Pornic, France
BioBayes 2019

Meta-analysis with Bayesian statistical methods

David Makowski
INRA
david.makowski@inra.fr

The ‘data synthesis challenge’

As more and more publications and data become available, how to conduct rigorous and comprehensive assessments?

Formal methods are needed to conduct rigorous and comprehensive quantitative synthesis

Meta-analysis: a statistical approach for quantitative synthesis

« The analysis of analyses »

« The statistical analysis of a large collection of results from individual studies for the purpose of integrating the findings »

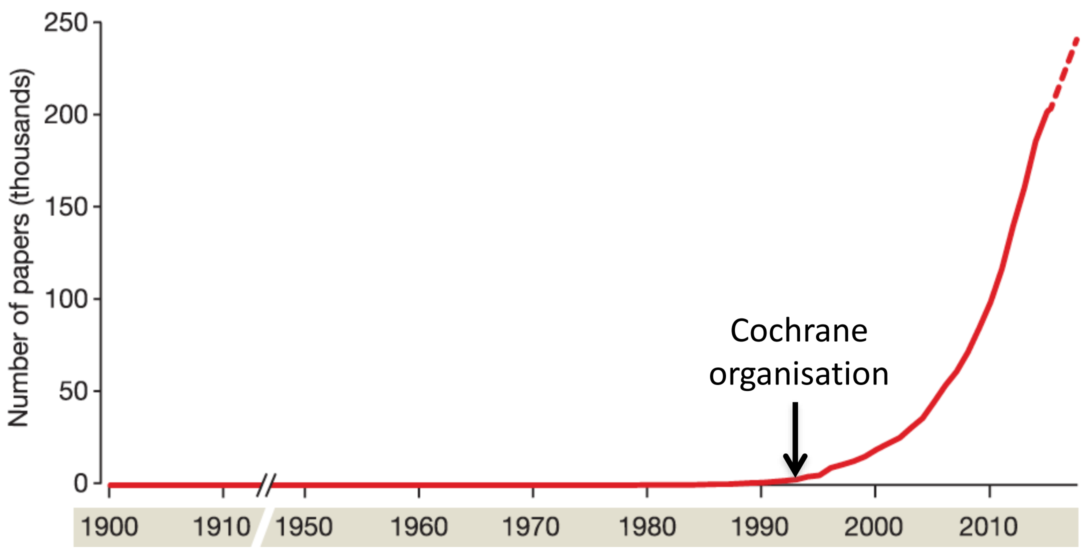
« Systematic review + statistical analysis »

Dictionary of epidemiology, 2001; Chalmers et al., 2002; Glass, 1976; Koricheva et al., 2013

Meta-analysis and the science of research synthesis

Jessica Gurevitch[1], Julia Koricheva[2], Shinichi Nakagawa[3,4] & Gavin Stewart[5]

Nature, March 2018



Cochrane

Trusted evidence.
Informed decisions.
Better health.

Harvard
Business
Review

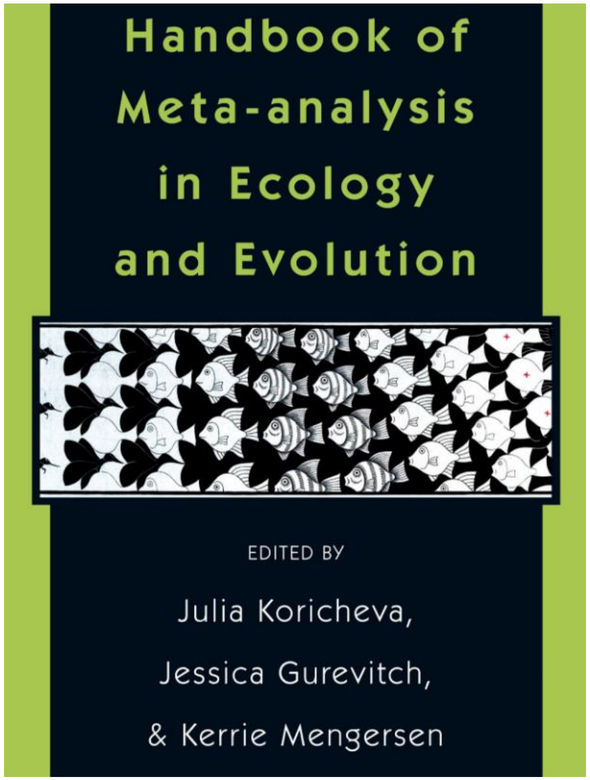
ECONOMICS & SOCIETY

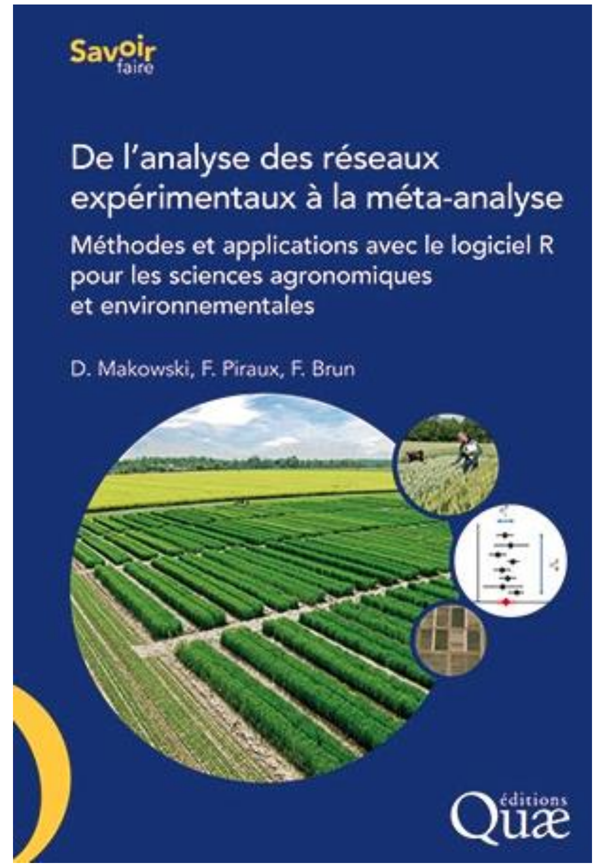
How to Be a Smart Consumer of Social Science Research

by Eva Vivalt

JULY 27, 2018

The immediate answer is to not rely too much on any one study. Whenever possible, look for meta-analyses or systematic reviews that synthesize results from many studies, as they can provide more-credible evidence and sometimes suggest reasons why results differ.


Handbook of
Meta-analysis
in Ecology
and Evolution
EDITED BY
Julia Koricheva,
Jessica Gurevitch,
& Kerrie Mengersen


Savoir faire
De l'analyse des réseaux
expérimentaux à la méta-analyse
Méthodes et applications avec le logiciel R
pour les sciences agronomiques
et environnementales
D. Makowski, F. Piraux, F. Brun
éditions
Quæ

Scoping

Literature search

Paper selection

Data extraction

Statistical analysis

Bias and uncertainty

Scoping

Literature search

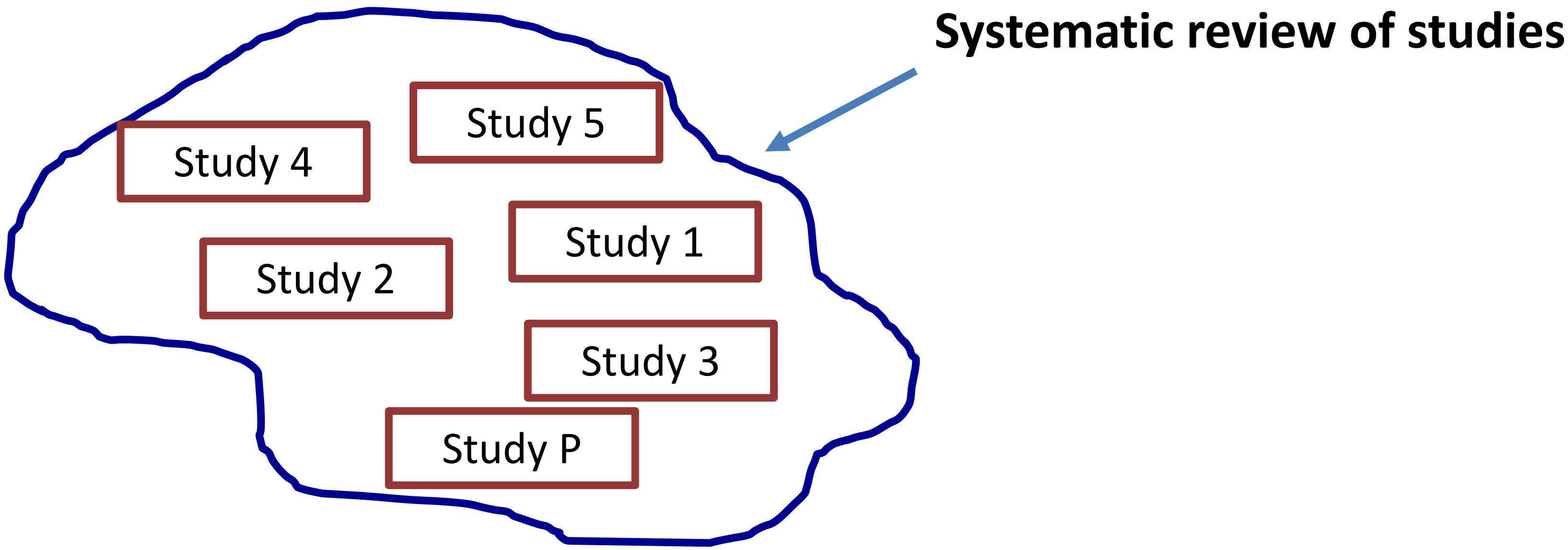
About 2 papers per day

Paper selection

Data extraction

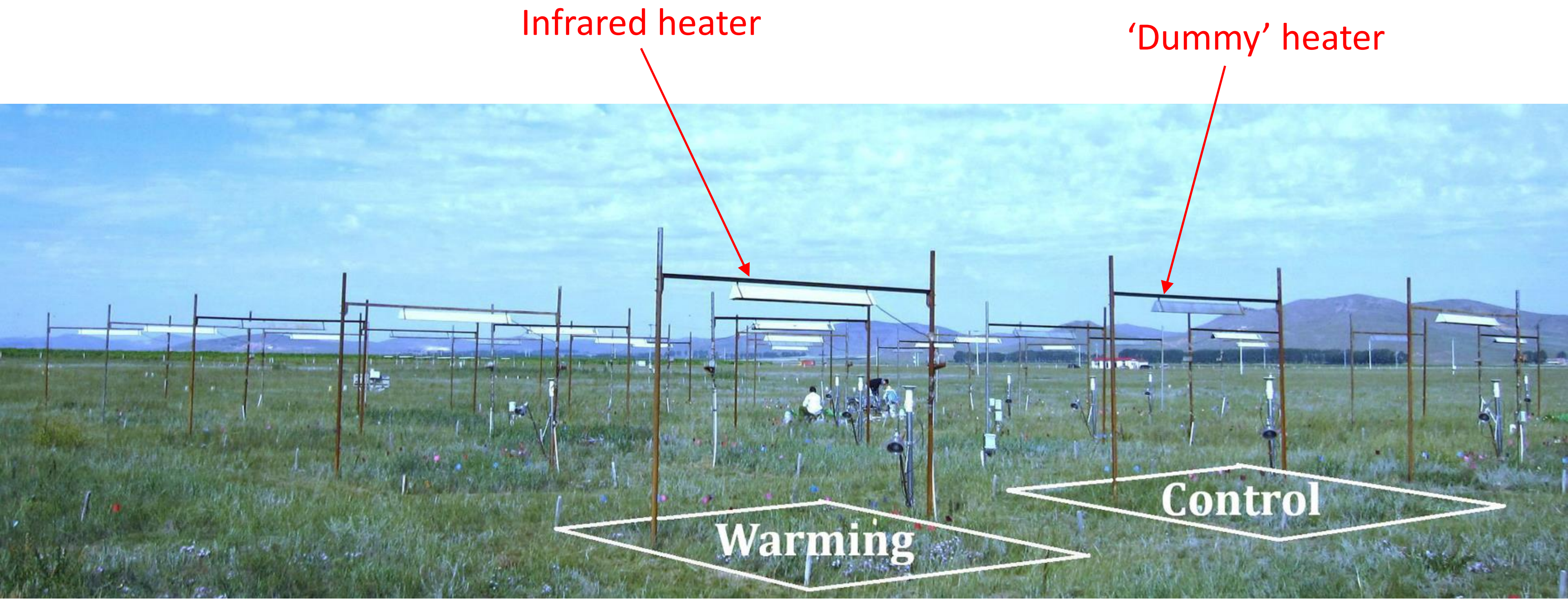
Statistical analysis

Bias and uncertainty



Set of studies dealing with a specific topic (e.g., %yield loss due to +1° C)

Field warming experiment

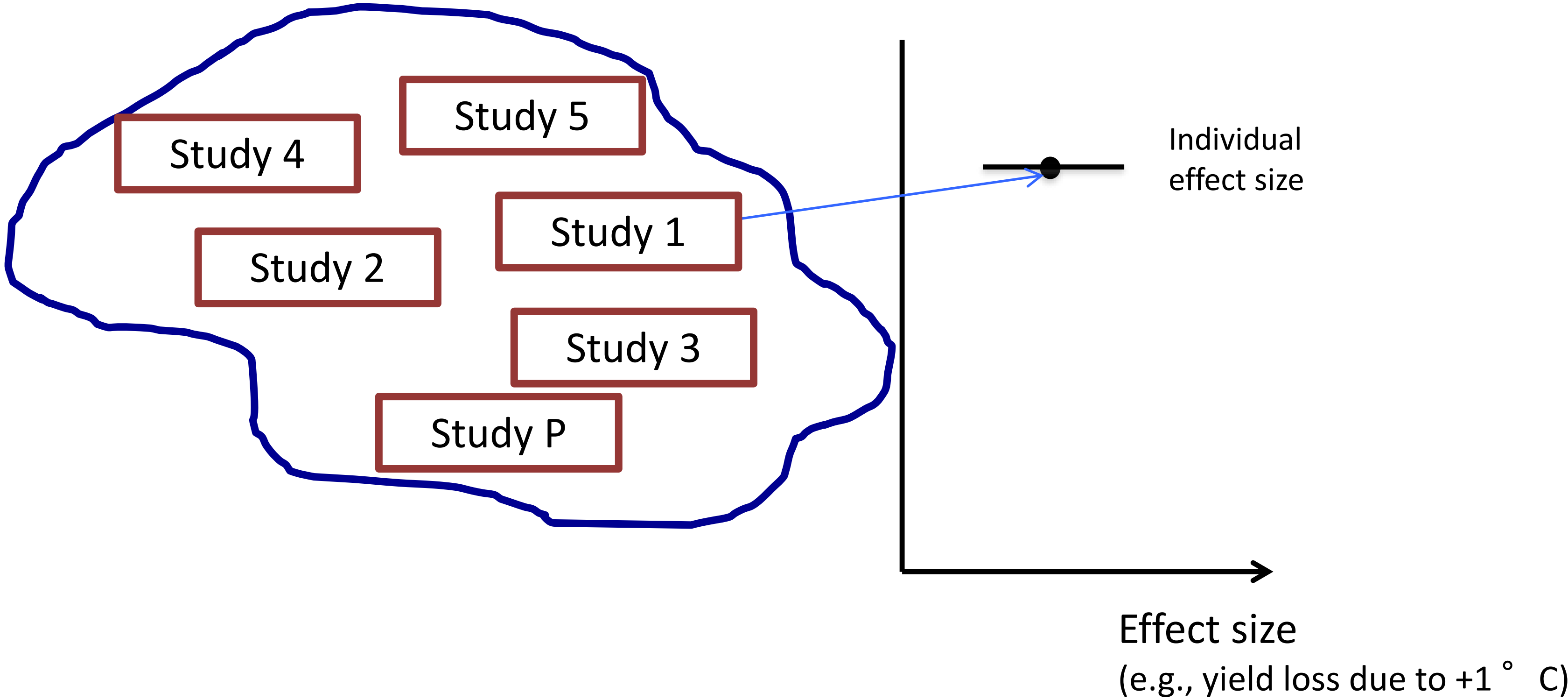


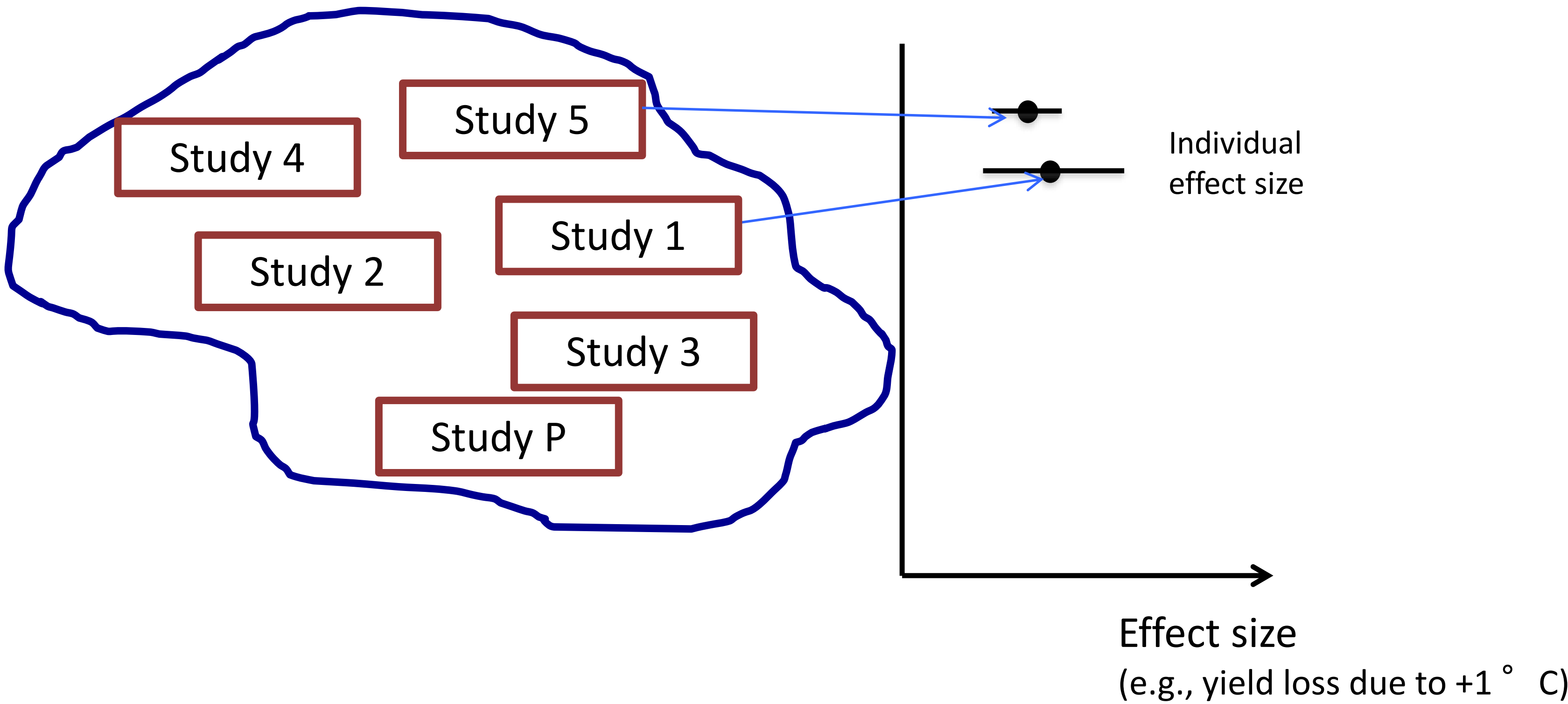
from Chi et al. 2013

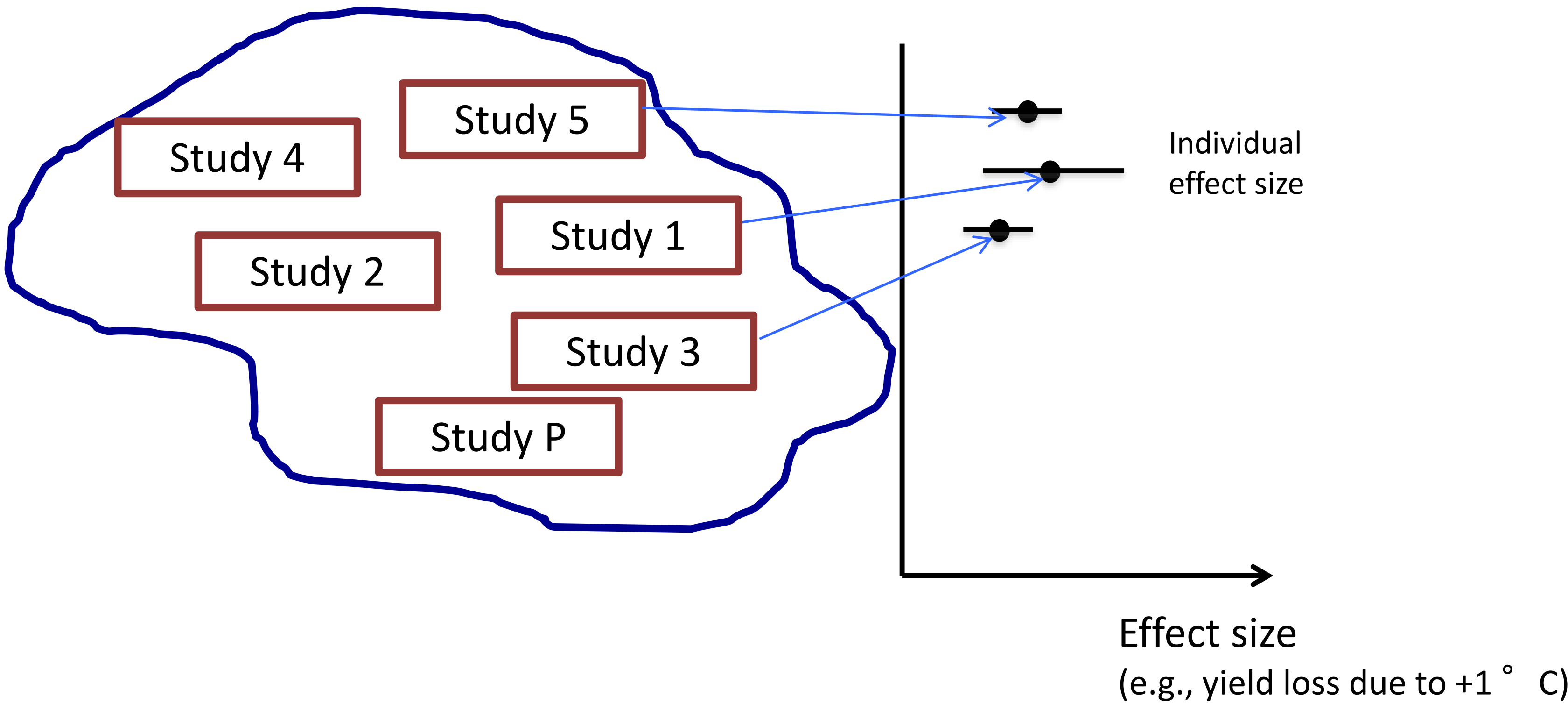
doi.org/10.1371/journal.pone.0056482

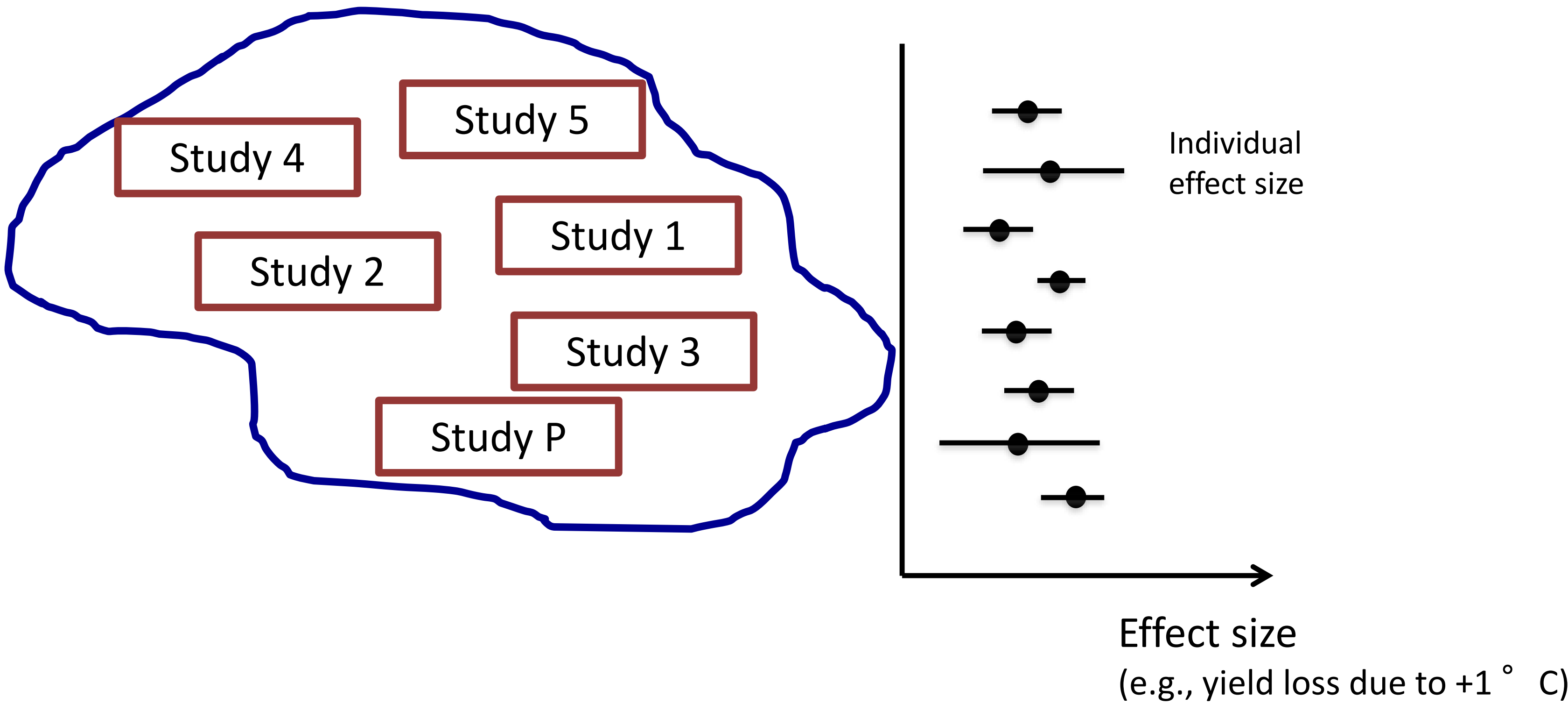
$$\Delta Y = (Yield_{warm}\text{-}\ Yield_{control})/Yield_{control}$$

$$Sensitivity = Yield\ \%\ change\ per\ {}^{\circ}\mathrm{C} = 100\,\frac{\Delta Y}{\Delta T}$$

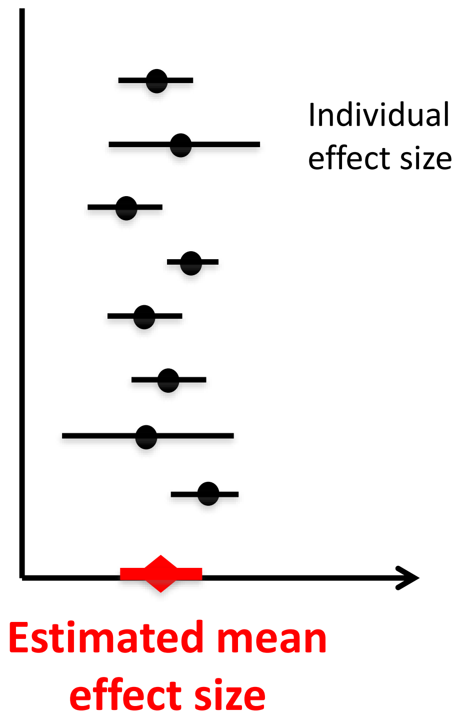

Study 5
Study 4
Study 1
Study 2
Study 3
Study P
Individual
effect size
Effect size
(e.g., yield loss due to +1 ° C)

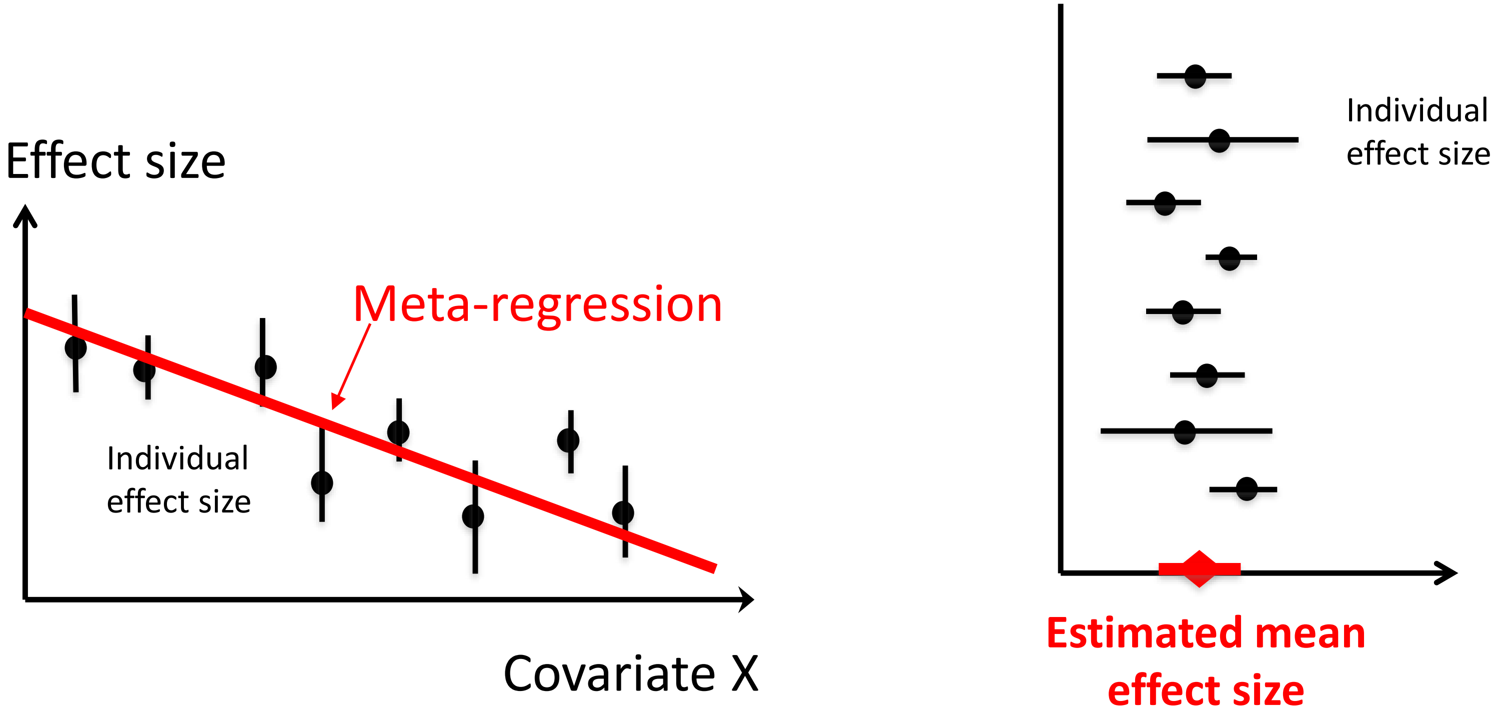

Study 5
Study 4
Study 1
Study 2
Study 3
Study P
Individual
effect size
Effect size
(e.g., yield loss due to +1 ° C)


Study 5
Study 4
Study 1
Study 2
Study 3
Study P
Individual
effect size
Effect size
(e.g., yield loss due to +1 ° C)



Study 5
Study 4
Study 1
Study 2
Study 3
Study P
Individual effect size
Effect size
(e.g., yield loss due to +1 ° C)


Individual
effect size
Estimated mean
effect size



Effect size
Meta-regression
Individual
effect size
Covariate X
Individual
effect size
Estimated mean
effect size

nature plants

Letter | Published: 19 December 2016

Plausible rice yield losses under future climate warming

Chuang Zhao, Shilong Piao, Xuhui Wang, Yao Huang, Philippe Ciais, Joshua Elliott, Mengtian Huang, Ivan A. Janssens, Tao Li, Xu Lian, Yongwen Liu, Christoph Müller, Shushi Peng, Tao Wang, Zhenzhong Zeng & Josep Peñuelas

Compilation of results of 83 field warming experiments located in 14 sites in the world

Field experiment Ambient CO_2

$$\Delta Y = (Yield_{warm} - Yield_{control})/Yield_{control}$$

$$Sensitivity = Yield\ \%\ change\ per\ ^{\circ}\mathrm{C} = 100\,\frac{\Delta Y}{\Delta T}$$

83 values of yield sensitivity (% yield loss per° C) in 14 sites

Country	Site name	Latitude	Longtitude	Research time (year)	Nitrogen (kg ha^{-1})	Warming design	Warming type	Temperature change (ΔK)	Growing season temperature (K)	$S_{Y,T}^{obs}$ (% K^{-1})
Philippines	Los Banos	14.22	121.25	1994	110	Open top camber	Passive	4.0	299.7	-6.4
Philippines	Los Banos	14.22	121.25	1995	220	Open top camber	Passive	4.0	299.1	-4.1
Nepal	Khumaltar	27.65	85.33	2001	N.A.	Open top camber	Passive	6.8	296.8	1.1
Nepal	Khumaltar	27.65	85.33	2002	N.A.	Open top camber	Passive	4.4	296.2	2.3
Nepal	Khumaltar	27.65	85.33	2003	N.A.	Open top camber	Passive	5.8	296.7	6.6
Nepal	Khumaltar	27.65	85.33	2004	N.A.	Open top camber	Passive	7.3	296.4	2.1

…

from Zhao et al., 2016

83 values of yield sensitivity in 14 sites

Country	Site name	Latitude	Longtitude	Research time (year)	Nitrogen (kg ha^{-1})	Warming design	Warming type	Temperature change (ΔK)	Growing season temperature (K)	$S_{Y,T}^{obs}$ (% K^{-1})
Philippines	Los Banos	Study 1		1994	110	Open top camber	Passive	4.0	299.7	-6.4
Philippines	Los Banos			1995	220	Open top camber	Passive	4.0	299.1	-4.1
Nepal	Khumaltar			2001	N.A.	Open top camber	Passive	6.8	296.8	1.1
Nepal	Khumaltar	Study 2		2002	N.A.	Open top camber	Passive	4.4	296.2	2.3
Nepal	Khumaltar			2003	N.A.	Open top camber	Passive	5.8	296.7	6.6
Nepal	Khumaltar			2004	N.A.	Open top camber	Passive	7.3	296.4	2.1

...

from Zhao et al., 2016

83 values of yield sensitivity in 14 sites

Country	Site name	Latitude	Longtitude	Research time (year)	Nitrogen (kg ha^{-1})	Warming design	Warming type	Temperature change (ΔK)	Growing season temperature (K)	$S_{Y,T}^{obs}$ (% K^{-1})
Philippines	Los Banos	Study 1		1994	110	Open top camber	Passive	4.0	299.7	-6.4
Philippines	Los Banos			1995	220	Open top camber	Passive	4.0	299.1	-4.1
Nepal	Khumaltar			2001	N.A.	Open top camber	Passive	6.8	296.8	1.1
Nepal	Khumaltar	Study 2		2002	N.A.	Open top camber	Passive	4.4	296.2	2.3
Nepal	Khumaltar			2003	N.A.	Open top camber	Passive	5.8	296.7	6.6
Nepal	Khumaltar			2004	N.A.	Open top camber	Passive	7.3	296.4	2.1

…

from Zhao et al., 2016

Two levels of variability:
- Within study
- Between studies

Country	Site name	Latitude	Longtitude	Research time (year)	Nitrogen (kg ha^{-1})	Warming design	Warming type	Temperature change (ΔK)	Growing season temperature (K)	$S^{obs}_{Y,T}$ (% K^{-1})
Philippines	Los Banos	**Study 1**		1994	110	Open top camber	Passive	4.0	299.7	-6.4
Philippines	Los Banos			1995	220	Open top camber	Passive	4.0	299.1	-4.1
Nepal	Khumaltar	**Study 2**		2001	N.A.	Open top camber	Passive	6.8	296.8	1.1
Nepal	Khumaltar			2002	N.A.	Open top camber	Passive	4.4	296.2	2.3
Nepal	Khumaltar			2003	N.A.	Open top camber	Passive	5.8	296.7	6.6
Nepal	Khumaltar			2004	N.A.	Open top camber	Passive	7.3	296.4	2.1

…

from Zhao et al., 2016

Hierarchical statistical model
« Random-effect model »

Overall mean

Effect of study i

Within-study level: $Y_{ij} = \mu + b_i + \varepsilon_{ij}$ $\varepsilon_{ij} \sim N\left(0, \sigma_{\varepsilon i}^2\right)$

Yield sensitivity in
study i, replicate j

Residual for jth data in study i

Hierarchical statistical model
« Random-effect model »

Overall mean

Effect of study i

Within-study level: $$Y_{ij} = \mu + b_i + \varepsilon_{ij} \qquad \varepsilon_{ij} \sim N\left(0, \sigma_{\varepsilon i}^2\right)$$

Yield sensitivity in study i, replicate j

Residual for jth data in study i

Between-study level: $$b_i \sim N\left(0, \sigma_b^2\right)$$

Prior: Gaussian and Gamma

Hierarchical statistical model
« Random-effect model »

Overall mean
Effect of study i

Within-study level: $$Y_{ij} = \mu + b_i + \varepsilon_{ij} \qquad \varepsilon_{ij} \sim N\left(0, \sigma_{\varepsilon i}^{2}\right)$$

Yield sensitivity in study i, replicate j

Residual for jth data in study i

Between-study level: $$b_i \sim N\left(0, \sigma_b^{2}\right)$$

Prior: Gaussian and Gamma

Fitting algorithm: MCMC

Hierarchical statistical model (with covariate)
« Random-effect model »

Covariate value for study i replicate j

Within-study level: $Y_{ij} = \mu + \alpha X_{ij} + b_i + \varepsilon_{ij}$ $\qquad \varepsilon_{ij} \sim N\left(0, \sigma_{\varepsilon i}^2\right)$

Yield sensitivity in
study i, replicate j

Between-study level: $b_i \sim N\left(0, \sigma_b^2\right)$

Prior: Gaussian and Gamma

Fitting algorithm: MCMC

A simpler model (frequently inappropriate): « Fixed-effect model »

Within-study level: $Y_{ij} = \mu + \varepsilon_{ij}$ $\varepsilon_{ij} \sim N\left(0, \sigma_{\varepsilon i}^{2}\right)$

Yield sensitivity in
study i, replicate j

Prior: Gaussian and Gamma

Fitting algorithm: MCMC

nature plants

Letter | Published: 19 December 2016

Plausible rice yield losses under future climate warming

Chuang Zhao, Shilong Piao, Xuhui Wang, Yao Huang, Philippe Ciais, Joshua Elliott, Mengtian Huang, Ivan A. Janssens, Tao Li, Xu Lian, Yongwen Liu, Christoph Müller, Shushi Peng, Tao Wang, Zhenzhong Zeng & Josep Peñuelas

Compilation of results of 83 field warming experiments located in 14 sites in the world

Field experiment Ambient CO_2

Hierarchical statistical model
« Random-effect model (1) »

Overall mean

Effect of study i

Within-study level: $S_{ij} = \mu + b_i + \varepsilon_{ij}$ $\qquad \varepsilon_{ij} \sim N\left(0, \sigma^2_{\varepsilon i}\right)$

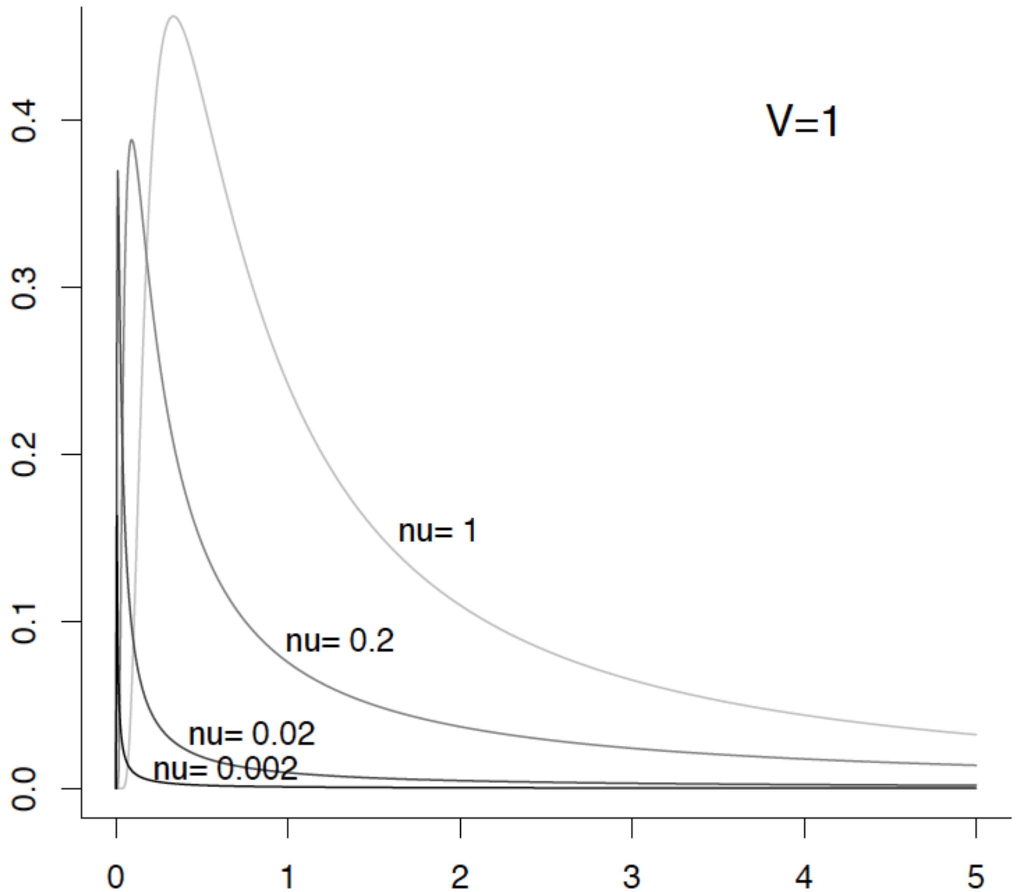
Yield sensitivity in study i, replicate j

Residual for jth data in study i

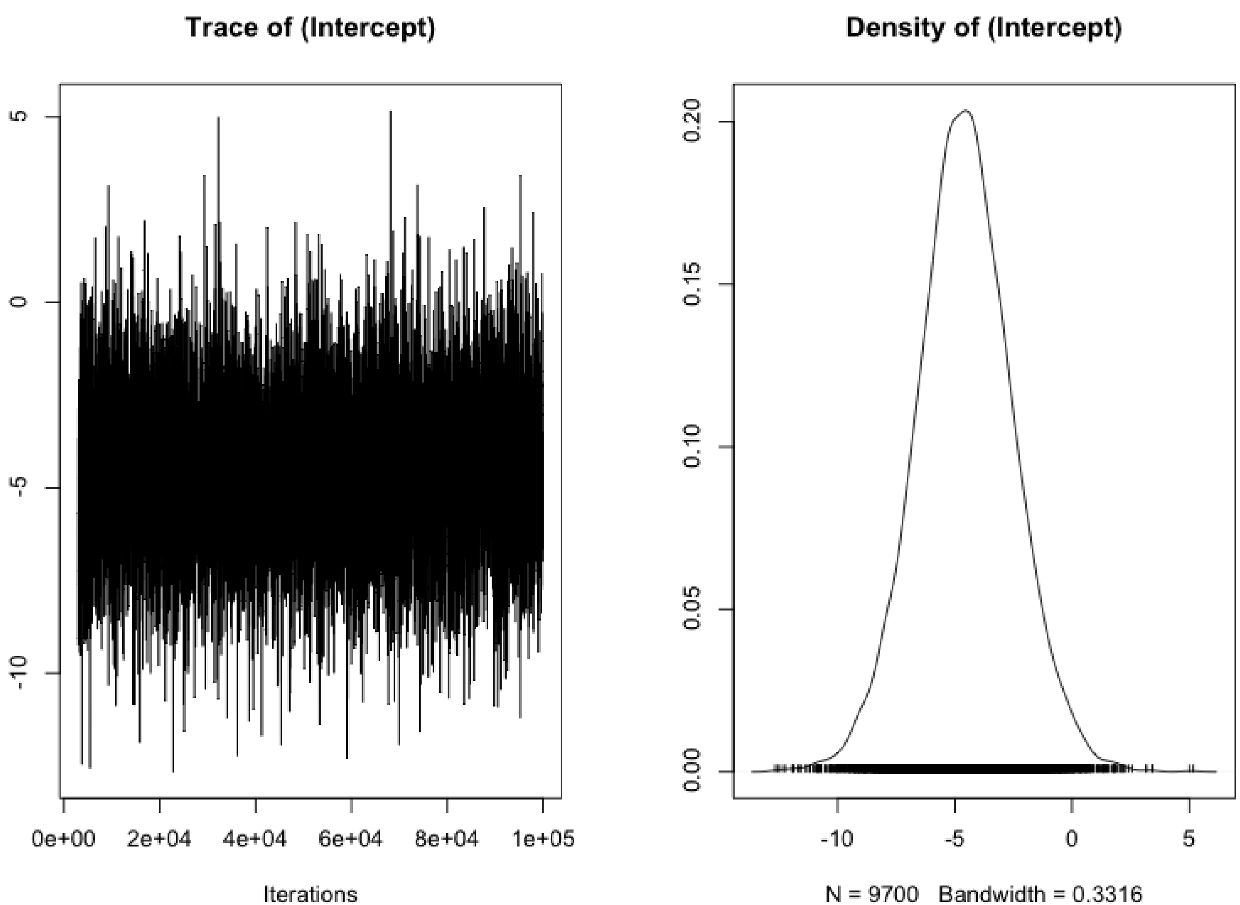
Between-study level: $b_i \sim N\left(0, \sigma^2_b\right)$

Prior: Gaussian and InvGamma

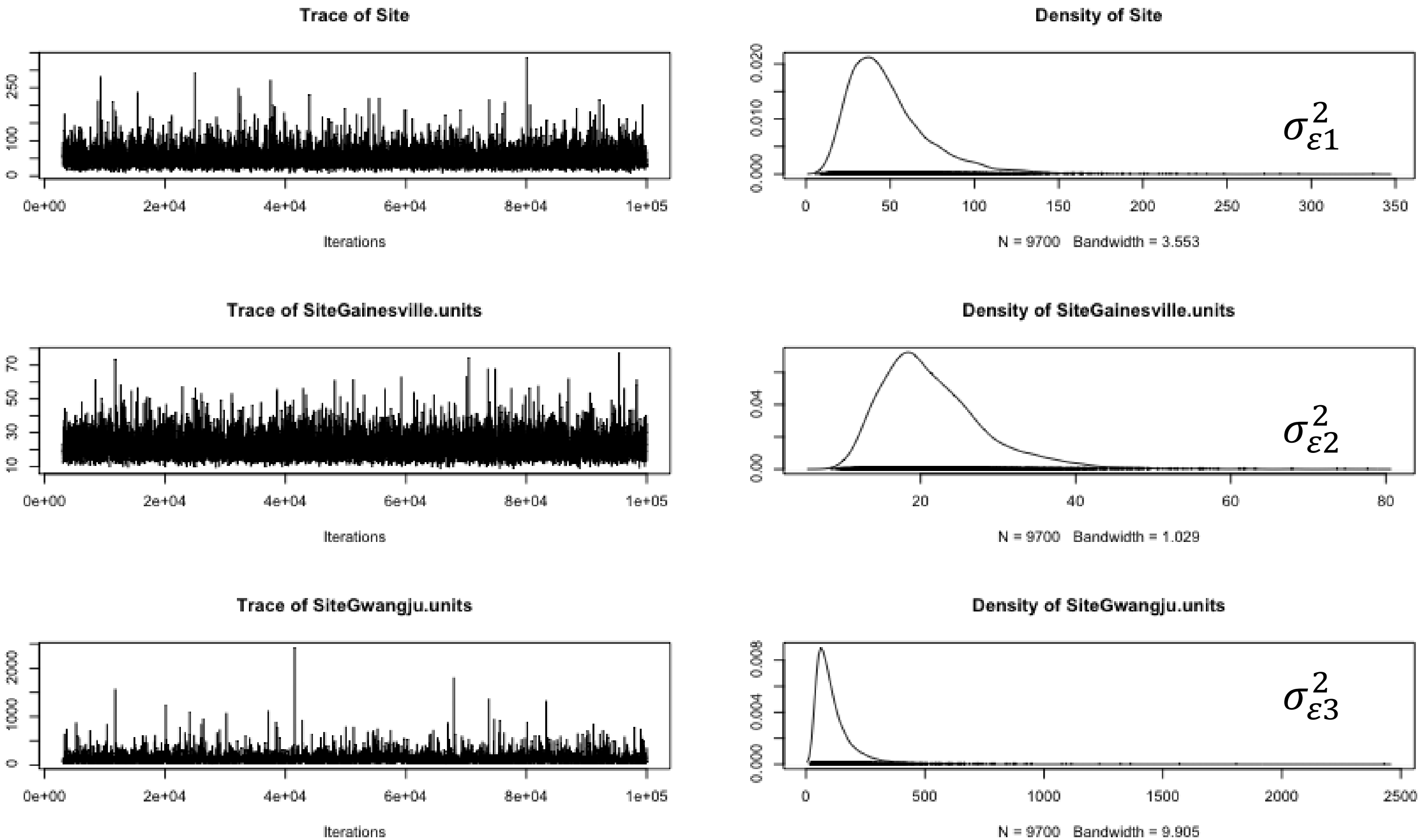
$$\left\{\begin{array}{l} \mu \sim N(mu, V) \\ \sigma^2_{\varepsilon i} \sim InvGamma(scaleR, shapeR) \\ \sigma^2_b \sim InvGamma(scaleB, shapeB) \end{array}\right.$$



scale = nu/2
shape=V*nu/2

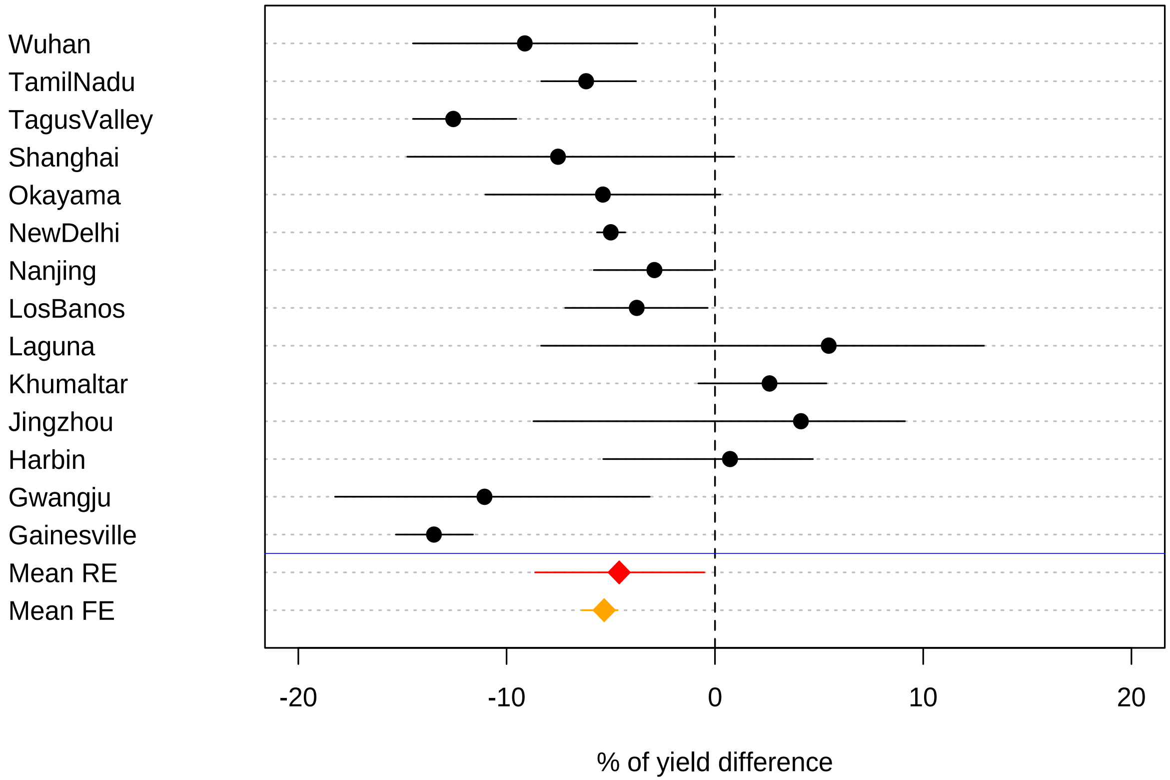

Trace of (Intercept)
5
0
-5
-10
0e+00
2e+04
4e+04
6e+04
8e+04
1e+05
Iterations
Density of (Intercept)
0.20
0.15
0.10
0.05
0.00
-10
-5
0
5
N = 9700 Bandwidth = 0.3316

μ



Trace of Site
Density of Site
Iterations
N = 9700 Bandwidth = 3.553
$\sigma^2_{\varepsilon 1}$
Trace of SiteGainesville.units
Density of SiteGainesville.units
Iterations
N = 9700 Bandwidth = 1.029
$\sigma^2_{\varepsilon 2}$
Trace of SiteGwangju.units
Density of SiteGwangju.units
Iterations
N = 9700 Bandwidth = 9.905
$\sigma^2_{\varepsilon 3}$

Meta-analysis of field warming experiments: Rice yield sensitivity to +1° C (ambient $[CO_2]$)



nature COMMUNICATIONS

Article | OPEN | Published: 17 November 2016

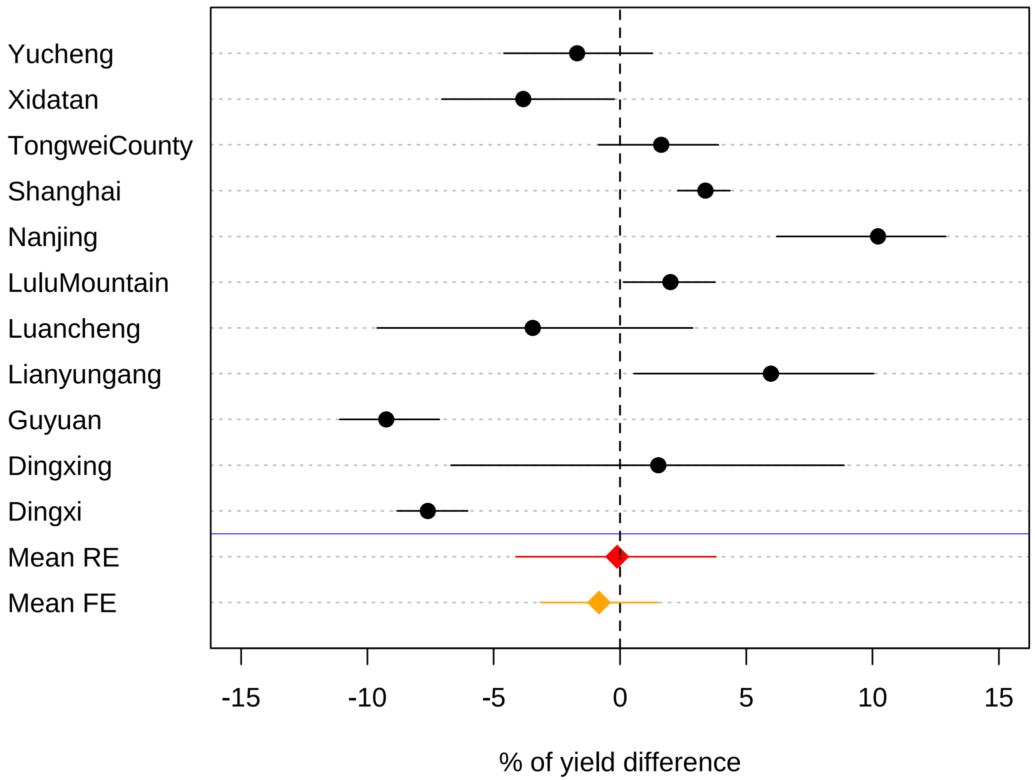
Field warming experiments shed light on the wheat yield response to temperature in China

Chuang Zhao, Shilong Piao, Yao Huang, Xuhui Wang, Philippe Ciais, Mengtian Huang, Zhenzhong Zeng & Shushi Peng

Compilation of 46 results of field warming experiments located in 11 sites in China

Field experiment
Ambient CO_2

Meta-analysis of field warming experiments: Wheat yield sensitivity to +1° C (ambient [CO_2])



Hierarchical statistical model (with covariate)
« Random-effect model »

Mean temperature during the growing season for study i replicate j

Within-study level: $$Y_{ij} = \mu + \alpha X_{ij} + b_i + \varepsilon_{ij} \qquad \varepsilon_{ij} \sim N\left(0, \sigma_{\varepsilon i}^2\right)$$

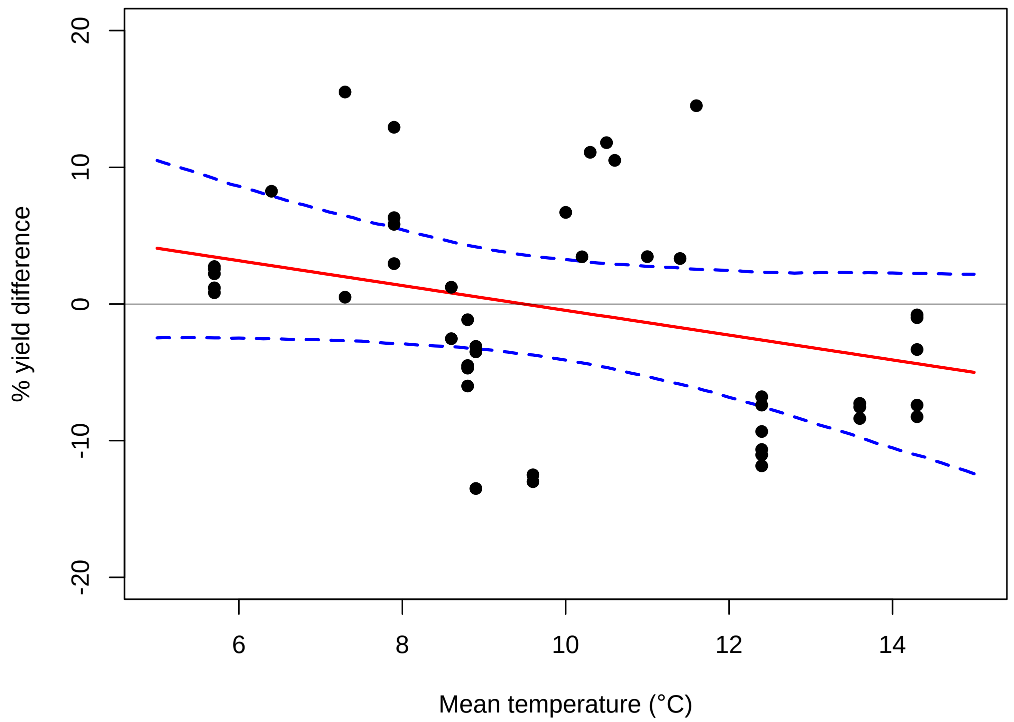
Yield sensitivity in
study i, replicate j

Between-study level: $$b_i \sim N\left(0, \sigma_b^2\right)$$

Prior: Gaussian and Gamma

Fitting algorithm: MCMC

Meta-regression:
Wheat yield sensitivity vs. Mean temperature



Meta-analysis with
Bayesian generalized random effect linear models

Eur J Plant Pathol (2014) 139:79–94
DOI 10.1007/s10658-013-0365-6

Comparison of statistical models in a meta-analysis of fungicide treatments for the control of citrus black spot caused by *Phyllosticta citricarpa*

**D. Makowski · A. Vicent · M. Pautasso ·
G. Stancanelli · T. Rafoss**

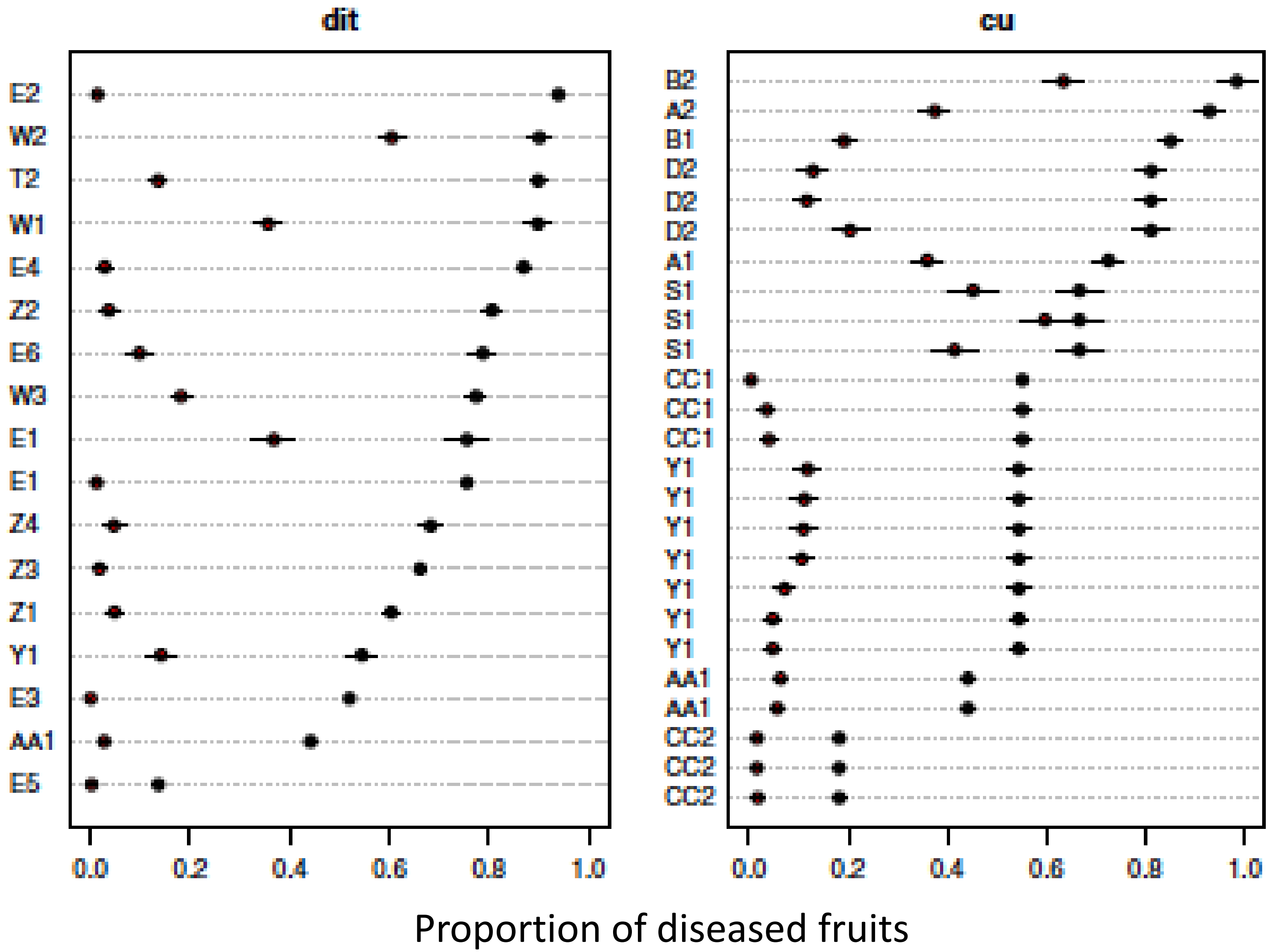
Untreated plot

n_0 fruits
y_0 diseased fruits

Treated plot

n_T fruits
y_T diseased fruits

Effect of fungicide treatments on citrus black spot incidence



Makowski et al., 2014

GLM

$$Y_{ij} \sim Binomial\left(n_{ij}, \pi_{ij}\right)$$

$$\operatorname{logit}\left(\pi_{ij}\right) = log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \alpha_0 + \alpha_T X_{ij}$$

GLM

Proba of diseased fruit in site i, plot j

$$Y_{ij} \sim Binomial(n_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \alpha_0 + \alpha_T X_{ij}$$

0 (untreated), 1 (treated)

GLM

Proba of diseased fruit in site i, plot j

$$Y_{ij} \sim Binomial(n_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \alpha_0 + \alpha_T X_{ij}$$

Log Odds ratio

GLMM

$$Y_{ij} \sim Binomial \quad (n_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \alpha_{0i} + \alpha_{Ti} X_{ij}$$

$$\begin{pmatrix} \alpha_{0i} \\ \alpha_{Ti} \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_0 \\ \mu_T \end{pmatrix}, \Sigma\right]$$

$$\Sigma = \begin{bmatrix} \sigma_0^2 & c \\ c & \sigma_T^2 \end{bmatrix}$$

GLMM

$$Y_{ij} \sim Binomial \quad (n_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \alpha_{0i} + \alpha_{Ti}X_{ij}$$

$$\begin{pmatrix} \alpha_{0i} \\ \alpha_{Ti} \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_0 \\ \mu_T \end{pmatrix}, \Sigma\right]$$

$$\Sigma = \begin{bmatrix} \sigma_0^2 & c \\ c & \sigma_T^2 \end{bmatrix}$$

GLMM

$$Y_{ij} \sim Binomial \quad (n_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \alpha_{0i} + \alpha_{Ti} X_{ij}$$

$$\begin{pmatrix} \alpha_{0i} \\ \alpha_{Ti} \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_0 \\ \mu_T \end{pmatrix}, \Sigma\right]$$

$$\Sigma = \begin{bmatrix} \sigma_0^2 & c \\ c & \sigma_T^2 \end{bmatrix}$$

Prior $\mu_0, \mu_T \sim N(0, 10^6)$

$\Sigma \sim InvWish(\psi, v)$ with $\psi = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and v=2

GLMM

$$Y_{ij} \sim Binomial \quad (n_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \alpha_{0i} + \alpha_{Ti} X_{ij}$$

$$\begin{pmatrix} \alpha_{0i} \\ \alpha_{Ti} \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_0 \\ \mu_T \end{pmatrix}, \Sigma\right]$$

Proba of diseased fruit (untreated) $\frac{exp(\alpha_{0i})}{1+exp(\alpha_{0i})}$

Proba of diseased fruit (treated) $\frac{exp(\alpha_{0i}+\alpha_{Ti})}{1+exp(\alpha_{0i}+\alpha_{Ti})}$

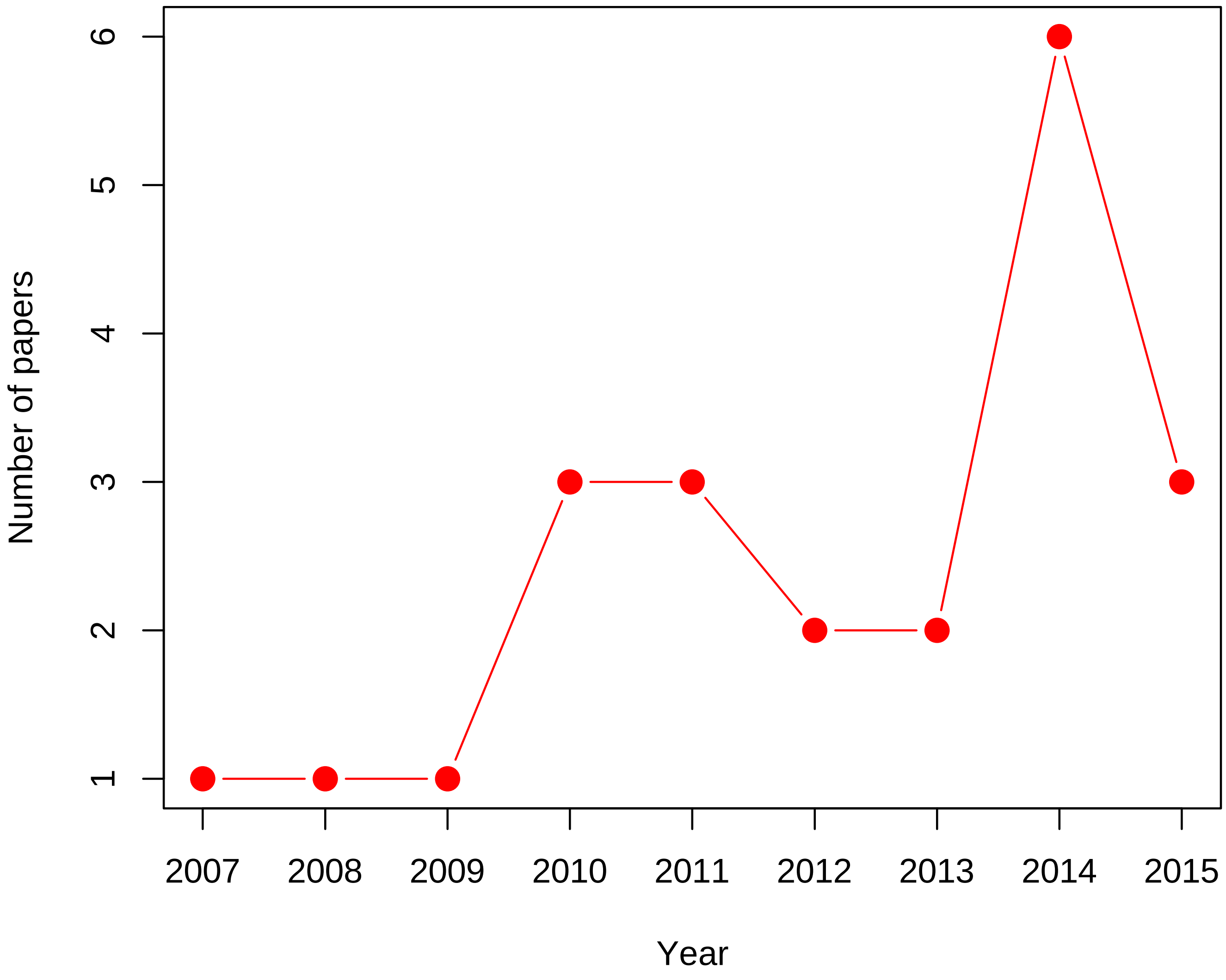
Type of fungicide	Probability of diseased fruit	
	Untreated fruits	Treated fruits
"cu"	0.74 (0.53-0.87)	0.16 (0.072-0.32)
"dit"	0.74 (0.62-0.83)	0.053 (0.022-0.12)

Makowski et al., 2014

Search from Web of Science Core Collection (2015)

TOPIC:
((meta-analy* OR (meta AND analy*)) AND bayes* AND (agronomy OR agriculture OR plant OR ecology))

- 41 documents found
- 22 documents showed applications of Bayesian methods



Number of papers
1
2
3
4
5
6
2007
2008
2009
2010
2011
2012
2013
2014
2015
Year

Area of science	Number of papers
Ecology	16
Plant pathology	2
Agronomy	2
Plant biology/Genetics	2

Software	Number of papers
winBUGS/openBUGS	10
R packages	6
JAGS	2
Python	1
Not specified	3

Advantages of Bayesian methods according to the reviewed papers

Advantage	Number of papers
No advantage mentioned	6
Take into account the hierarchical structure of the data	12
Accomodate missing data and expand the size of the dataset	8
Explicitly quantify uncertainty	2
Incorporate prior information	1

Conclusion

- Meta-analyses can be easily performed using Bayesian methods
- Bayesian methods potentially useful to
  - Handle missing data
  - Increase the size of the dataset
  - Deal with complex dataset structure
  - Deal with uncertainty