Ranking crop species from direct and indirect evidences

D. Makowski\textsuperscript{1}, I. Albert\textsuperscript{2}

\textsuperscript{1} UMR Agronomie INRA AgroParisTech Paris-Saclay
\textsuperscript{2} UMR MIA INRA AgroParisTech Paris-Saclay
Which species is the most productive?
One experiment comparing two species

Experiment 1

Species 1: Switchgrass

Species 2: Miscanthus x giganteus
One experiment comparing two species

Experiment 1

Species 1: Switchgrass

Species 2: Miscanthus x giganteus

Yield of species 1

Yield of species 2
Two experiments comparing two species

Experiment 1
Species 1  Species 2

Experiment 2
Species 1  Species 2
Dataset including $N$ experiments comparing $P$ species

<table>
<thead>
<tr>
<th>Experiment 1</th>
<th>Experiment 2</th>
<th>Experiment 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species 1</td>
<td>Species 2</td>
<td>Species 1</td>
</tr>
<tr>
<td>Species 2</td>
<td>Species 1</td>
<td>Species 3</td>
</tr>
<tr>
<td>Experiment 4</td>
<td>Experiment 7</td>
<td>Experiment N</td>
</tr>
<tr>
<td>Species 3</td>
<td>Species 2</td>
<td>Species 4</td>
</tr>
<tr>
<td>Species 2</td>
<td>Species 1</td>
<td>Species 2</td>
</tr>
<tr>
<td>Experiment 8</td>
<td>Experiment 9</td>
<td></td>
</tr>
<tr>
<td>Species P</td>
<td>Species 2</td>
<td>Species 4</td>
</tr>
<tr>
<td>Species 5</td>
<td>Species 4</td>
<td></td>
</tr>
</tbody>
</table>
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:

\[
\log(Y_{ij}) = \mu_{\text{ref}} + \alpha_i + b_j + \varepsilon_{ij}
\]

- \(Y_{ij}\) = Yield of the crop \(i\) for the « site-year » \(j\)
- \(\mu_{\text{ref}}\) = Average yield for the reference crop
- \(\alpha_i\) = Fixed effect (\(i\) = crop index)
- \(b_j\) = Random effect (\(j\) = site-year index)
- \(\varepsilon_{ij}\) = 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 sacchariflorus
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=\( \text{Yield species} \times / \text{Yield triticosecale} \))

Laurent et al., 2015
Indirect comparison

Ex: switchgrass vs. alfalfa

Experiment 1

Switchgrass

Yield switchgrass

Yield Miscanthus

Experiment 2

Miscanthus × giganteus

Yield alfalfa

Yield Miscanthus

Yield Alfalfa

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: TriJcosecale
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*
- Switchgrass
- *Miscanthus x giganteus*

*Site-year 2*
- Alfalfa
- *Miscanthus x giganteus*

Yield switchgrass
Yield *Miscanthus*

Yield alfalfa
Yield *Miscanthus*
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)

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

\[
\mu_{ij} = \mu_{\text{Ref}_i}^{ij} + d_{\text{Ref}_j} \times 1_{j \neq \text{Ref}_i}
\]

- \( Y_{ijk} \) is the \( k^{th} \) yield measured in site-year \( i \) for crop species \( j \),
- \( \mu_{\text{Ref}_i}^{ij} \) is the mean log-yield value of the reference species \( \text{Ref}_i \) in site-year \( i \) (reference species may differ across site-years),
- \( d_{\text{Ref}_j} \) is the mean effect of species \( j \) over site-years relative to the species \( \text{Ref}_i \) (baseline contrasts)
- \( \sigma \) is the log-yield standard deviation.
Fixed-effect model (model 1)

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

\[
\mu_{ij} = \mu_{i}^{\text{Ref}_i} + d_{\text{Ref}_i,j} \times 1_{j \neq \text{Ref}_i}
\]

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

\[
d_{\text{Ref}_i,j} = d_{\text{REF}_j} - d_{\text{REFRef}_i}
\]

- \(d_{\text{REF}_j}\) is the mean effect of species \(j\) over site-years relative to an overall species baseline noted \(\text{REF}\)
- \(d_{\text{REFRef}_i}\) is the mean effect of species \(\text{Ref}_i\) over site-years relative to the overall species baseline \(\text{REF}\)
Fixed-effect model (model 1)

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

\[
\mu_{ij} = \mu_{i}^{Ref} + d_{Ref_{i}j} \times 1_{j \neq Ref_{i}}
\]

Assumption of consistency:

\[
d_{Ref_{i}j} = d_{REFj} - d_{REFRef_{i}}
\]

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)

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

\[
\mu_{ij} = \mu_i^{Ref} + d_{Ref,j} \times 1_{j \neq Ref_i}
\]

Assumption of consistency:

\[
d_{Ref,i,j} = d_{REFj} - d_{REFRef_i}
\]

\[
\mu_i^{Ref} \sim N(0, 10^4) \text{ for } i = 1, \ldots, N
\]

\[
d_{REFj} \sim N(0, 10^4) \text{ for } j = 2, \ldots, S
\]

\[
\sigma \sim Unif(0, 2)
\]
Random-effect model (model 2)

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

\[
\mu_{ij} = \mu_{i}^{\text{Ref}_i} + \delta_{i\text{Ref}_i, j} \times 1_{j \neq \text{Ref}_i}
\]

\[
\delta_{i\text{Ref}_i, j} \sim N\left(d_{\text{Ref}_i j}, \tau^2\right)
\]

\[
d_{\text{Ref}_i, j} = d_{\text{REF}_j} - d_{\text{REFRef}_i}
\]

\[
\mu_{i}^{\text{Ref}_i} \sim N\left(0, 10^4\right) \text{ for } i = 1, \ldots, N
\]

\[
d_{\text{REF}_j} \sim N\left(0, 10^4\right) \text{ for } j = 2, \ldots, S
\]

\[
\sigma \sim \text{Unif}(0, 2)
\]

\[
\tau \sim \text{Unif}(0, 2)
\]
Random-effect model (model 2)

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

\[
\mu_{ij} = \mu_{i}^{\text{Ref}_i} + \delta_{i\text{Ref}_i,j} \times 1_{j \neq \text{Ref}_i}
\]

\[
\delta_{i\text{Ref}_i,j} \sim N\left(d_{\text{Ref}_i,j}, \tau^2\right)
\]

\[
d_{\text{Ref}_i,j} = d_{\text{REF}_j} - d_{\text{REFRef}_i}
\]

\[
\text{var}\left(\delta_{ij}^{ij'}\right) = \text{var}\left(\delta_{i\text{Ref}_i,j'} - \delta_{i\text{Ref}_i,j}\right) = \text{var}\left(\delta_{i\text{Ref}_i,j'}\right) + \text{var}\left(\delta_{i\text{Ref}_i,j}\right) - 2 \text{cov}\left(\delta_{i\text{Ref}_i,j'}, \delta_{i\text{Ref}_i,j}\right)
\]

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

from Dias et al. (2010)
Random-effect model (model 2)

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

\[
\mu_{ij} = \mu_{i}^{Ref_i} + \delta_{iRef_i,j} \times 1_{j \neq Ref_i}
\]

\[
\delta_{iRef_i,j} \sim N\left(d_{Ref_i,j}, \tau^2\right)
\]

\[
d_{Ref_i,j} = d_{REFj} - d_{REFRef_i}
\]

\[
\delta_{iRef_i,j} \begin{pmatrix} \delta_{iRef_i,2} \\ \vdots \\ \delta_{iRef_i,j-1} \end{pmatrix} \sim N\left(d_{REFj} - d_{REFRef_i} + \frac{1}{j-1} \sum_{s=1}^{j-1} \left[ \delta_{iRef_i,s} - (d_{REFs} - d_{REFRef_i}) \right], \frac{j}{2(j-1)} \tau^2 \right)
\]

from Dias et al. (2010)
Random-effect model with species-specific residual standard deviations (model 3)

\[ \sigma_j \sim Unif(0, 2) \quad j = 1, \ldots, s \]
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)

\[ D = \frac{(\log (Y) - \mu)^2}{\sigma^2} \]

\[ p_{val_k} = \Pr[D_{k}^{rep} > D_{k}^{obs}] \]

\[ PPP = \Pr[\sum_k D_{k}^{rep} > \sum_k D_{k}^{obs}] \]
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, \( j = B \) or C

MTC excluding direct evidence on BC

\[
\delta^*_{iBC} \sim N(d^\text{Dir}_{BC}, \sigma^*_d)
\]

\[
\delta_{(i)\text{RC}} \sim N(d^\text{Ind}_{BC}, \sigma^2_d)
\]

\[
\sim N(d^\text{AC}_{BC} - d^\text{AB}_{BC}, \sigma^2_d)
\]
Comparison with a two-way mixed model (model 4) (adapted from Piepho et al., 2012; Madden et al., 2016)

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

\[
\mu_{ij} = \theta + \beta_i + \gamma_j + u_{ij}
\]

\[
\beta_i \sim N\left(0, \sigma_{\beta}^2\right) \quad u_{ij} \sim N\left(0, \sigma_{u}^2\right)
\]

\(\theta\) is the mean yield of the baseline species,
\(\beta_i\) is a site-year random effect,
\(\gamma_j\) 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
<table>
<thead>
<tr>
<th>DIC/AIC/BIC</th>
<th>Bayesian fixed-effect model (model 1)</th>
<th>Bayesian random-effect model (model 2)</th>
<th>Bayesian random-effect model with species-specific variances (model 3)</th>
<th>Two-way mixed effect model (model 4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIC</td>
<td>DIC=912</td>
<td>DIC=347</td>
<td>DIC=287</td>
<td>AIC=750.4 BIC=919.9</td>
</tr>
<tr>
<td>Dbar</td>
<td>809</td>
<td>145</td>
<td>175</td>
<td></td>
</tr>
<tr>
<td>pD</td>
<td>103</td>
<td>202</td>
<td>112</td>
<td></td>
</tr>
<tr>
<td>PPP</td>
<td>0.51</td>
<td>0.51</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td>tau</td>
<td></td>
<td>0.69 (0.60, 0.80)</td>
<td>0.33 (0.18, 0.46)</td>
<td>0.58</td>
</tr>
<tr>
<td>sigma</td>
<td>0.46 (0.43, 0.48)</td>
<td>0.27 (0.25, 0.29)</td>
<td></td>
<td>0.27</td>
</tr>
<tr>
<td>sigma1</td>
<td></td>
<td></td>
<td></td>
<td>0.31 (0.26, 0.38)</td>
</tr>
<tr>
<td>sigma2</td>
<td></td>
<td></td>
<td></td>
<td>0.24 (0.22, 0.27)</td>
</tr>
<tr>
<td>sigma3</td>
<td></td>
<td></td>
<td></td>
<td>0.08 (0.03, 0.25)</td>
</tr>
<tr>
<td>sigma4</td>
<td></td>
<td></td>
<td></td>
<td>0.31 (0.24, 0.41)</td>
</tr>
<tr>
<td>sigma5</td>
<td></td>
<td></td>
<td></td>
<td>3.09 (0.31, 9.29)</td>
</tr>
<tr>
<td>sigma10</td>
<td></td>
<td></td>
<td></td>
<td>0.15 (0.02, 0.52)</td>
</tr>
<tr>
<td>sigma11</td>
<td></td>
<td></td>
<td></td>
<td>0.72 (0.22, 1.26)</td>
</tr>
<tr>
<td>sigma14</td>
<td></td>
<td></td>
<td></td>
<td>0.22 (0.16, 0.31)</td>
</tr>
</tbody>
</table>
Yield ratios (compared to Miscantus x giganteus)

A. Model 2

B. Model 4
Model 2 vs. Model 4

A. log yield ratio (Bayesian) vs. log yield ratio (mixed model)

B. log lower bound (Bayesian) vs. log lower bound (mixed model)

C. log upper bound (Bayesian) vs. log upper bound (mixed model)
Species ranking (Model 2)

- Pennisetum purpureum (8)
- Arundo donax (40)
- Sida hermaphrodita (7)
- Miscanthus x giganteus (89)
- Saccharum arundinaceum (8)
- Saccharum spp (12)
- Salix schwerinii E.Wolf x vinifera (3)
- Zea mays (3)
- Salix (7)
- Panicum arundinaceum (16)
- Spartina cynosuroides (3)
- Cannabis sativa (8)
- Populus maximowiczii x P.nigra (16)
- Salix viminalis (16)
- Sorghum bicolor (10)
- Panicum virgatum (177)
- Secale cereale (26)
- Pennisetum flaccidum (8)
- Dactylis glomerata (8)
- Saccharum officinarum (2)
- Secale montanum (8)
- Triticosecale (34)
- Triticum aestivum (18)
- Eragrostis curvula (8)
- Cynodon dactylon (8)
- Populus maximowiczii x P.trichocarpa (8)
- Cynara cardunculus (15)
- Medicago sativa (8)
- Festuca arundinacea (26)
- Miscanthus sinensis (4)
- Miscanthus sacchariflorus (4)
- Helianthus tuberosus (8)
- Phalaris arundinacea (16)
- Sorghum halepense (2)
- Phragmites australis (4)
- Erianthus (2)
Deviance of model 2 based on direct evidence

Deviance of model 2
The diagrams illustrate the density functions for different scenarios denoted by $d_{1,10}$, $d_{10,14}$, and $d_{1,17}$. Each scenario includes three components: Direct, Indirect, and MTC. The graphs show the distribution of these components across a range of values, with the density on the y-axis and the range on the x-axis.
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.


