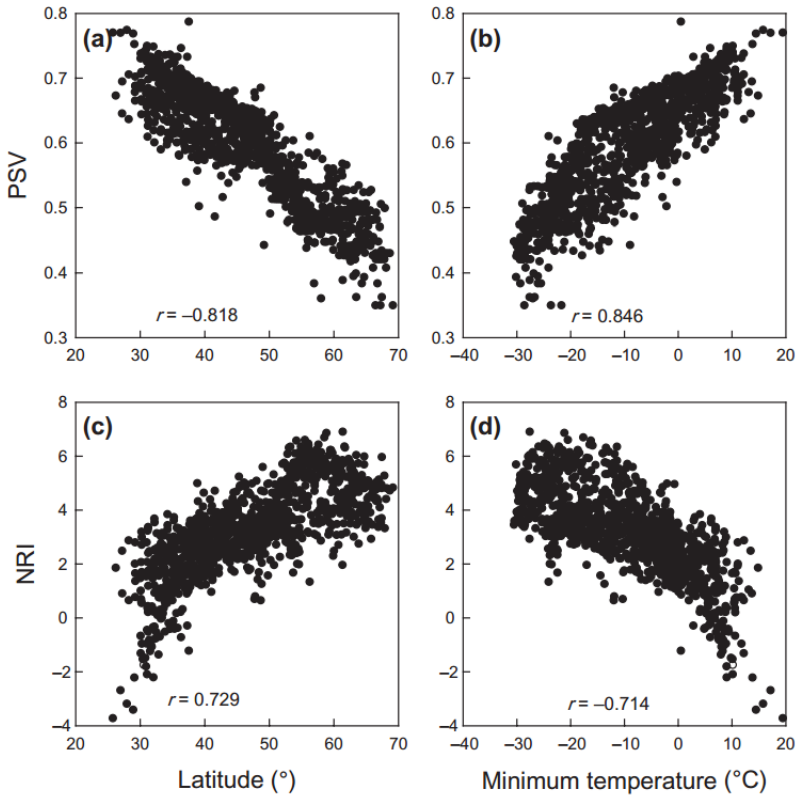


Figure 1 Spatial variation in phylogenetic species variability (PSV) (a) and net relatedness index (NRI) (b).



Qian et al. 2013 GEB

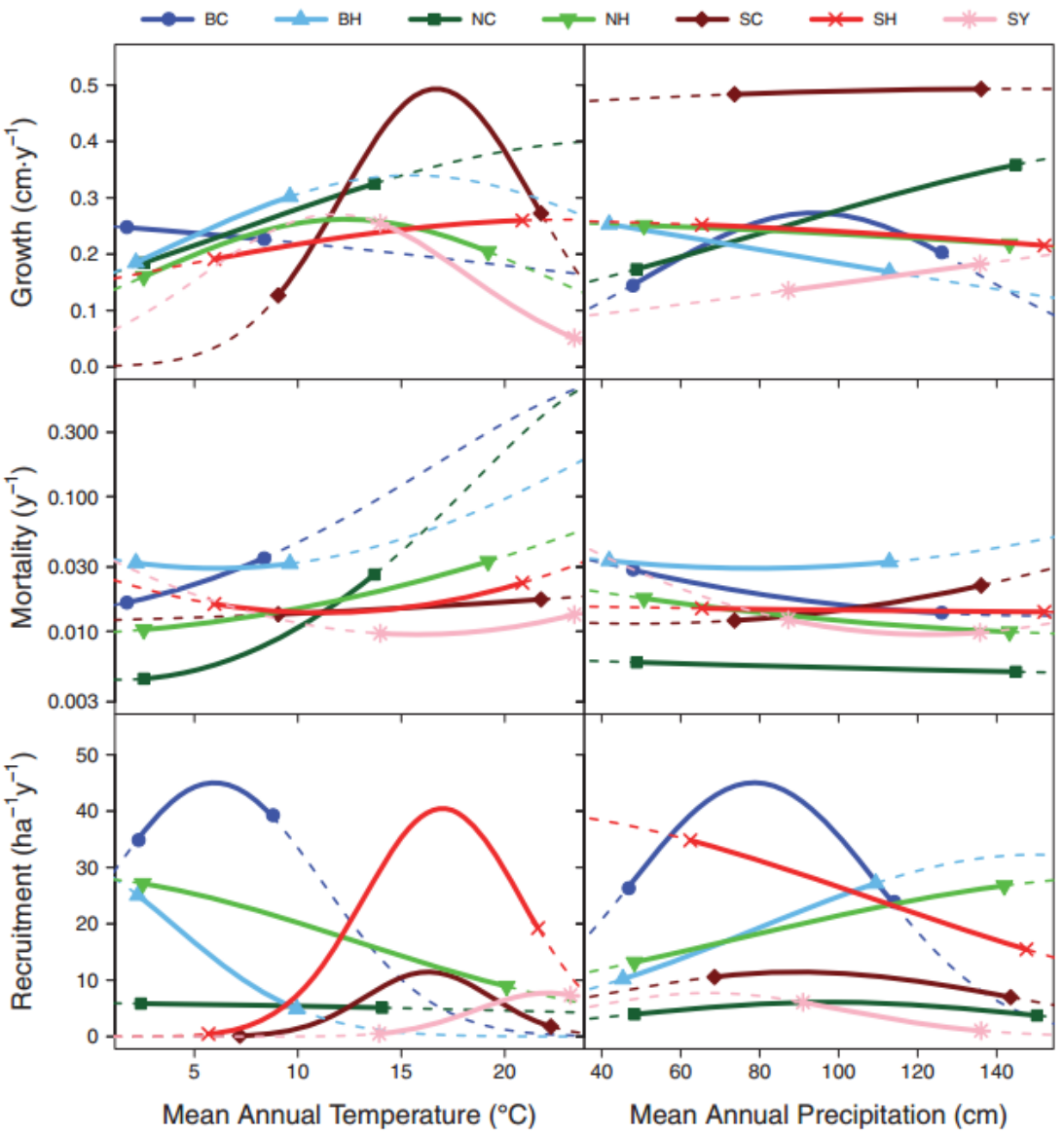
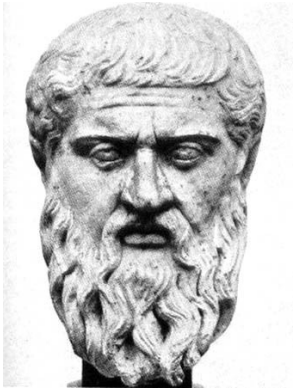


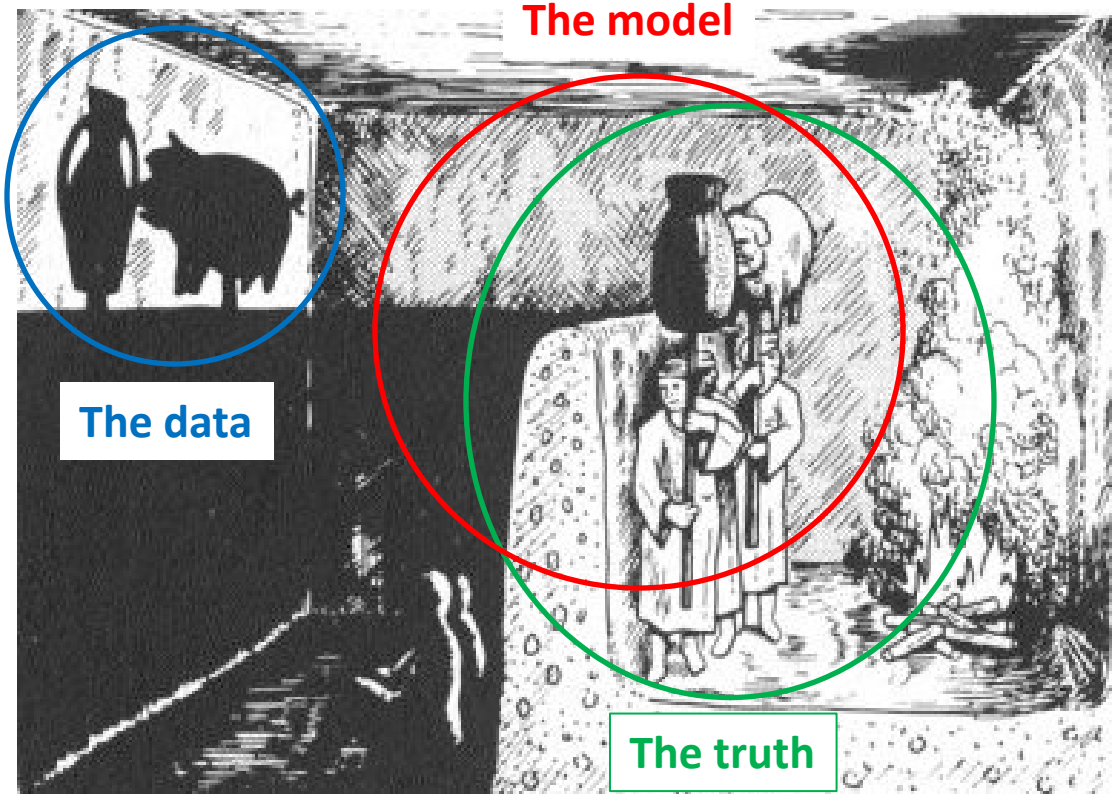
Figure 2 Predicted tree diameter growth rate, mortality rate and recruitment rate (ingrowth of stems surpassing 2.54 cm diameter) for seven plant functional types (PFTs) in the eastern USA against mean annual temperature and precipitation. The climate dependency of each rate was modelled as a Gaussian-shaped response to a linear combination of temperature and precipitation. The solid portion of each line indicates the interval between the 1st and 99th percentiles of temperature or precipitation for each PFT; the dashed portion represents extrapolations outside this range. All other covariates were fixed to their mean values in calculating these responses. PFT legend abbreviations are as in Fig. 3.

Vanderwell et al. 2013 GEB

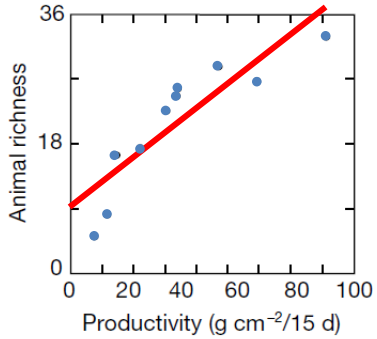


Plato

Karl Pearson



Chase & Leibold 2002 Nature



P(D|H) : the LIKELIHOOD

Probability (frequency) of the data, given the model.

Is only useful for indirect statements about the hypothesis.

Easy to calculate.

P(H|D)

Strength of our belief in the hypothesis (model), given the data.

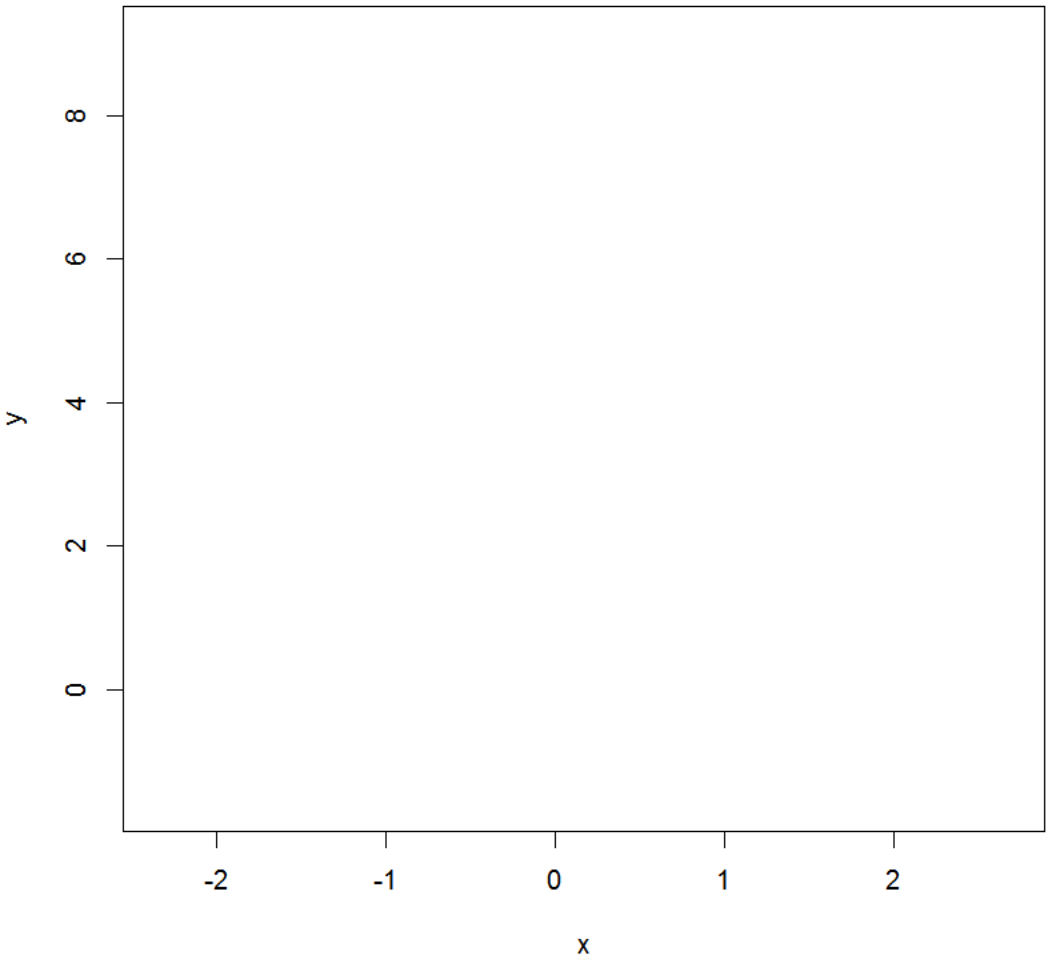
Can be used for direct statements about the hypotheses.

Difficult to calculate.

$$P(H|D) = \frac{P(H) \times P(D|H)}{P(D)}$$

Normal linear regression (GLM family “Gaussian”)

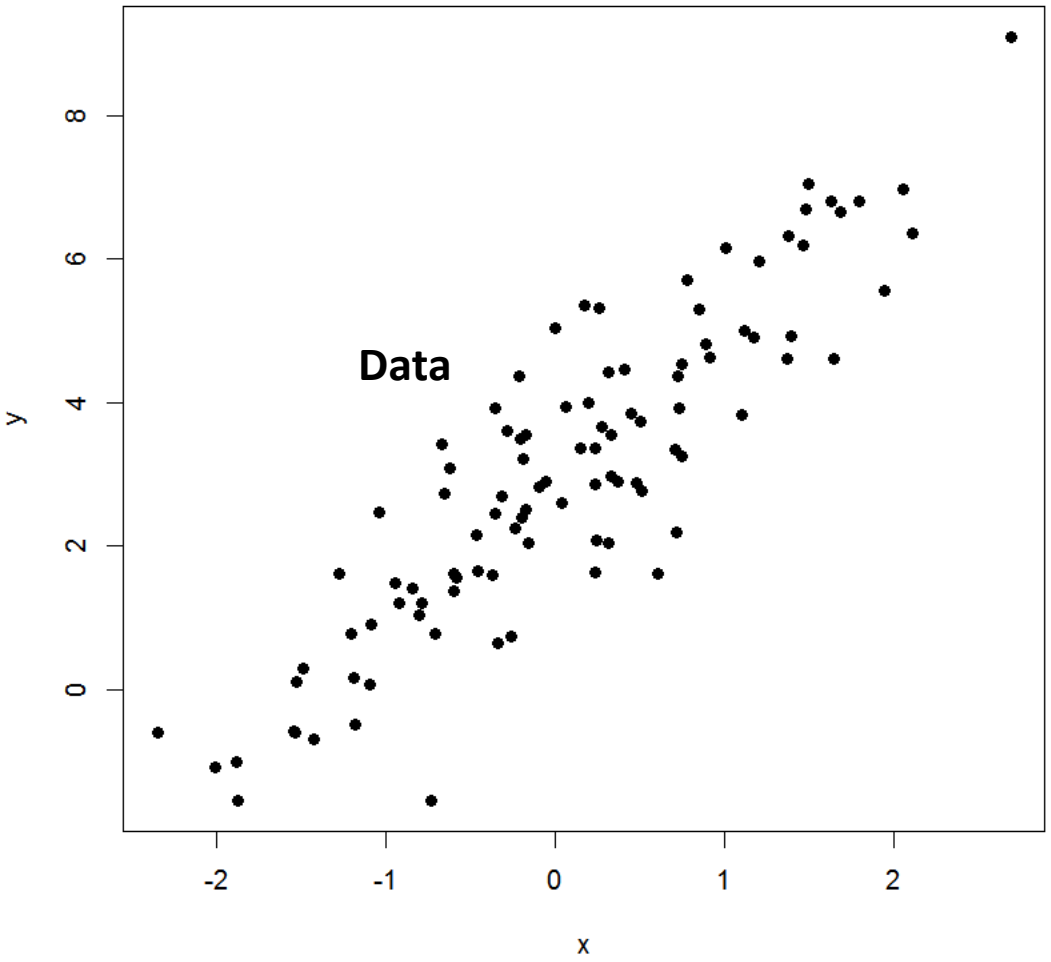
Response, dependent variable, explained variable



Predictor, independent variable, explanatory variable

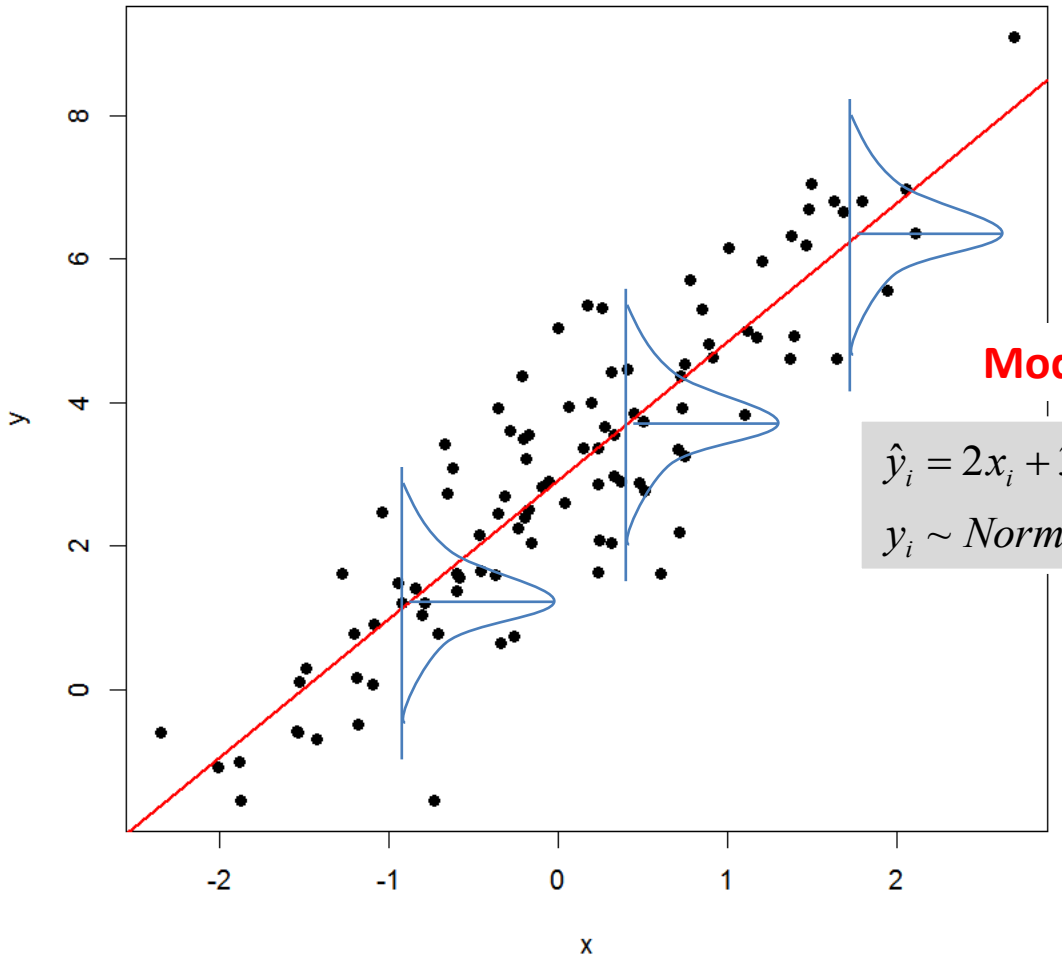
Normal linear regression (GLM family “Gaussian”)

Response, dependent variable, explained variable



Predictor, independent variable, explanatory variable

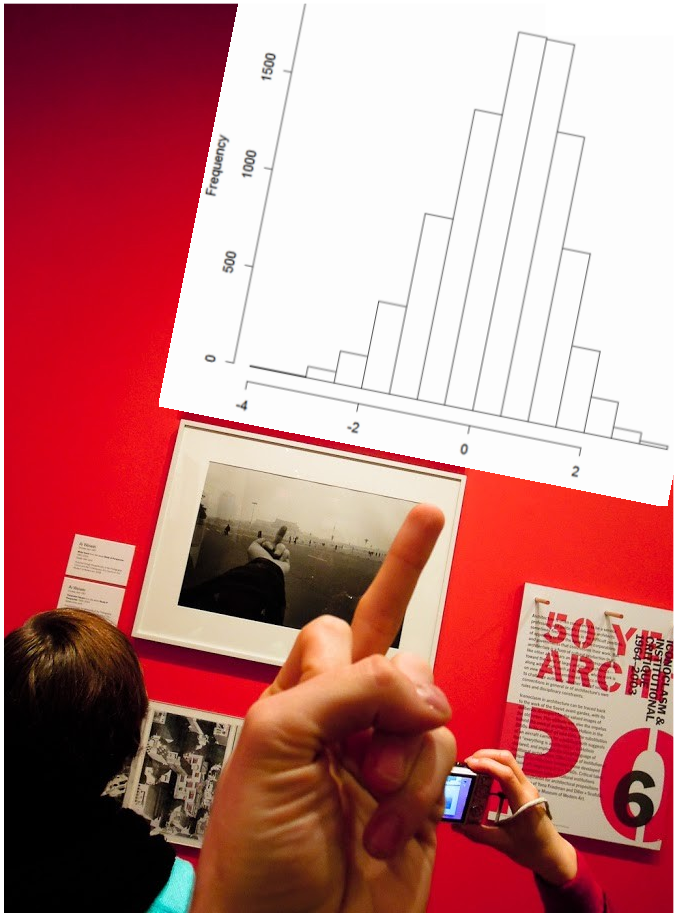
Normal linear regression (GLM family “Gaussian”)



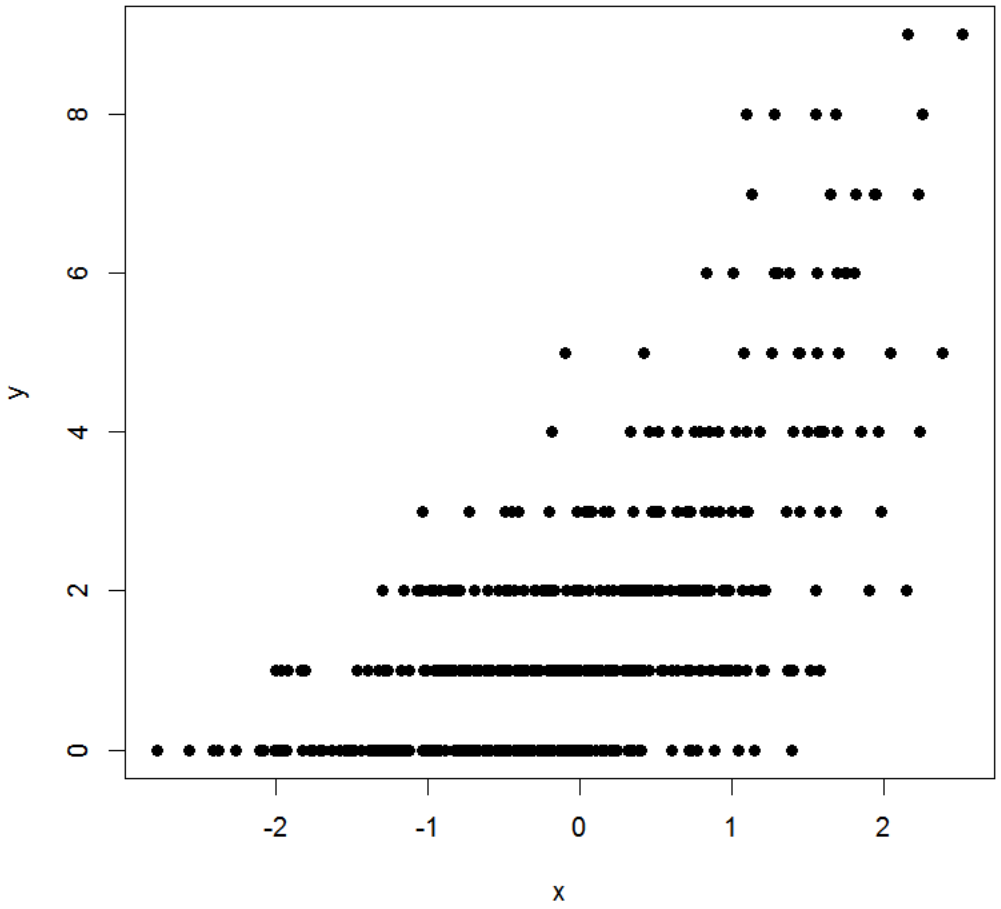
Model

$$\hat{y}_i = 2x_i + 3$$

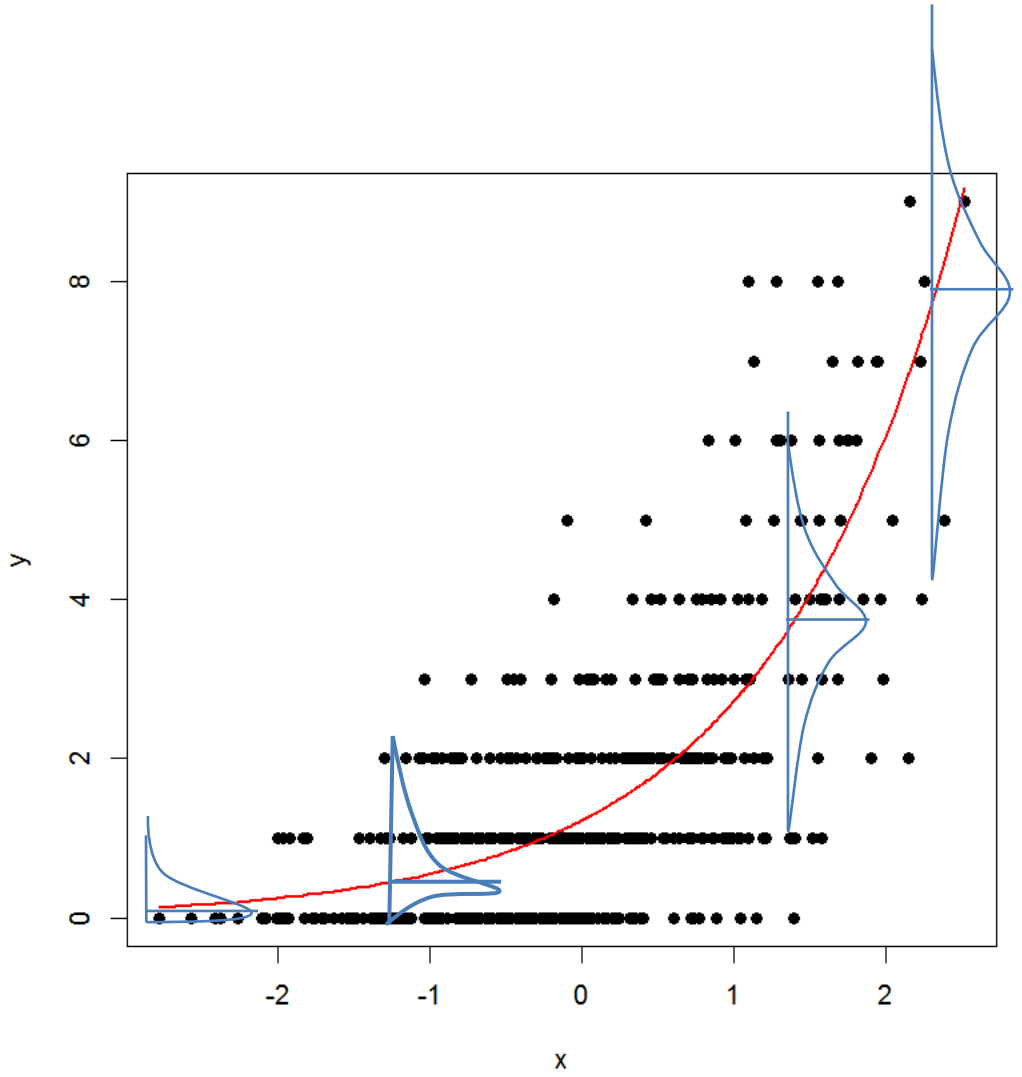
$$y_i \sim Normal(\hat{y}_i, \sigma^2)$$


Frequency
0
500
1000
1500
-4
-2
0
2
50 Y
ARCE
6

Poisson log-linear regression (GLM family “Poisson”)



Poisson log-linear regression (GLM family “Poisson”)

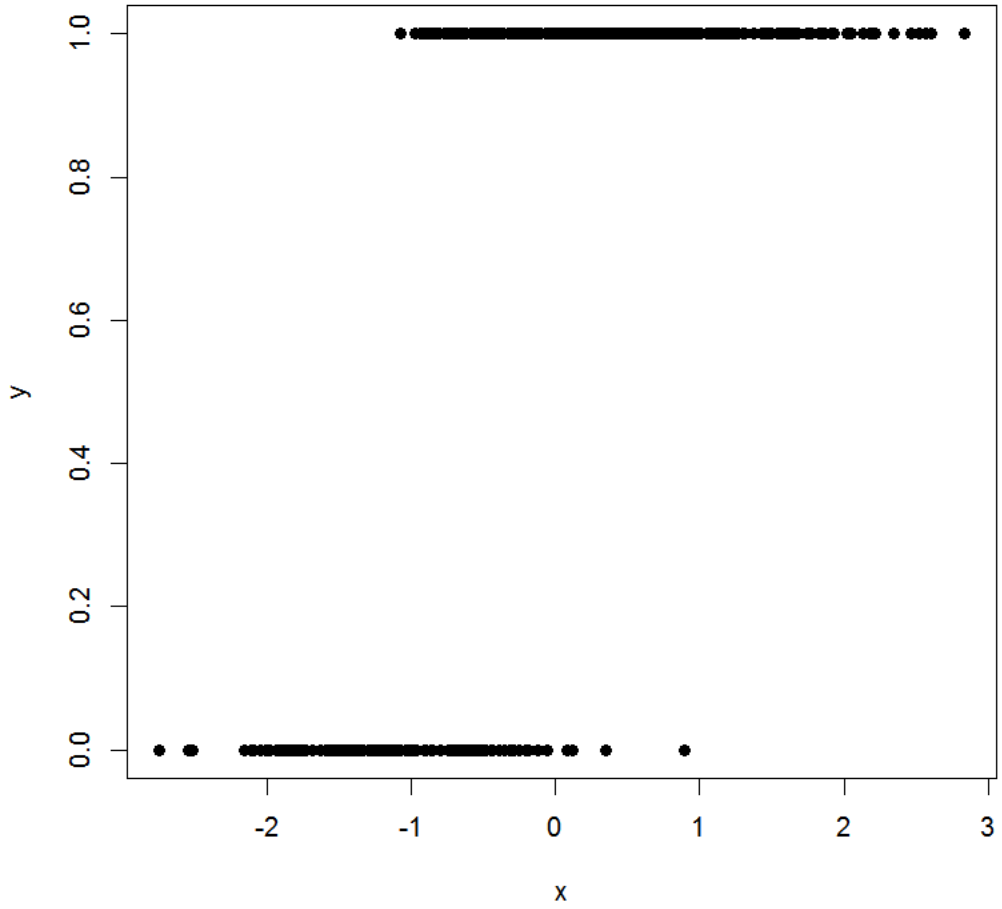


Model

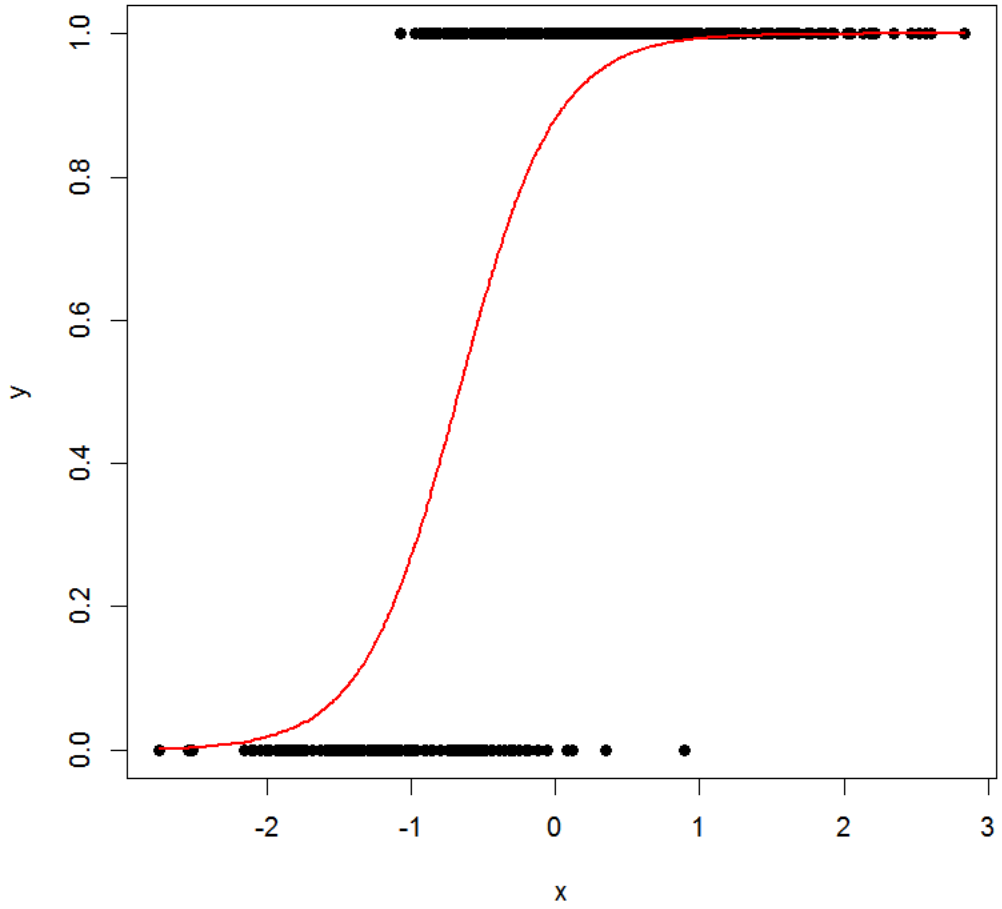
$$\log \hat{y}_i = 0.2x_i + 0.8$$

$$y_i \sim Poisson(\hat{y})$$

Binomial (logistic) regression (GLM family “binomial”)



Binomial (logistic) regression (GLM family “binomial”)



Model

$$\log\left(\frac{p_i}{1-p_i}\right) = 3x_i + 2$$

$$y_i \sim Bernoulli(p_i)$$