MetaForest

Using random forests to explore heterogeneity in meta-analysis
Caspar J. van Lissa, Utrecht University NL

c.j.vanlissa@uu.nl
Applied meta-analysis

- Considered “golden standard” of evidence Crocetti, 2016

- “Superstitions” that it is somehow immune to small-sample problems because each data point is based on an entire study

- Often small N, but many moderators (either measured or ignored)
Dealing with heterogeneity

1. Studies are too different
   * Do not meta-analyze

2. Studies are similar, but not ‘identical’
   * Random-effects meta-analysis

3. There are known differences between studies
   * Code differences as moderating variables
   * Control for moderators using meta-regression (Higgins et al., 2009)
* **Fixed-Effect meta-analysis:**
  * One “true” effect size
  * Observed effect sizes differ due to sampling error
  * Weighted “mean” of effect sizes
  * Big N → more influence

![Impact of Statin Dose On Death and Myocardial Infarction](chart.png)
Types of meta-analysis

* Random-Effects meta-analysis:
  * Distribution of true effect sizes
  * Observed effect sizes differ due to:
    * Sampling error (as before)
    * The variance of this distribution of effect sizes
  * Weights based on precision **and** heterogeneity
    * Study weights become more equal, the more between-studies heterogeneity there is
* True effect size is a function of moderators
* Weighted regression
  * Fixed-effects or random-effects weights
Problem with heterogeneity

* Differences in terms of samples, operationalizations, and methods might all introduce heterogeneity Liu, Liu, & Xie, 2015

* When the number of studies is small, meta-regression lacks power to test more than a few moderators

* We often lack theory to whittle down the list of moderators to a manageable number Thompson & Higgins, 2002

* If we include too many moderators, we might overfit the data
How can we weed out which study characteristics influence effect size?
Dusseldorp and colleagues (2014) used “Classification Trees” to explore which combinations of study characteristics jointly predict effect size.

- The Dependent Variable is Effect Size.
- The Independent Variables are Study Characteristics (moderators).
How do tree-based models work?

- They predict the DV by splitting the data into groups, based on the IV’s.
They predict the DV by splitting the data into groups, based on the IV’s.
How do tree-based models work?

- They predict the DV by splitting the data into groups, based on the IV’s.
How do tree-based models work?

- They predict the DV by splitting the data into groups, based on the IV’s.
Advantages of trees over regression

* Trees easily handle situations where there are many predictors relative to observations

* Trees capture interactions and non-linear effects of moderators

* Both these conditions are likely to be the case when performing meta-analysis in a heterogeneous body of literature
Limitations of single trees

- Single trees are very prone to overfitting
Random Forests

1. Draw many (+/−1000) bootstrap samples
2. Grow a trees on each bootstrap sample
3. To make sure each tree learns something unique, they are only allowed to choose the best moderator from a small random selection of moderators at each split
4. Average the predictions of all these trees
Benefits of random forests

- Random forests are **robust to overfitting**
  - Each tree captures some “true” effects and some idiosyncratic noise
  - Noise averages out across bootstrap samples

- Random forests make **better predictions** than single trees
  - Single trees predict a constant value for each “node”
  - Forests average predictions of many trees, leading to smooth prediction curves
How does MetaForest work?

- Apply random-effects weights to random forests
- Just like in classic meta-analysis, more precise studies are more influential in building the model
What do I report in my paper?

* An “$R^2_{oob}$”: An estimate of how well this model predicts new data

* Variable importance metrics, indicating which moderators most strongly predict effect size

* Partial dependence plots:
  Marginal relationship between moderators and effect size
Several simulation studies examining:
- Predictive performance
- Power
- Ability to identify relevant / irrelevant moderators
- Van Lissa, 2017: https://osf.io/khjgb/
Design factors:

- $k$: Number of studies in meta-analysis (20, 40, 80, and 120)
- $N$: Average within-study sample size (40, 80, and 160)
- $M$: Number of irrelevant/noise moderators (1, 2, and 5)
- $\beta$: Population effect size (.2, .5, and .8)
- $\tau^2$: Residual heterogeneity (0, .04, and .28) Van Erp et al., 2017 (0, 50 and 80th percentile)

Model:

- (a) main effect of one moderator
- (b) two-way interaction
- (c) three-way interaction
- (d) two two-way interactions
- (e) non-linear, cubic relationship

Focusing on one simulation study
To determine practical guidelines, we examined under what conditions MetaForest achieved a positive $R^2$ in new data at least 80% of the time.
Results

- MetaForest had sufficient power in most conditions, even for as little as 20 studies,
  - Except when the effect size was small ($\beta = 0.2$), and residual heterogeneity was high ($\tau^2 = 0.28$)

- Power was most affected by true effect size and residual heterogeneity, followed by the true underlying model
MetaForest is a comprehensive approach to Meta-Analysis.

You could just report:
- Variable importance
- Partial prediction plots
- Residual heterogeneity

Alternatively, add it to your existing Meta-Analysis workflow
- Use it to check for relevant moderators
- Follow up with classic meta-analysis
Can you get it published?

**Methodological journal:**
- Received positive Reviews
- Editor: “the field of psychology is simply not ready for this technique”

**Applied journal:** (Journal of Experimental Social Psychology, 2018)
- Included MetaForest as a check for moderators
- Accepted WITHOUT QUESTIONS about this new technique
- Editor: “I see the final manuscript as having great potential to inform the field.”
- Manuscript, data, and syntax at [https://osf.io/sey6x/](https://osf.io/sey6x/)


**Small sample: 17 studies** (79 effect sizes)

**Dependent variable:** Intervention effect (Cohen’s D)

**Moderators:**
- DV_Aligned: Outcome variable aligned with training content?
- Location: Conducted in childcare center or elsewhere?
- Curriculum: Fixed curriculum?
- Train_Knowledge: Focus on teaching knowledge?
- Pre_Post: Is it a pre-post design?
- Blind: Were researchers blind to condition?
- Journal: Is this study published in a peer-reviewed journal?
WeightedScatter(data, yi="di")
```r
res <- rma.mv(d, vi, random = ~ 1 | study_id, mods = moderators, data=data)

<table>
<thead>
<tr>
<th></th>
<th>estimate</th>
<th>se</th>
<th>zval</th>
<th>pval</th>
<th>ci.lb</th>
<th>ci.ub</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>-0.0002</td>
<td>0.2860</td>
<td>-0.0006</td>
<td>0.9995</td>
<td>-0.5607</td>
<td>0.5604</td>
</tr>
<tr>
<td>sex</td>
<td>-0.0028</td>
<td>0.0058</td>
<td>-0.4842</td>
<td>0.6282</td>
<td>-0.0141</td>
<td>0.0085</td>
</tr>
<tr>
<td>age</td>
<td>0.0049</td>
<td>0.0053</td>
<td>0.9242</td>
<td>0.3554</td>
<td>-0.0055</td>
<td>0.0152</td>
</tr>
<tr>
<td>donorcodeTypical</td>
<td>0.1581</td>
<td>0.2315</td>
<td>0.6831</td>
<td>0.4945</td>
<td>-0.2956</td>
<td>0.6118</td>
</tr>
<tr>
<td>interventioncodeOther</td>
<td>0.4330</td>
<td>0.1973</td>
<td>2.1952</td>
<td>0.0281</td>
<td>0.0464</td>
<td>0.8196</td>
</tr>
<tr>
<td>interventioncodeProsocial Spending</td>
<td>0.2869</td>
<td>0.1655</td>
<td>1.7328</td>
<td>0.0831</td>
<td>-0.0376</td>
<td>0.6113</td>
</tr>
<tr>
<td>controlcodeNothing</td>
<td>-0.1136</td>
<td>0.1896</td>
<td>-0.5989</td>
<td>0.5492</td>
<td>-0.4852</td>
<td>0.2581</td>
</tr>
<tr>
<td>controlcodeSelf Help</td>
<td>-0.0917</td>
<td>0.0778</td>
<td>-1.1799</td>
<td>0.2380</td>
<td>-0.2442</td>
<td>0.0607</td>
</tr>
<tr>
<td>outcomecodeLife Satisfaction</td>
<td>0.0497</td>
<td>0.0968</td>
<td>0.5134</td>
<td>0.6077</td>
<td>-0.1401</td>
<td>0.2395</td>
</tr>
<tr>
<td>outcomecodeOther</td>
<td>-0.0300</td>
<td>0.0753</td>
<td>-0.3981</td>
<td>0.6906</td>
<td>-0.1777</td>
<td>0.1177</td>
</tr>
<tr>
<td>outcomecodePN Affect</td>
<td>0.0063</td>
<td>0.0794</td>
<td>0.0795</td>
<td>0.9367</td>
<td>-0.1493</td>
<td>0.1619</td>
</tr>
</tbody>
</table>
```
PartialDependence(res, rawdata = TRUE, pi = .95)
mf <- ClusterMF(d ~ ., study = "study_id", data)

Call:
ClusterMF(formula = d ~ ., data = data, study = "study_id")

R squared (OOB): -0.0489
Residual heterogeneity (tau2): 0.0549
plot(mf)
PartialDependence(mf, rawdata = TRUE, pi = .95)
PartialDependence(mf, rawdata = TRUE, pi = .95)
PartialDependence(mf, vars = c("interventioncode", "age"), interaction = TRUE)
# Meta-analysis using random forests

## MetaForest results:

<table>
<thead>
<tr>
<th>Type of analysis:</th>
<th>ClusterMF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of studies:</td>
<td>Forest 1: 38, Forest 2: 40</td>
</tr>
<tr>
<td>Number of moderators:</td>
<td>7</td>
</tr>
<tr>
<td>Number of trees in forest:</td>
<td>Two forests of length 1000</td>
</tr>
<tr>
<td>Candidate variables per split:</td>
<td>2</td>
</tr>
<tr>
<td>Minimum terminal node size:</td>
<td>0</td>
</tr>
<tr>
<td>OOB prediction error (MSE):</td>
<td>0.25</td>
</tr>
<tr>
<td>R squared (OOB):</td>
<td>0.20</td>
</tr>
</tbody>
</table>

## Tests for Heterogeneity:

<table>
<thead>
<tr>
<th></th>
<th>tau²</th>
<th>tau² SE</th>
<th>I²</th>
<th>H²</th>
<th>Q-test</th>
<th>df</th>
<th>Q_p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw effect sizes:</td>
<td>0.22</td>
<td>0.05</td>
<td>75.64</td>
<td>4.10</td>
<td>284.46</td>
<td>77</td>
<td>0.00</td>
</tr>
<tr>
<td>Residuals (after MetaForest):</td>
<td>0.13</td>
<td>0.04</td>
<td>64.99</td>
<td>2.85</td>
<td>198.13</td>
<td>77</td>
<td>0.00</td>
</tr>
</tbody>
</table>

## Random intercept meta-analyses:

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>se</th>
<th>cl.lb</th>
<th>cl.ub</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw effect sizes:</td>
<td>0.41</td>
<td>0.05</td>
<td>0.38</td>
<td>0.44</td>
<td>0.00</td>
</tr>
<tr>
<td>Residuals (after MetaForest):</td>
<td>-0.01</td>
<td>0.05</td>
<td>-0.12</td>
<td>0.09</td>
<td>0.80</td>
</tr>
</tbody>
</table>

## Convergence plot:

[Convergence plot image]
**Get MetaForest**

- `install.packages("metaforest")`
- `MetaForest`  
- [www.developmentaldatascience.org/metaforest](http://www.developmentaldatascience.org/metaforest)

**Other cool features:**
- Functions for model tuning using the `caret` package