Paris, 2016

Ranking crop species from direct and indirect evidences

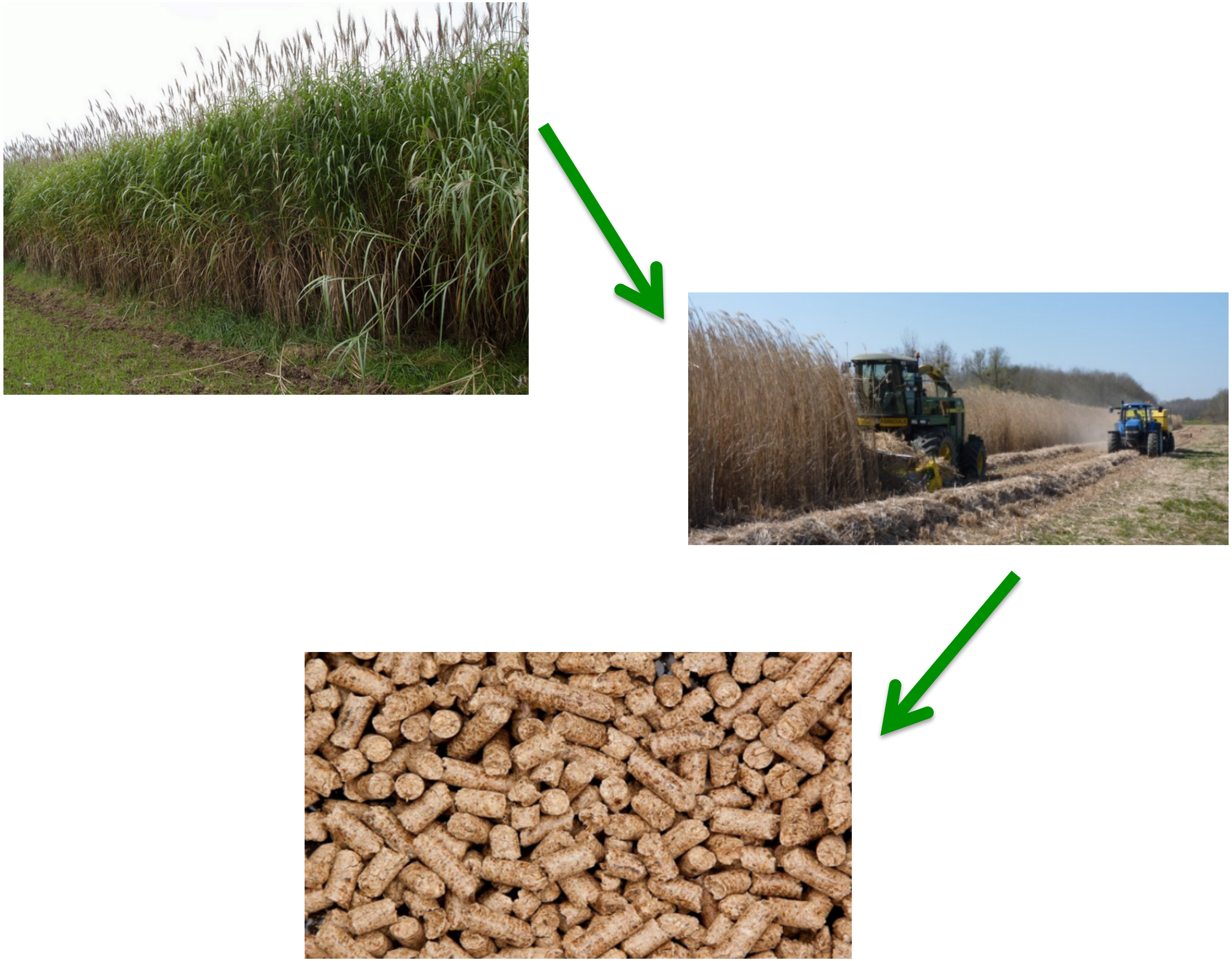
D. Makowski[1], I. Albert[2]

[1] UMR Agronomie INRA AgroParisTech Paris-Saclay

[2]UMR MIA INRA AgroParisTech Paris-Saclay

Which species is the most productive?

One experiment comparing two species



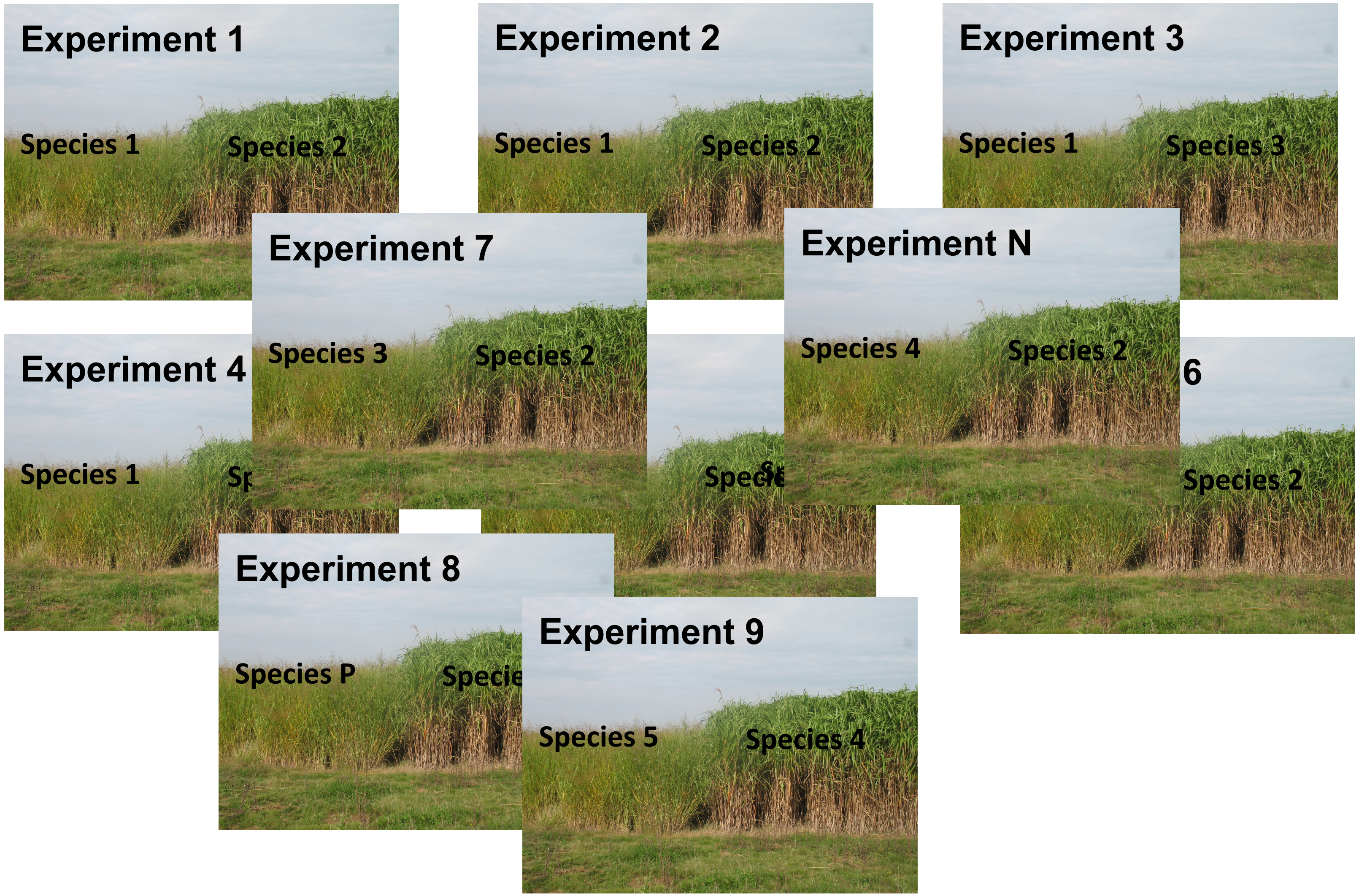
One experiment comparing two species

Yield of species 1

Yield of species 2

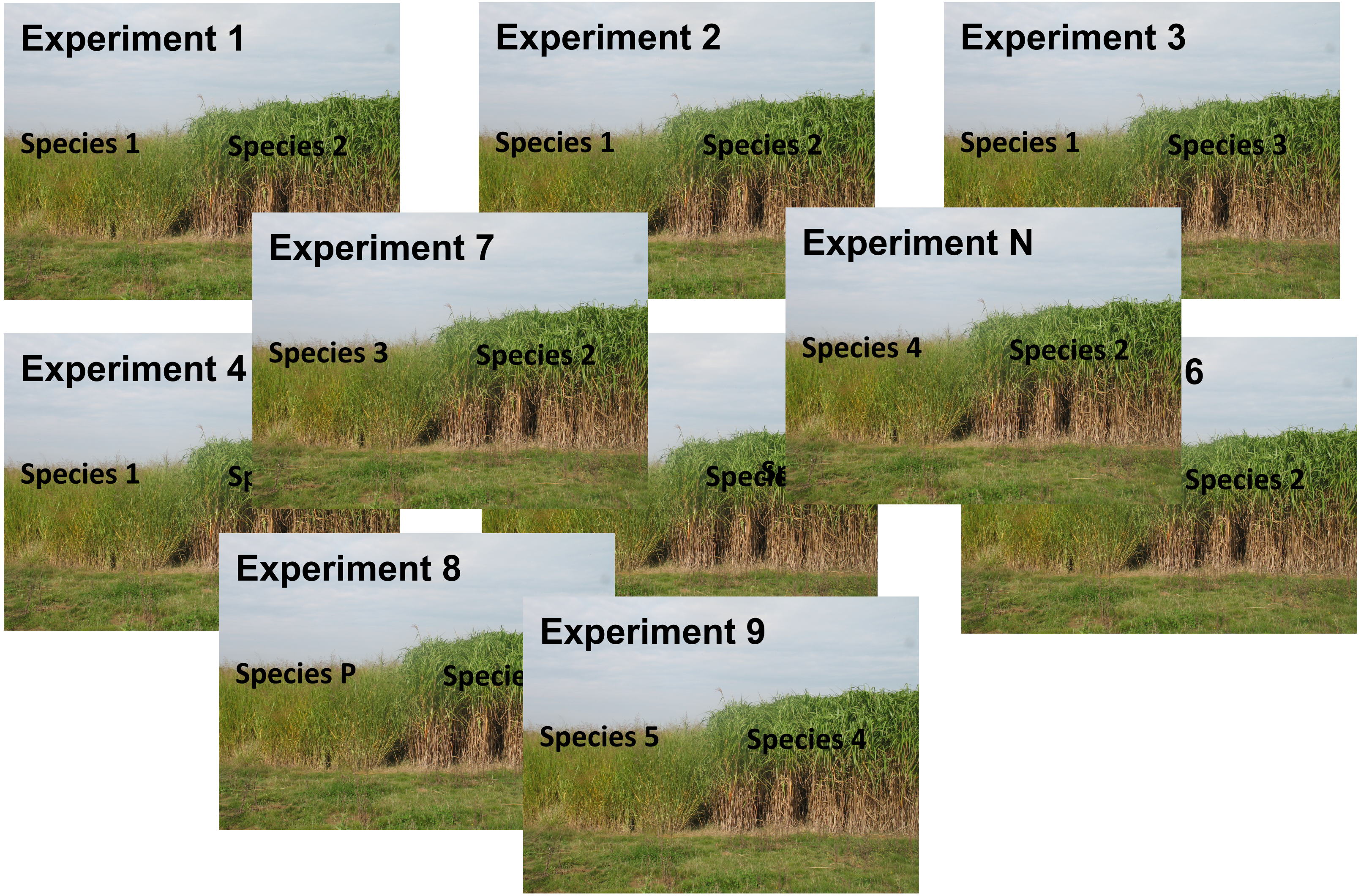
Two experiments comparing two species

Dataset including *N* experiments comparing *P* species



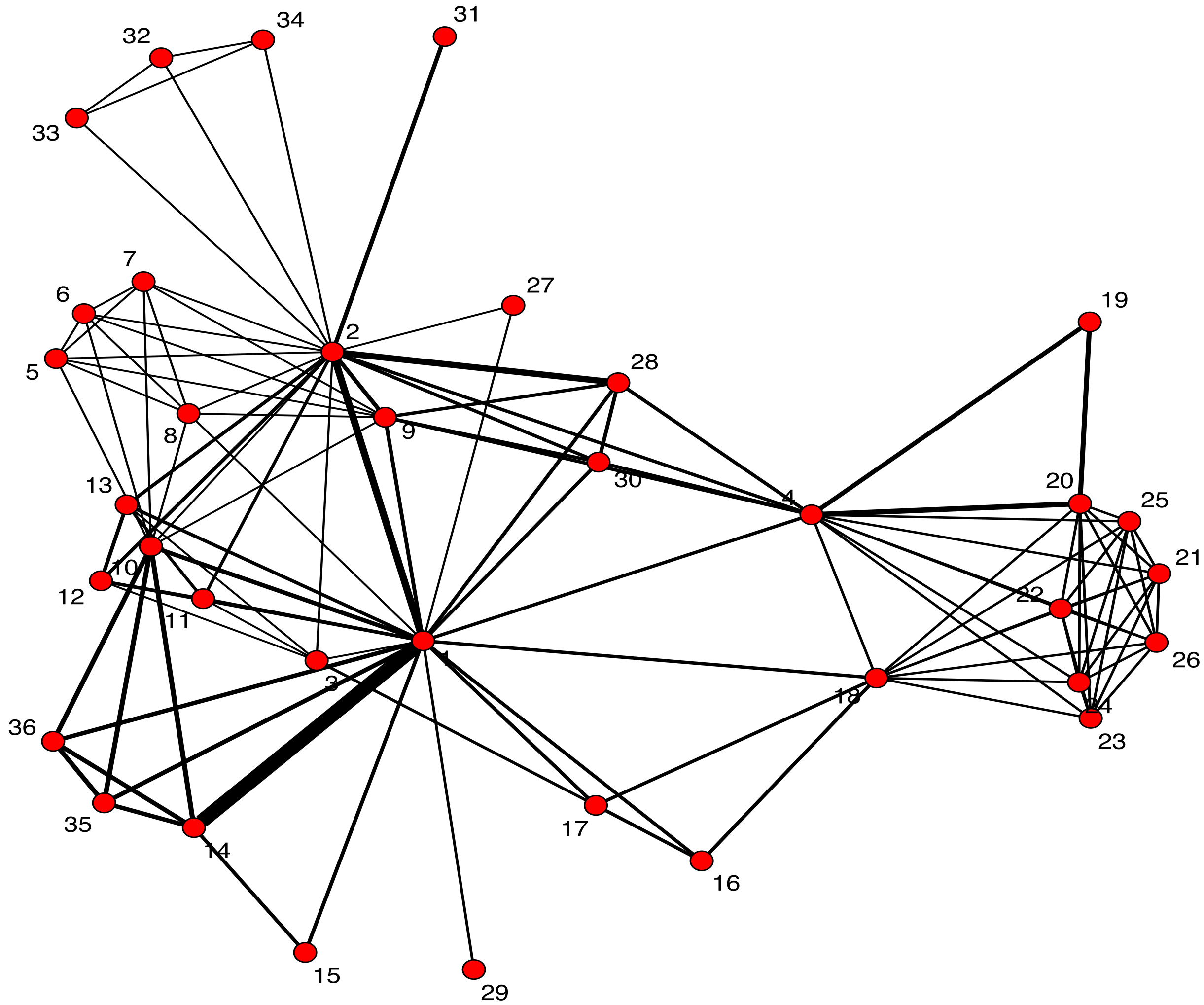
How could we rank the P species according to their productivity from such dataset?

Dataset including *N* experiments comparing *P* species



Network of experimental data

28 papers, 67 site-years, 36 species, 639 yield data



1: Miscanthus x giganteus
2: Panicum virgatum
3: Salix
4: Triticosecale
5: Erianthus
6: Sorghum halepense
7: Saccharum officinarum
8: Zea mays
9: Sorghum bicolor
10: Pennisetum purpureum
11: Phalaris arundinacea
12: Miscanthus sinensis
13: Phragmites australis
14: Arundo donax
15: Cynara cardunculus
16: Miscanthus sacchariflorus
17: Sida hermaphrodita
18: Salix viminalis
19: Triticum aestivum
20: Secale cereale

Meta-analysis to estimate yield ratio by direct comparison (Laurent et al. 2015)

Linear random-effects model:

$$\log(Y_{ij}) = \mu_{ref} + \alpha_i + b_j + \varepsilon_{ij}$$

Y_{ij} = Yield of the crop *i* for the « site-year » *j*

μ_{ref} = Average yield for the reference crop

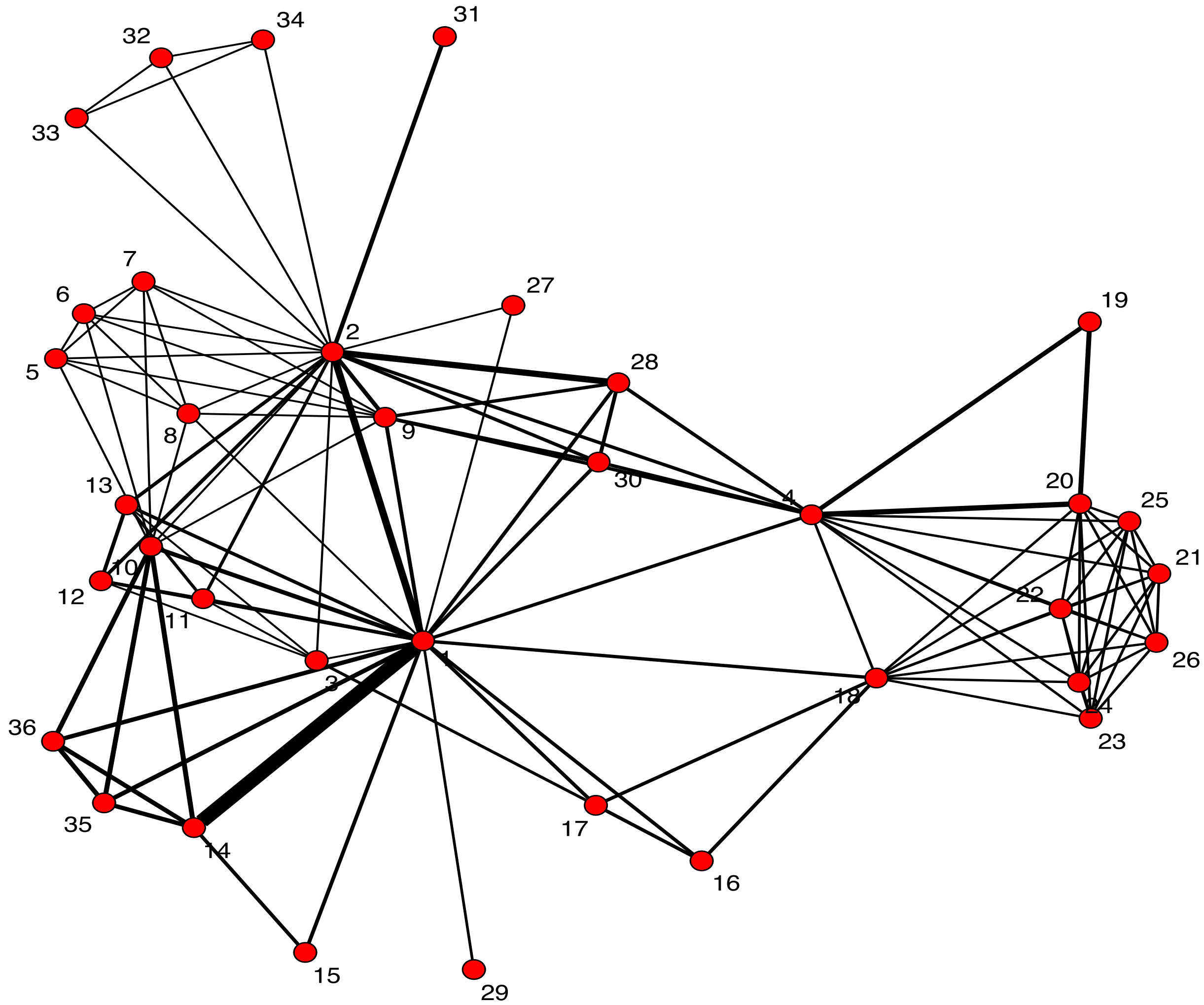
α_i = Fixed effect (*i* = crop index)

b_j = Random effect (*j* = site-year index)

ε_{ij} = Residual

Network of experimental data

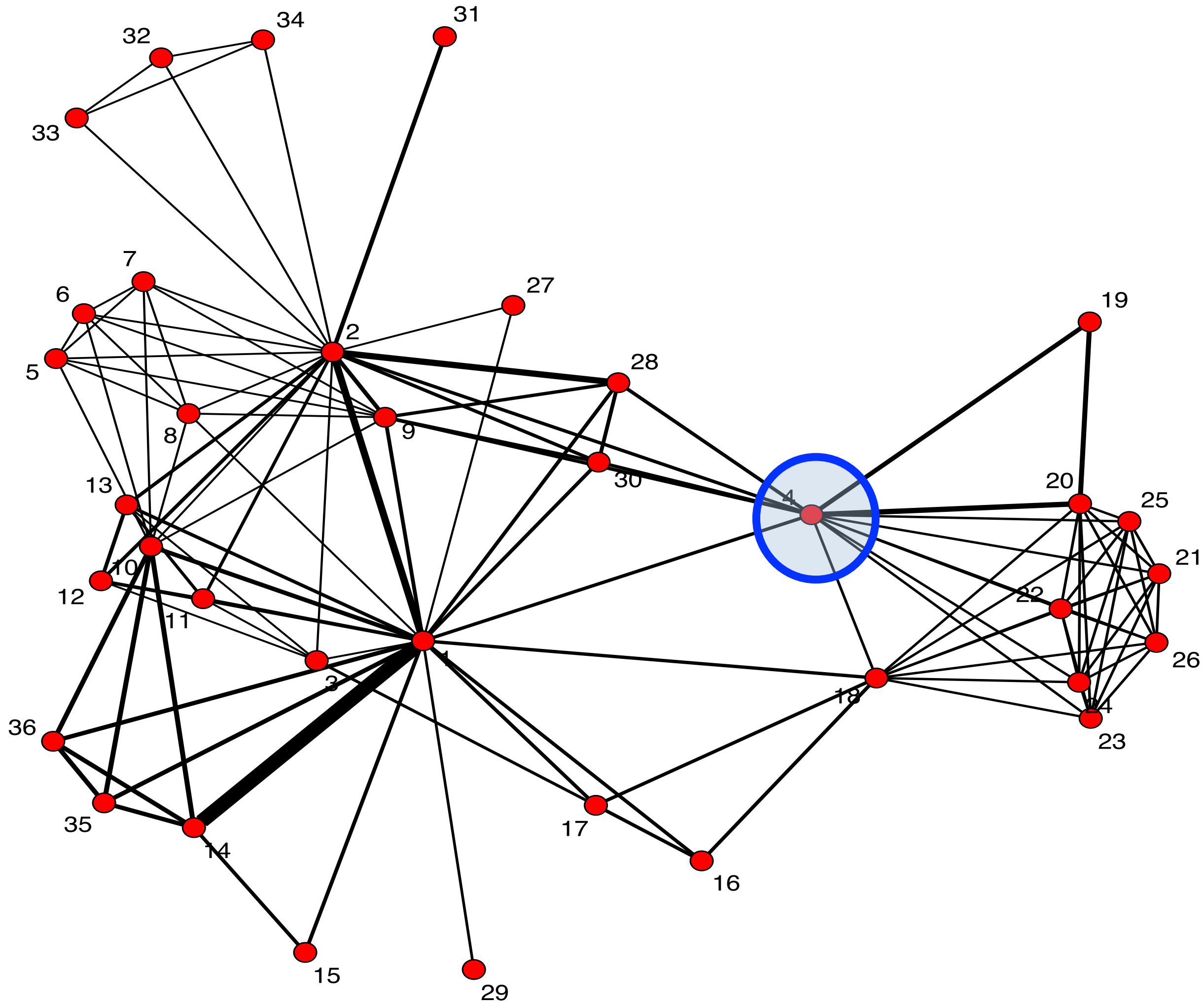
28 papers, 67 site-years, 36 species, 639 yield data



1: Miscanthus x giganteus
2: Panicum virgatum
3: Salix
4: Triticosecale
5: Erianthus
6: Sorghum halepense
7: Saccharum officinarum
8: Zea mays
9: Sorghum bicolor
10: Pennisetum purpureum
11: Phalaris arundinacea
12: Miscanthus sinensis
13: Phragmites australis
14: Arundo donax
15: Cynara cardunculus
16: Miscanthus saccharifloru
17: Sida hermaphrodita
18: Salix viminalis
19: Triticum aestivum
20: Secale cereale

Network of experimental data

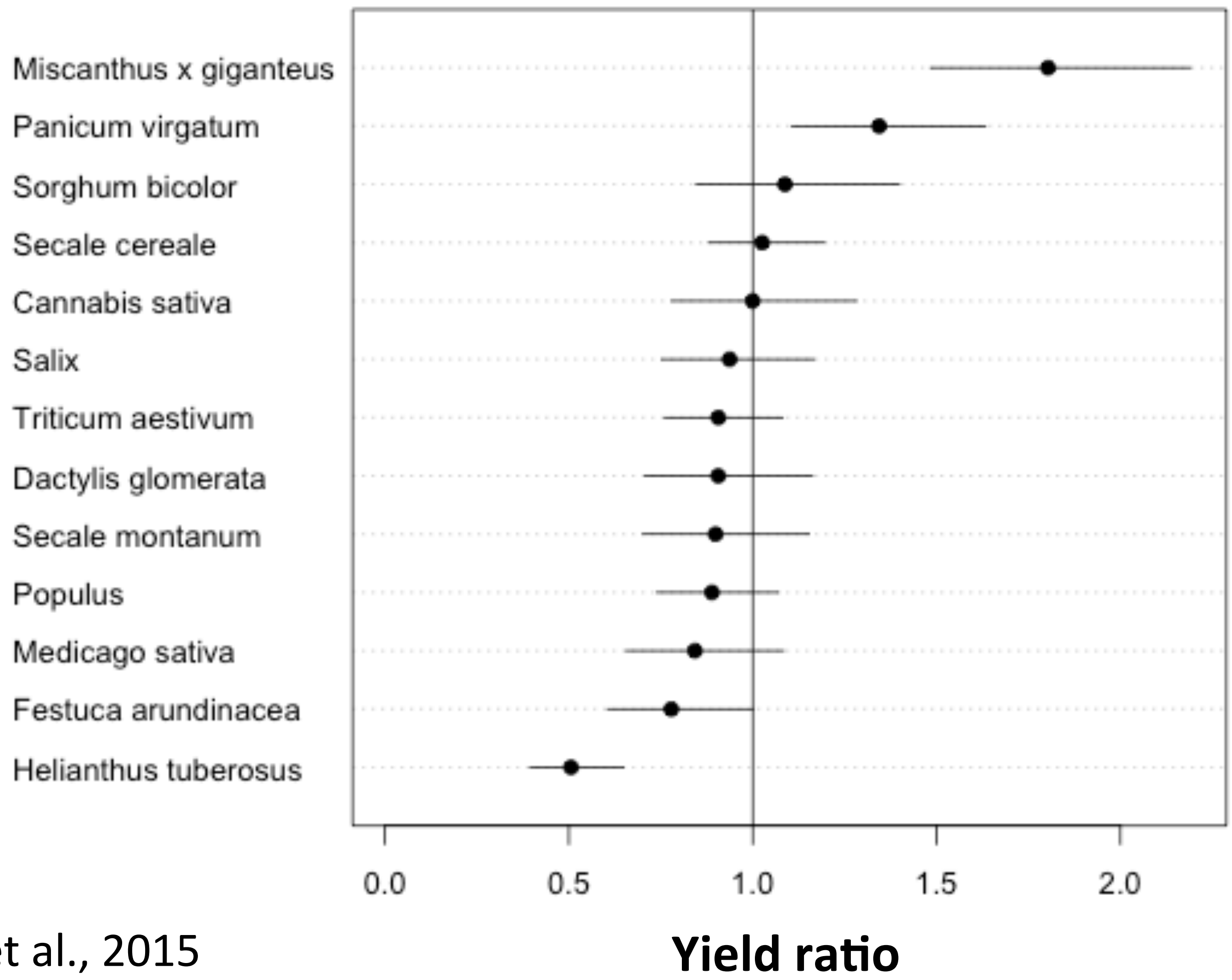
28 papers, 67 site-years, 36 species, 639 yield data



1: Miscanthus x giganteus
2: Panicum virgatum
3: Salix
4: Triticosecale
5: Erianthus
6: Sorghum halepense
7: Saccharum officinarum
8: Zea mays
9: Sorghum bicolor
10: Pennisetum purpureum
11: Phalaris arundinacea
12: Miscanthus sinensis
13: Phragmites australis
14: Arundo donax
15: Cynara cardunculus
16: Miscanthus sacchariflorus
17: Sida hermaphrodita
18: Salix viminalis
19: Triticum aestivum
20: Secale cereale

Yield ratio of crop species compared to *Triticosecale*

(Ratio=Yield species *X* / Yield triticosecale)



Laurent et al., 2015

Indirect comparison

Ex: switchgrass *vs*. alfalfa

Experiment 1

Experiment 2

$$\frac{\text{Yield switchgrass}}{\text{Yield } \textit{Miscanthus}}$$

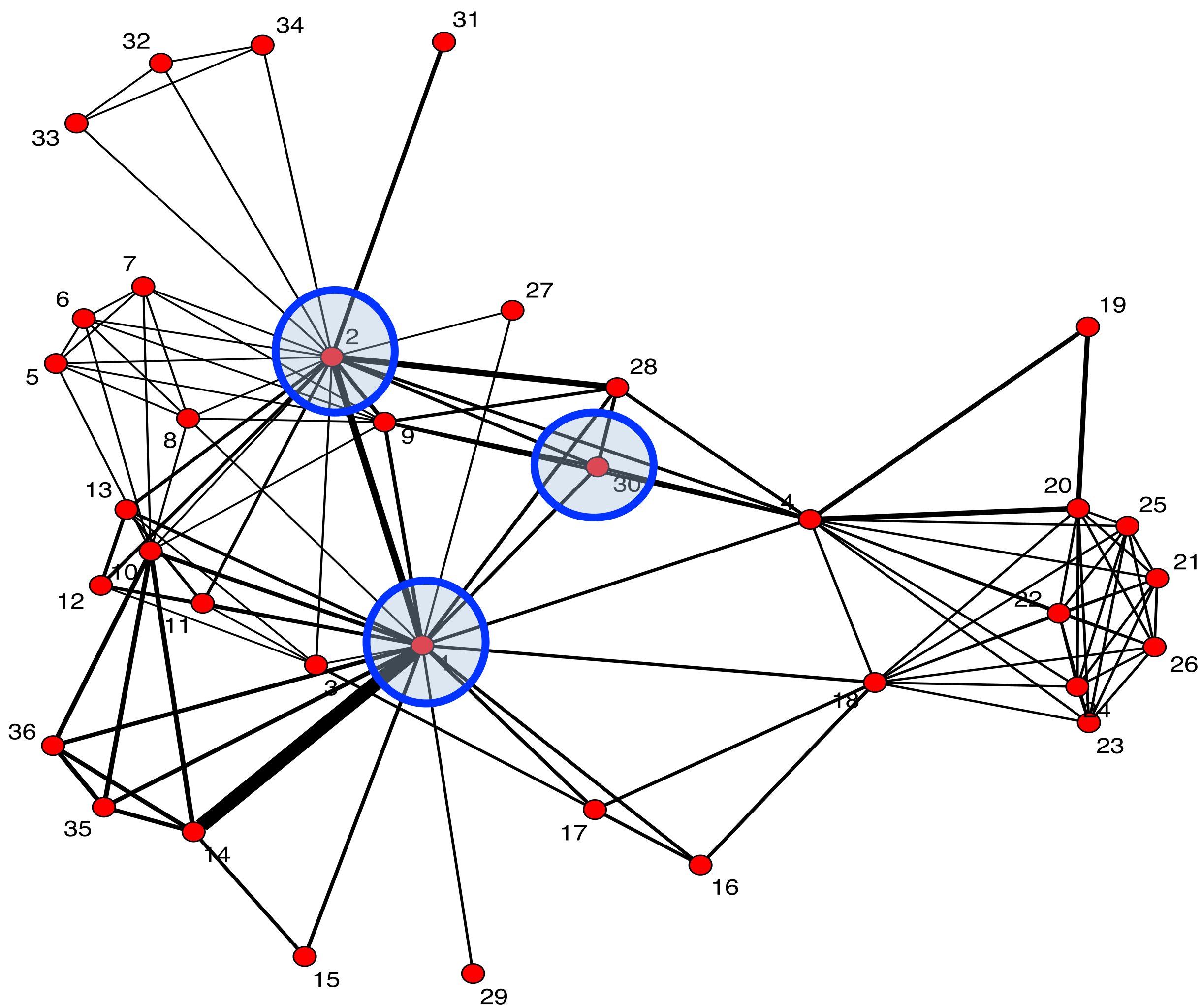
$$\frac{\text{Yield alfalfa}}{\text{Yield } \textit{Miscanthus}}$$

→

$$\frac{\text{Yield Alfalfa}}{\text{Yield switchgrass}}$$

Network of experimental data

28 papers, 67 site-years, 36 species, 639 yield data



1: Miscanthus x giganteus
2: Panicum virgatum
3: Salix
4: Triticosecale
5: Erianthus
6: Sorghum halepense
7: Saccharum officinarum
8: Zea mays
9: Sorghum bicolor
10: Pennisetum purpureum
11: Phalaris arundinacea
12: Miscanthus sinensis
13: Phragmites australis
14: Arundo donax
15: Cynara cardunculus
16: Miscanthus sacchariflorus
17: Sida hermaphrodita
18: Salix viminalis
19: Triticum aestivum
30: Medicago sativa

Indirect comparison

Ex: switchgrass *vs*. alfalfa

Site-year 1

Site-year 2

$$\frac{\text{Yield switchgrass}}{\text{Yield } \textit{Miscanthus}}$$

$$\frac{\text{Yield alfalfa}}{\text{Yield } \textit{Miscanthus}}$$

→ $$\frac{\text{Yield Alfalfa}}{\text{Yield switchgrass}}$$

Mixed treatment comparison for combining direct and indirect evidence

- Fixed-effect model

- Random-effect model

Adapted versions of the Bayesian models presented by Dias et al. (2010)

Fixed-effect model (model 1)

$$\log(Y_{ijk}) \sim N\left(\mu_{ij}, \sigma^2\right), i = 1, \ldots, N,\ j = 1, \ldots, S, k = 1, \ldots, M_{ij}$$

$$\mu_{ij} = \mu_i^{\mathrm{Ref}_i} + d_{\mathrm{Ref}_i j} \times \mathbf{1}_{j \neq \mathrm{Ref}_i}$$

- Y_{ijk} is the k^{th} yield measured in site-year i for crop species j,
- $\mu_i^{\mathrm{Ref}_i}$ is the mean log-yield value of the reference species Ref_i in site-year i (reference species may differ across site-years),
- $d_{\mathrm{Ref}_i j}$ is the mean effect of species j over site-years relative to the species Ref_i (baseline contrasts)
- σ is the log-yield standard deviation.

Fixed-effect model (model 1)

$$\log(Y_{ijk}) \sim N\left(\mu_{ij},\sigma^2\right),\ i = 1,\ ...,\ N,\ \ j = 1,\ ...,\ S,\ k = 1,\ ...,\ M_{ij}$$

$$\mu_{ij} = \mu_i^{\mathrm{Ref}_i} + d_{\mathrm{Ref}_i j} \times 1_{j \neq \mathrm{Ref}_i}$$

Assumption of consistency (Dias et al., 2010):

$$d_{\mathrm{Ref}_i j} = d_{REFj} - d_{\mathrm{REFRef}_i}$$

- d_{REFj} is the mean effect of species *j* over site-years relative to an overall species baseline noted REF
- d_{REFRef_i} is the mean effect of species Ref_i over site-years relative to the overall species baseline REF

Fixed-effect model (model 1)

$$\log(Y_{ijk}) \sim N\left(\mu_{ij}, \sigma^2\right), i = 1, \ldots, N,\ j = 1, \ldots, S, k = 1, \ldots, M_{ij}$$

$$\mu_{ij} = \mu_i^{\mathrm{Ref}_i} + d_{\mathrm{Ref}_i j} \times 1_{j \neq \mathrm{Ref}_i}$$

Assumption of consistency:

$$d_{\mathrm{Ref}_i j} = d_{REFj} - d_{\mathrm{REFRef}_i}$$

The value of d_{REFj} is set equal to zero for j=1

The contrast between species j' and j: $d_{j'j} = d_{REFj'} - d_{REFj}$

Yield ratio: $R_{j'j} = \exp(d_{j'j}) = \exp(d_{REFj'} - d_{REFj})$

Fixed-effect model (model 1)

$$\log(Y_{ijk}) \sim N\left(\mu_{ij},\sigma^2\right), i = 1, ..., N,\ j = 1, ..., S, k = 1, ..., M_{ij}$$

$$\mu_{ij} = \mu_i^{\mathrm{Ref}_i} + d_{\mathrm{Ref}_i j} \times \mathbf{1}_{j \neq \mathrm{Ref}_i}$$

Assumption of consistency:

$$d_{\mathrm{Ref}_i j} = d_{REFj} - d_{\mathrm{REFRef}_i}$$

$$\mu_i^{\mathrm{Ref}} \sim N\left(0,10^4\right) \text{ for } i = 1, ..., N$$

$$d_{REFj} \sim N\left(0,10^4\right) \text{ for } j = 2, ..., S$$

$$\sigma \sim Unif\left(0,2\right)$$

Random-effect model (model 2)

$$\log(Y_{ijk}) \sim N\left(\mu_{ij}, \sigma^2\right),\ i = 1, ..., N,\ j = 1, ..., S,\ k = 1, ..., M_{ij}$$

$$\mu_{ij} = \mu_i^{\mathrm{Ref}_i} + \delta_{i\mathrm{Ref}_i j} \times 1_{j \neq \mathrm{Ref}_i}$$

$$\delta_{i\mathrm{Ref}_i \mathrm{j}} \sim N\left(d_{\mathrm{Ref}_i j}, \tau^2\right)$$

$$d_{\mathrm{Ref}_i j} = d_{REFj} - d_{\mathrm{REFRef}_i}$$

$$\mu_i^{\mathrm{Ref}} \sim N\left(0, 10^4\right) \text{ for } i = 1, ..., N$$

$$d_{REFj} \sim N\left(0, 10^4\right) \text{ for } j = 2, ..., S$$

$$\sigma \sim Unif\left(0, 2\right)$$

$$\tau \sim Unif\left(0, 2\right)$$

Random-effect model (model 2)

$$\log(Y_{ijk}) \sim N\left(\mu_{ij}, \sigma^2\right),\ i = 1, ..., N,\ j = 1, ..., S,\ k = 1, ..., M_{ij}$$

$$\mu_{ij} = \mu_i^{\mathrm{Ref}_i} + \delta_{i\,\mathrm{Ref}_i j} \times 1_{j \neq \mathrm{Ref}_i}$$

$$\delta_{i\mathrm{Ref}_i \mathrm{j}} \sim N\left(d_{\mathrm{Ref}_i j}, \tau^2\right)$$

$$d_{\mathrm{Ref}_i j} = d_{REFj} - d_{\mathrm{REFRef}_i}$$

$$\mathrm{var}\left(\delta_{ijj'}\right) = \mathrm{var}\left(\delta_{i\mathrm{Ref}_i\mathrm{j'}} - \delta_{i\mathrm{Ref}_i\mathrm{j}}\right) = \mathrm{var}\left(\delta_{i\mathrm{Ref}_i\mathrm{j'}}\right) + \mathrm{var}\left(\delta_{i\mathrm{Ref}_i\mathrm{j}}\right) - 2\,\mathrm{cov}\left(\delta_{i\mathrm{Ref}_i\mathrm{j'}}, \delta_{i\mathrm{Ref}_i\mathrm{j}}\right)$$

$$\Leftrightarrow \mathrm{cov}\left(\delta_{i\mathrm{Ref}_i\mathrm{j'}}, \delta_{i\mathrm{Ref}_i\mathrm{j}}\right) = -\left(\tau^2 - 2\tau^2\right)/\,2 = \tau^2 /\, 2$$

from Dias et al. (2010)

Random-effect model (model 2)

$$\log(Y_{ijk}) \sim N\left(\mu_{ij}, \sigma^2\right), i = 1, \ldots, N,\ j = 1, \ldots, S, k = 1, \ldots, M_{ij}$$

$$\mu_{ij} = \mu_i^{\mathrm{Ref}_i} + \delta_{i\,\mathrm{Ref}_i j} \times 1_{j \neq \mathrm{Ref}_i}$$

$$\delta_{i\mathrm{Ref}_i \mathrm{j}} \sim N\left(d_{\mathrm{Ref}_i j}, \tau^2\right)$$

$$d_{\mathrm{Ref}_i j} = d_{REFj} - d_{\mathrm{REFRef}_i}$$

$$\delta_{i\mathrm{Ref}_i \mathrm{j}} \left\| \begin{pmatrix} \delta_{i\mathrm{Ref}_i 2} \\ \ldots \\ \delta_{i\mathrm{Ref}_i j-1} \end{pmatrix} \right. \sim N\left(d_{REFj} - d_{REF\mathrm{Ref}_i} + \frac{1}{j-1}\sum_{s=1}^{j-1}\left[\delta_{i\mathrm{Ref}_i s} - (d_{REFs} - d_{REF\mathrm{Ref}_i})\right], \frac{j}{2(j-1)}\tau^2\right)$$

from Dias et al. (2010)

Random-effect model with species-specific residual standard deviations (model 3)

$$\sigma_j \sim Unif\left(0, 2\right) \qquad j = 1, \ldots, S$$

OpenBUGS

- MCMC simulations implemented with the OpenBUGS 3.2.3 software
- Three chains run until convergence
  - 20 000 iterations for model 1
  - 200 000 for model 2
  - 250 000 for model 3
- 10 000, 100 000, and 125 000 additional iterations

Model evaluation (deviance)

$$D = (\log(Y) - \mu)^2/\sigma^2$$

$$p_{valk} = Pr\left[D_k^{rep} > D_k^{obs}\right]$$

$$PPP = Pr\left[\sum_k D_k^{rep} > \sum_k D_k^{obs}\right]$$

Model evaluation (node splitting)

The value of $d_{jj'}$ is estimated in three different ways:

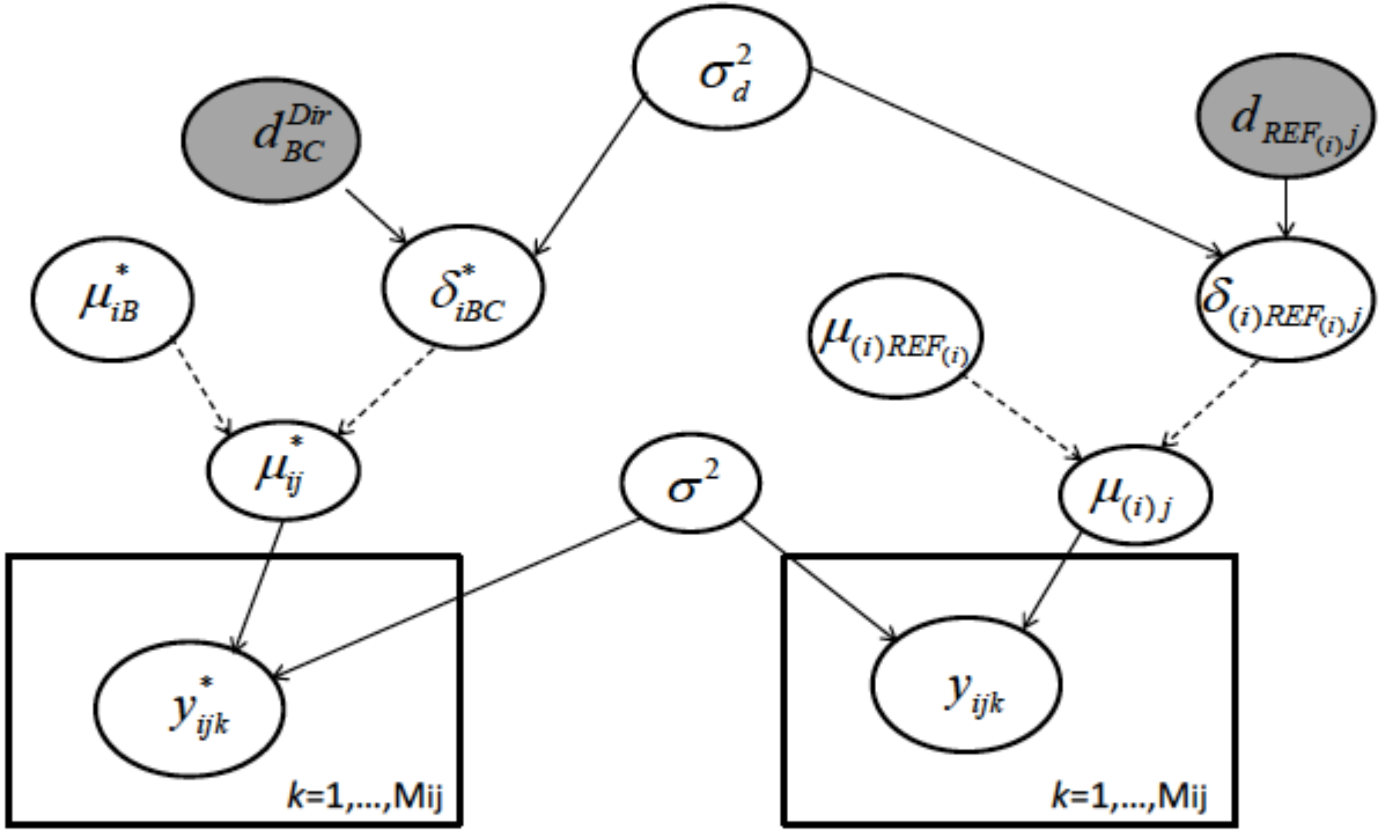
- from direct evidence only,
- from indirect evidence only,
- from both types of evidence with the MTC model.

The three posterior distributions of $d_{jj'}$ are compared graphically in order to detect possible inconsistencies.

Model evaluation (node splitting)

Direct : only studies which compare BC, j = B ou C

MTC excluding direct evidence on BC



$$\delta_{iBC}^* \sim N(d_{BC}^{Dir}, \sigma_d^{*2})$$

$$\delta_{(i)BC} \sim N(d_{BC}^{Ind}, \sigma_d^2)$$
$$\sim N(d_{AC} - d_{AB}, \sigma_d^2)$$

Comparison with a two-way mixed model (model 4) (adapted from Piepho et al., 2012; Madden et al., 2016)

$$\log(Y_{ijk}) \sim N\left(\mu_{ij}, \sigma^2\right)$$

$$\mu_{ij} = \theta + \beta_i + \gamma_j + u_{ij}$$

$$\beta_i \sim N\left(0, \sigma_\beta^2\right) \qquad u_{ij} \sim N\left(0, \sigma_u^2\right)$$

θ is the mean yield of the baseline species,

β_i is a site-year random effect,

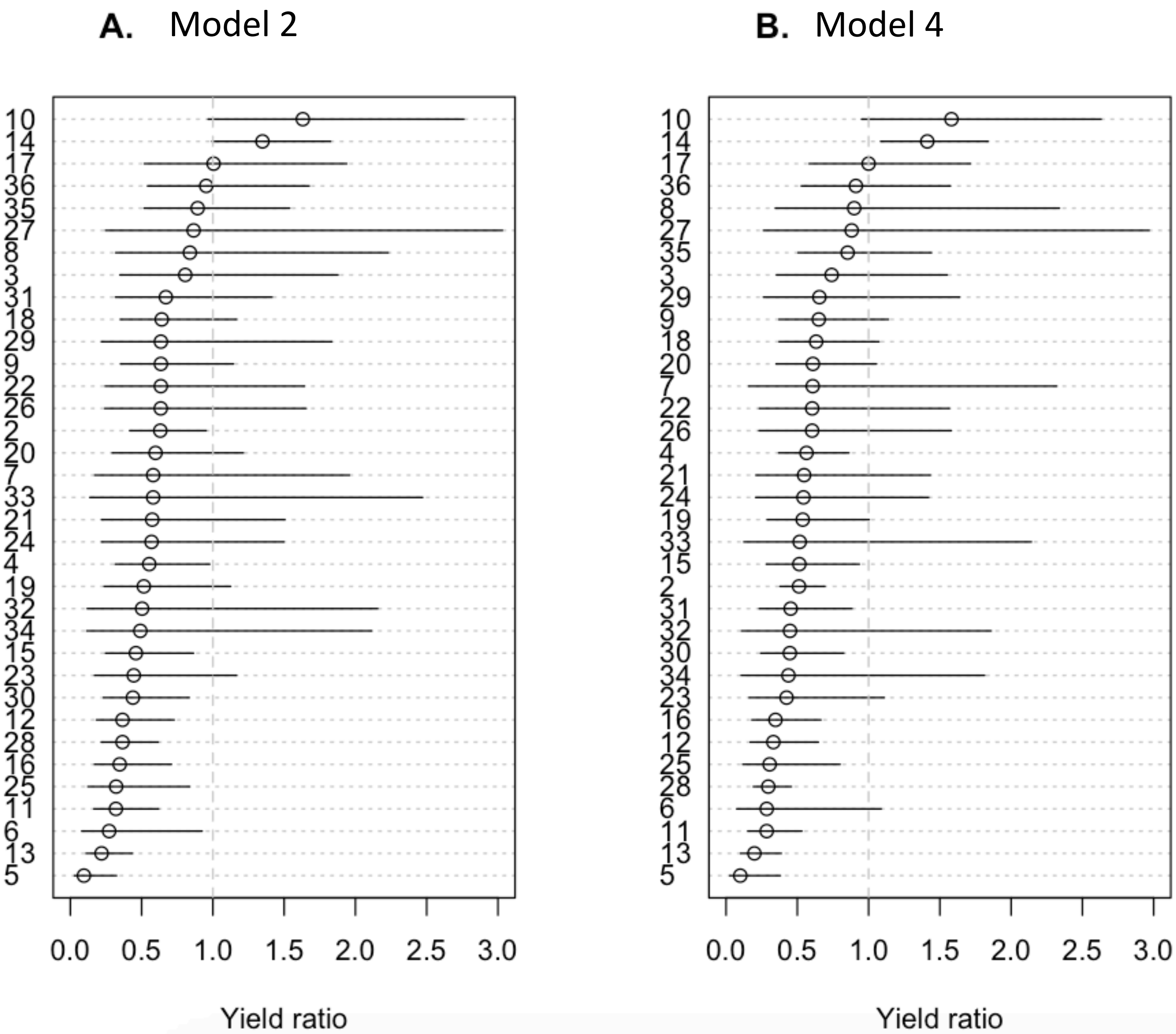
γ_j is the fixed main effect of the j^{th} species compared to the baseline species,

u_{ij} is a random effect describing the between site-year variability of the effect of the j^{th} species (interaction between species and site-year).

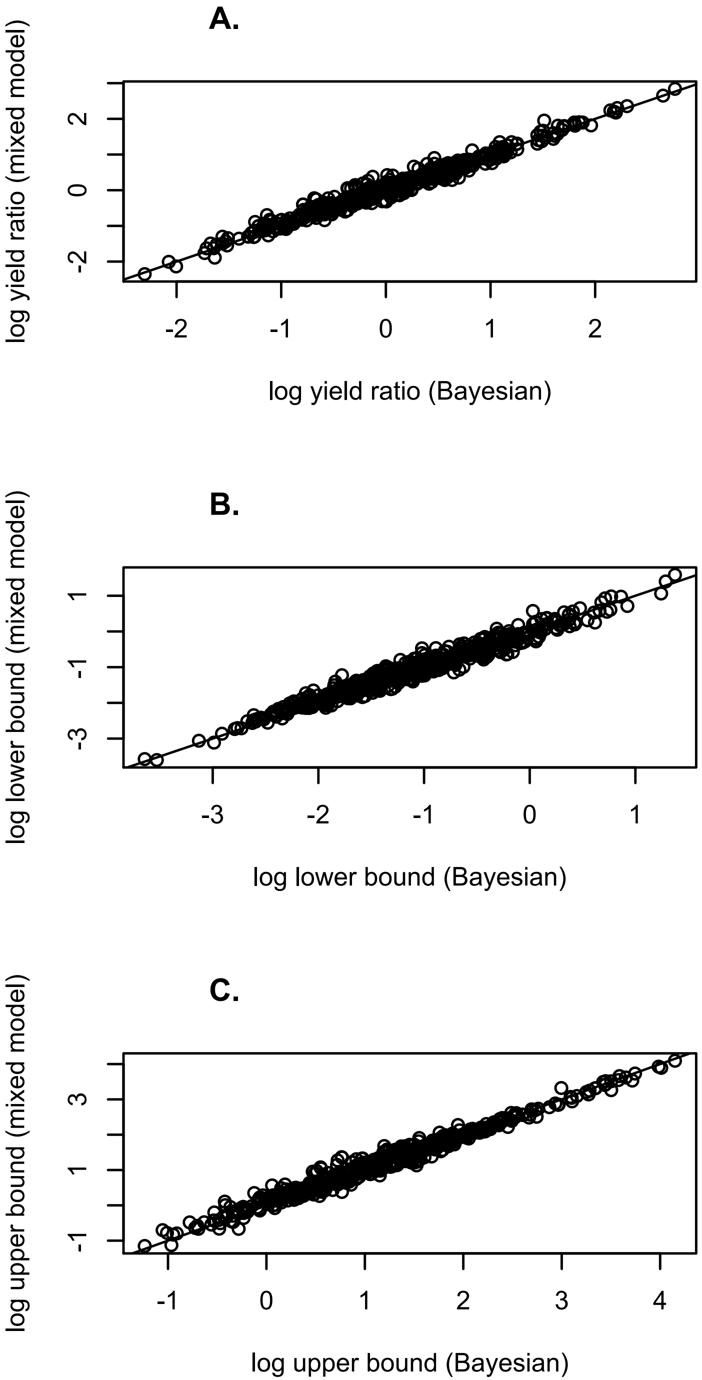
Results

	Bayesian fixed-effect model (model 1)	Bayesian random-effect model (model 2)	Bayesian random-effect model with species-specific variances (model 3)	Two-way mixed effect model (model 4)
DIC/AIC/BIC	DIC=912	DIC=347	DIC=287	AIC=750.4 BIC=919.9
Dbar	809	145	175	
pD	103	202	112	
PPP	0.51	0.51	0.51	
tau		0.69 (0.60, 0.80)	0.33 (0.18, 0.46)	0.58
sigma	0.46 (0.43, 0.48)	0.27 (0.25, 0.29)		0.27
sigma1			0.31 (0.26, 0.38)	
sigma2			0.24 (0.22, 0.27)	
sigma3			0.08 (0.03, 0.25)	
sigma4			0.31 (0.24, 0.41)	
sigma5			3.09 (0.31, 9.29)	
sigma10			0.15 (0.02, 0.52)	
sigma11			0.72 (0.22, 1.26)	
sigma14			0.22 (0.16, 0.31)	

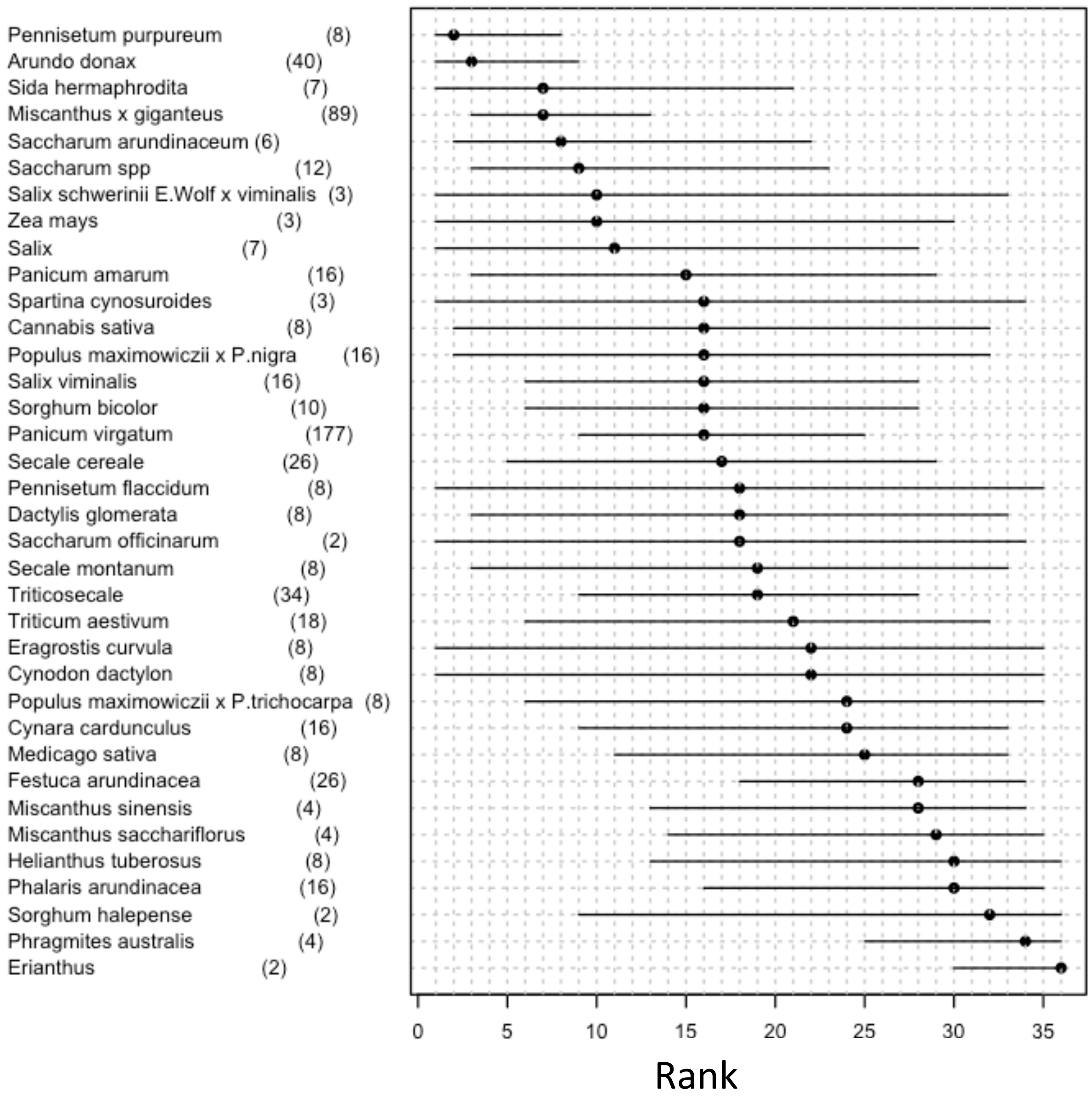
Yield ratios (compared to Miscantus x giganteus)

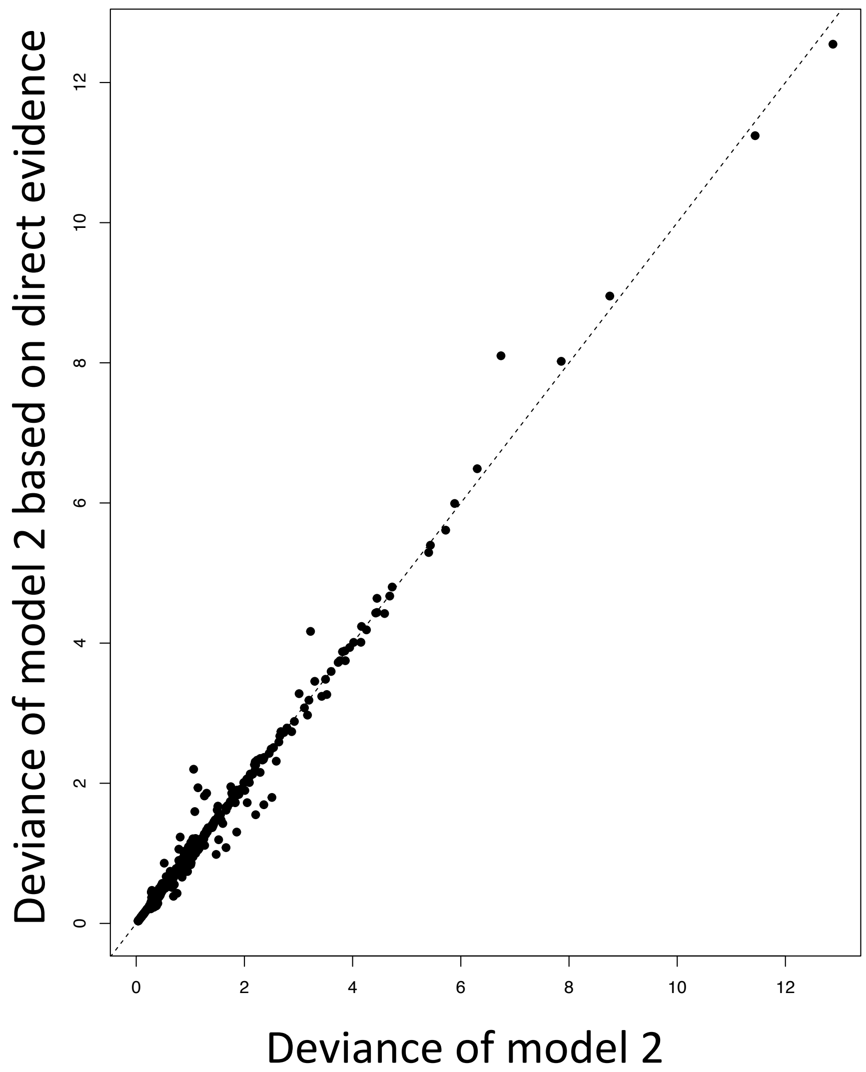


Model 2 *vs.* Model 4

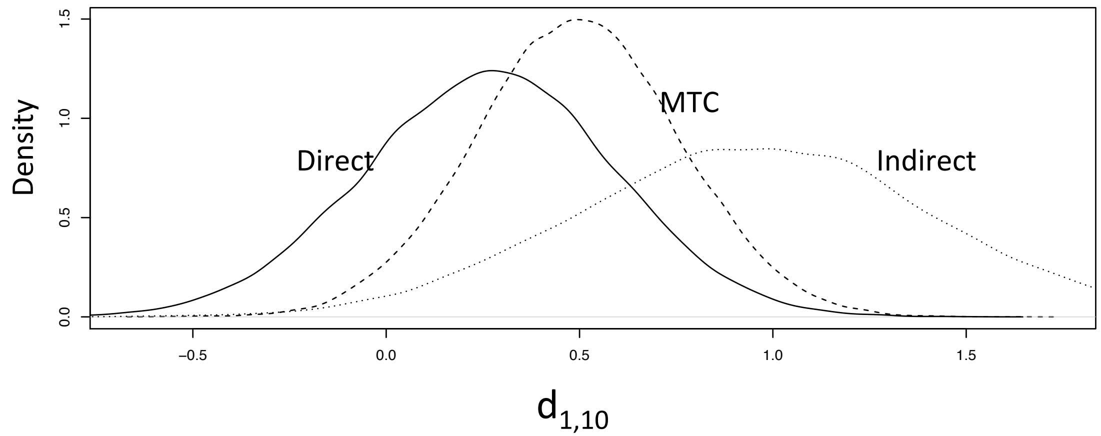


Species ranking (Model 2)

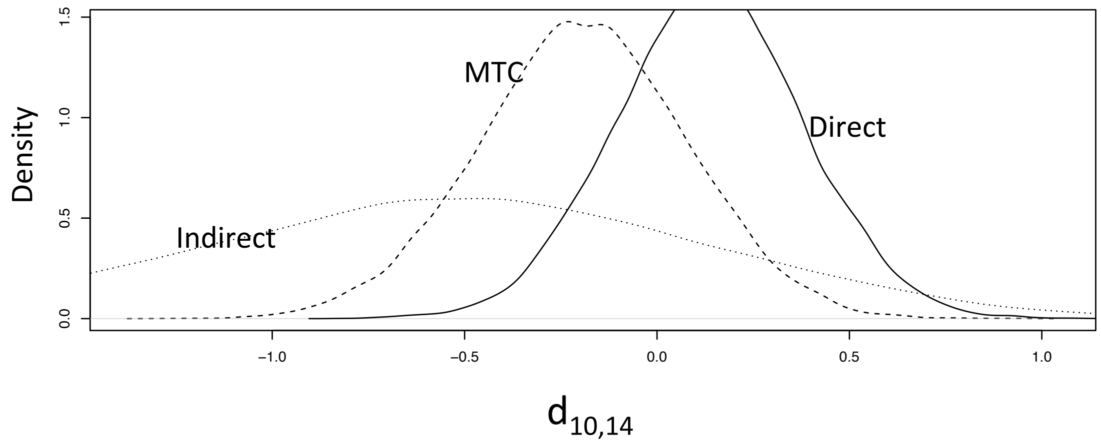




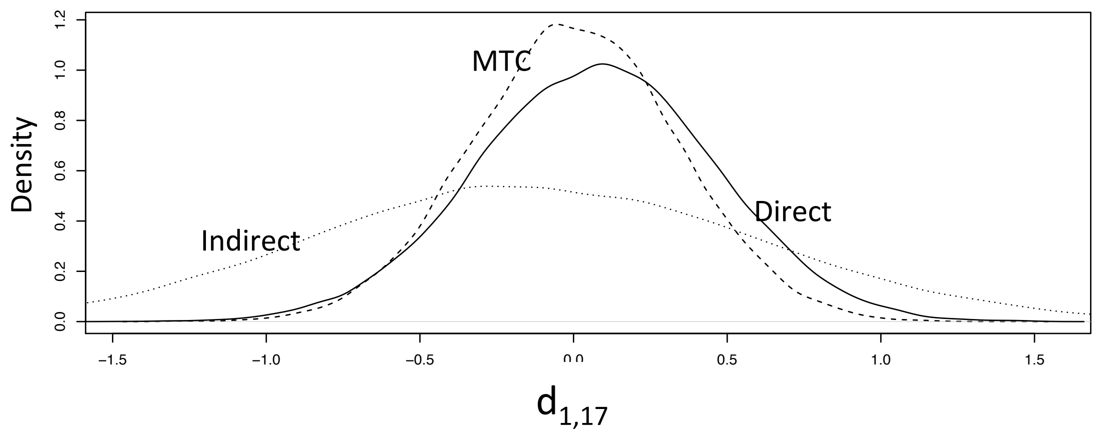
Deviance of model 2 based on direct evidence
0
2
4
6
8
10
12
Deviance of model 2



Density
MTC
Direct
Indirect
0.0
0.5
1.0
1.5
−0.5
0.0
0.5
1.0
1.5
$d_{1,10}$



Density
MTC
Direct
Indirect
0.0
0.5
1.0
1.5
−1.0
−0.5
0.0
0.5
1.0
$d_{10,14}$



Density
MTC
Direct
Indirect
0.0
0.2
0.4
0.6
0.8
1.0
1.2
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
$d_{1,17}$

Conclusion

- Mixed treatment comparison (MTC) can be used for ranking crop species from yield data collected for several species texted in field experiments.
- We introduce several Bayesian MTC models based on baseline treatment contrasts.
- The practical advantages of these models to produce yield ratio estimates in the context of manifold comparisons by study and with a sparse network adjacency matrix.
- Results reveal that the Bayesian and classical models lead to close yield ratio estimates.
- The Bayesian models allow an in-depth analysis of the uncertainty in the species ranking.

References

- Albert I., Makowski D. Ranking crop species using mixed treatment comparisons. In prep.
- Dias S., N. J. Welton, D. M. Caldwell and A. E. Adesa. 2010. Checking consistency in mixed treatment comparison meta-analysis. Statistics in Medicine 29, 932-944.
- Laurent A., Loyce C., Pelzer E., Makowski D. 2015a. Ranking yields of energy crops: a meta-analysis using direct and indirect comparisons. Renewable & Sustainable Energy Reviews 46, 41-50.
- Laurent A., Loyce C., Makowski D., Pelzer E. 2015b. Using site-specific data to estimate energy crop yield. Environmental Modelling & Software 74 104-113.
- Lu, G. and Ades, A. E. (2006). Assessing evidence inconsistency in mixed treatment comparisons. Journal of the American Statistical Association 101, 447–459.
- Madden L.V., Piepho H.-P., Paul P.A. 2016. Statistical models and methods for network meta-analysis. Phytopathology, in press.
- Piepho H.P., Williams E.R., Madden L.V. 2012. The use of two-way linear mixed models in multitreatment meta-analysis. Biometrics 68, 1269-1277.