Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

How many trials are needed in crop variety testing?

The 11th Working Seminar on Statistical Methods in Variety Testing
Słupia Wielka, June 26, 2018

Johannes Forkman

How many trials are needed in crop variety testing?

Objective: To investigate the effect of the number of trials on the ranking of the varieties

Precision depends on the number of trials

$$var(\bar{Y}) = \frac{var(Y)}{N}$$

- Variance decreases
- The least significant difference (LSD) decreases
- Power of tests increases
- Ranking becomes more correct

Optional strategies

Select the sample size such that

- **confidence intervals** in differences become small
- **LSD** becomes small
- **power** in statistical tests become large

However, we will instead focus on **ranking**

Ranking

1. The best variety has rank 1
2. The second best has rank 2
3. The third best has rank 3

...

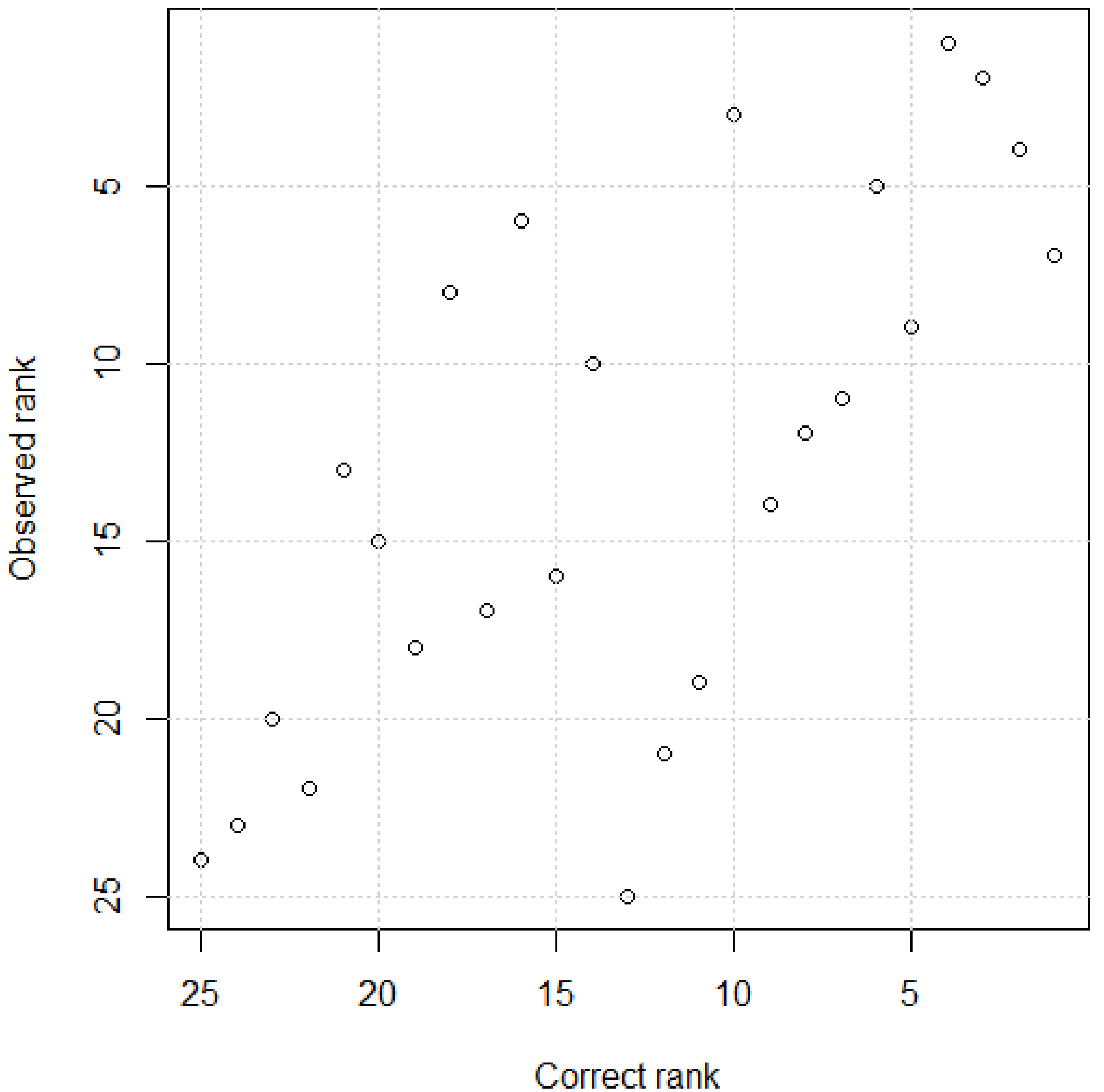
N trials	∞ trials
Ranks depend on N	Ranks are correct

The degree of correspondance

(between ranks at N trials and ranks at ∞ trials)

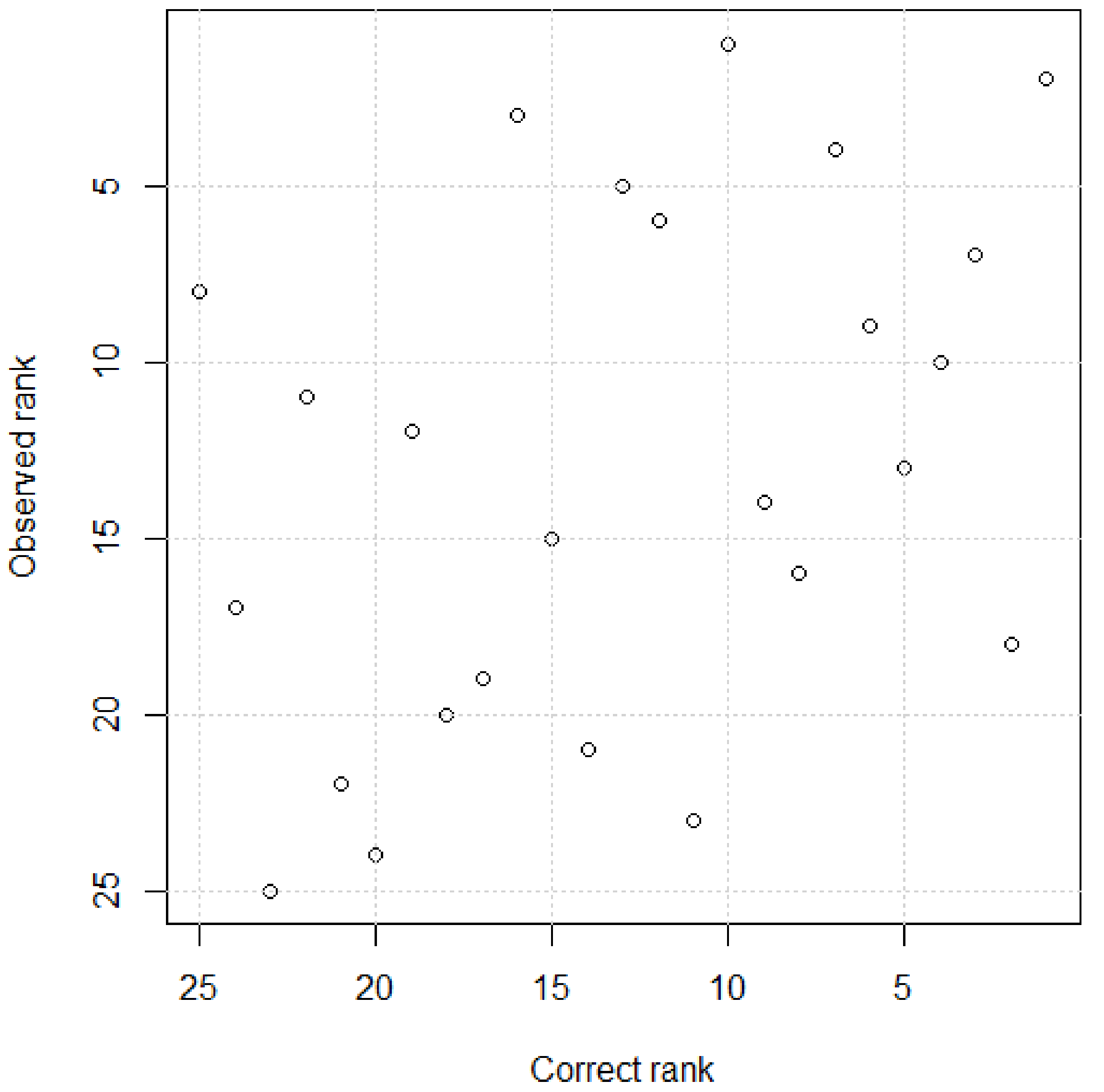
can be measured by the rank correlation coefficient

Rank correlation 0.70



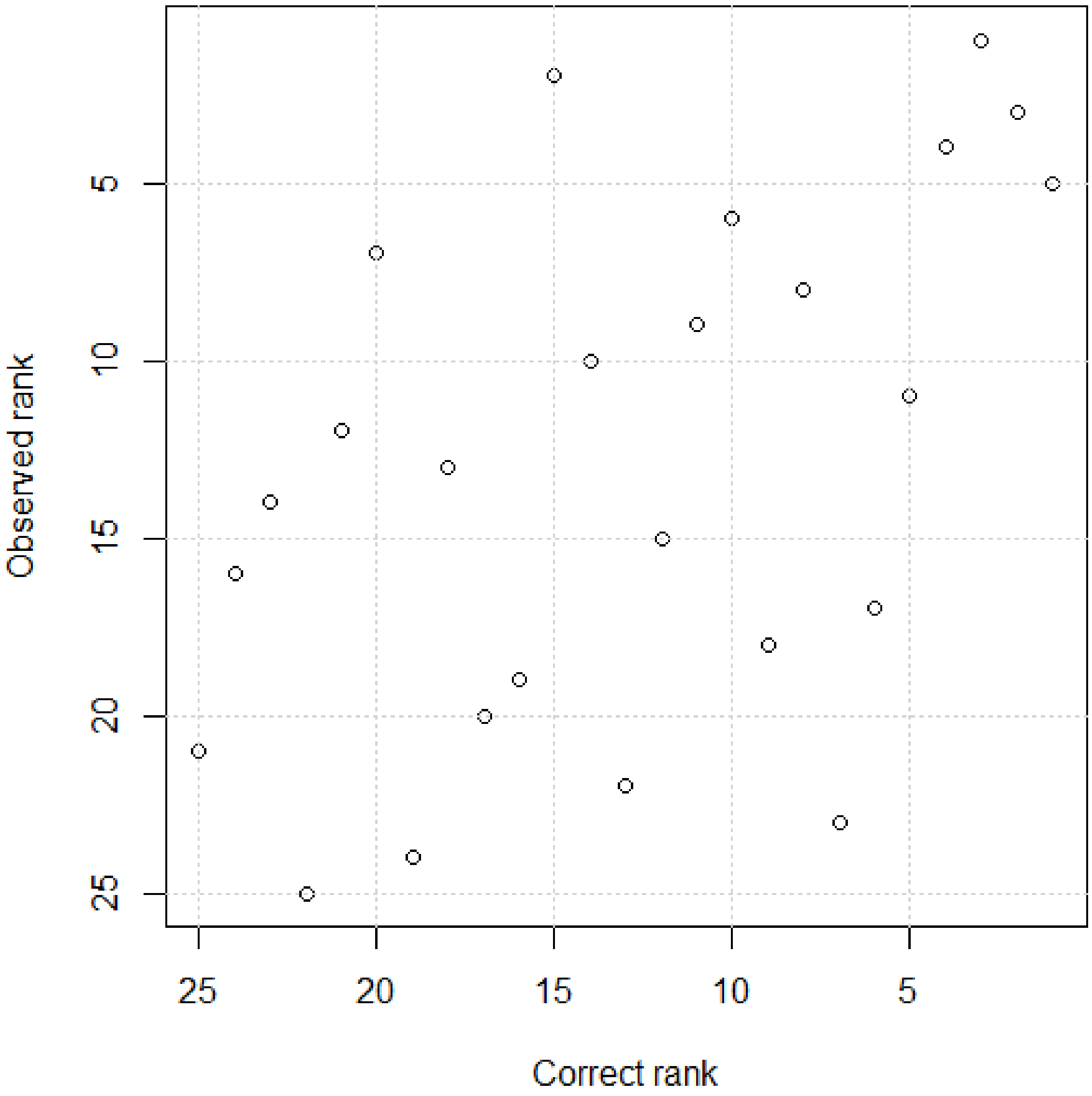
∞ trials

Rank correlation 0.40



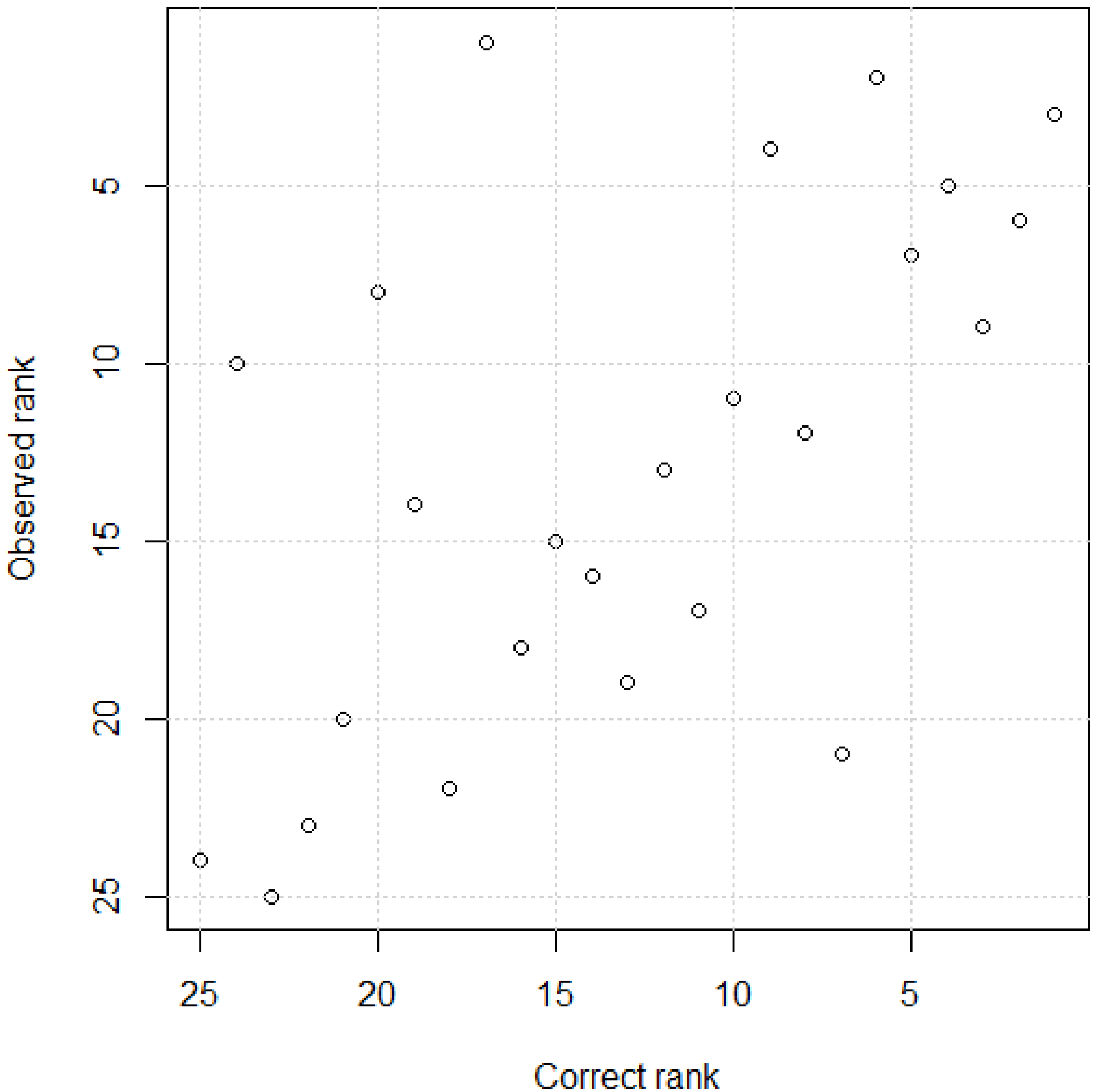
J. Forkman

Rank correlation 0.50

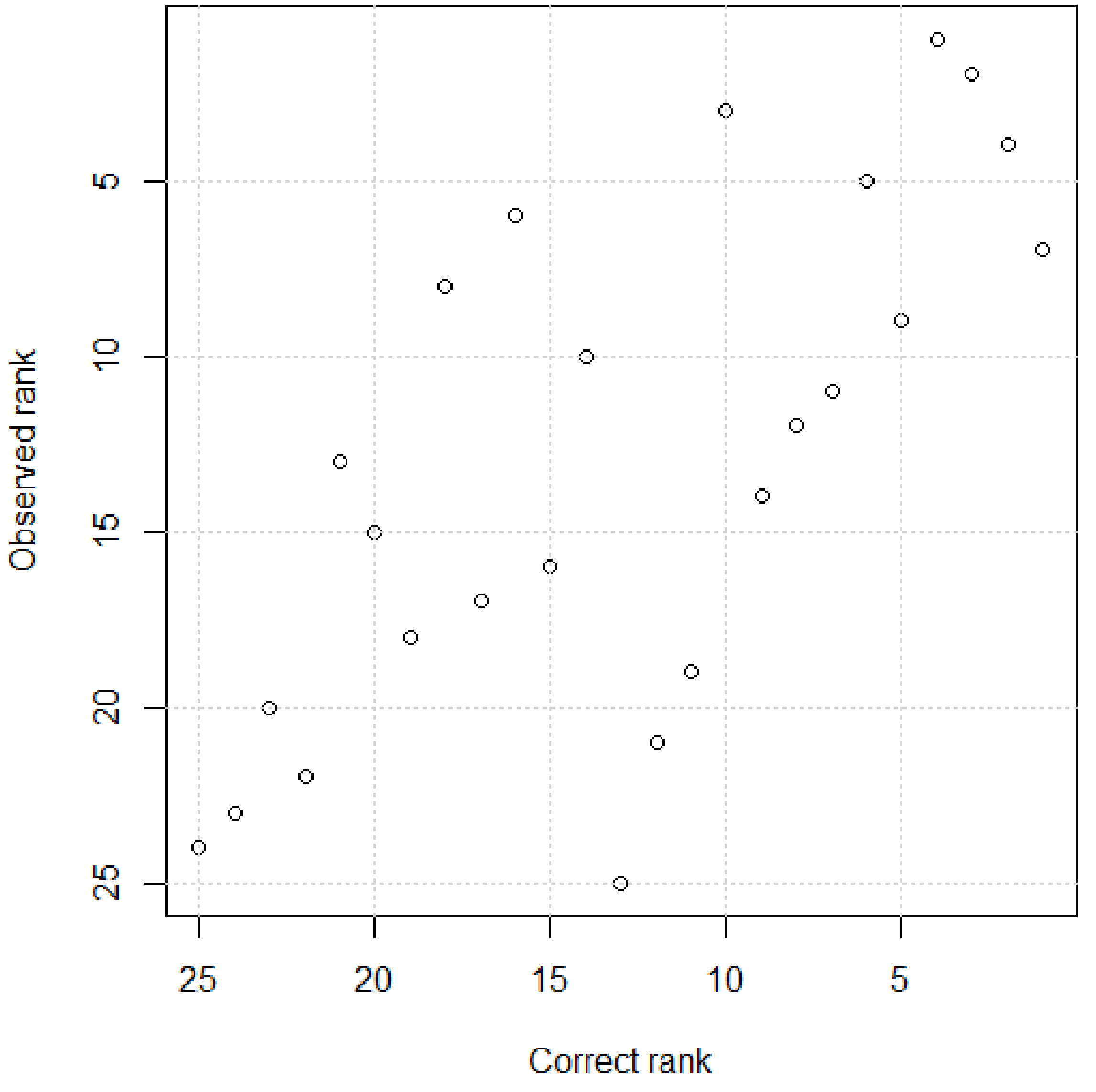


J. Forkman

Rank correlation 0.60

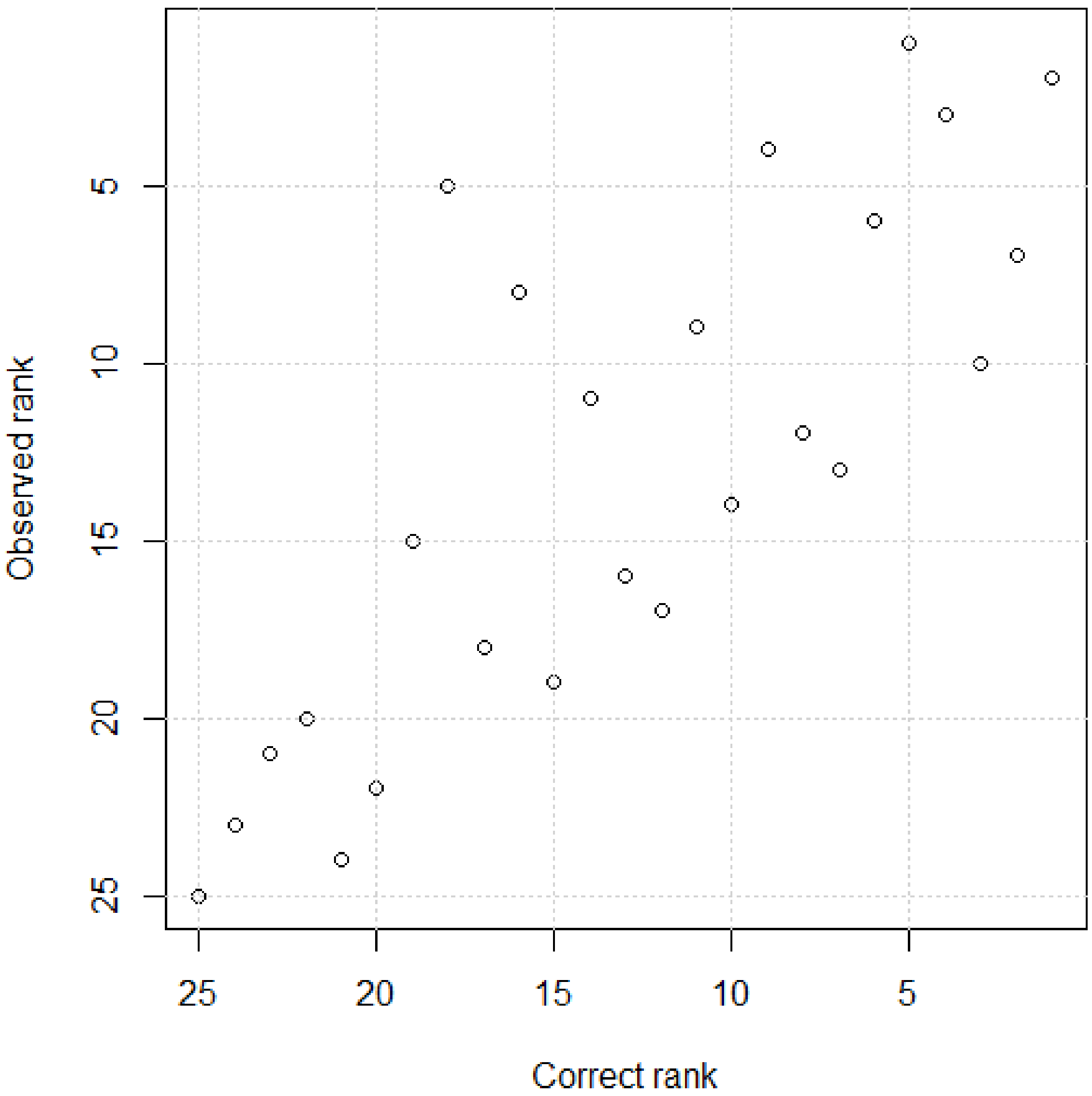


Rank correlation 0.70

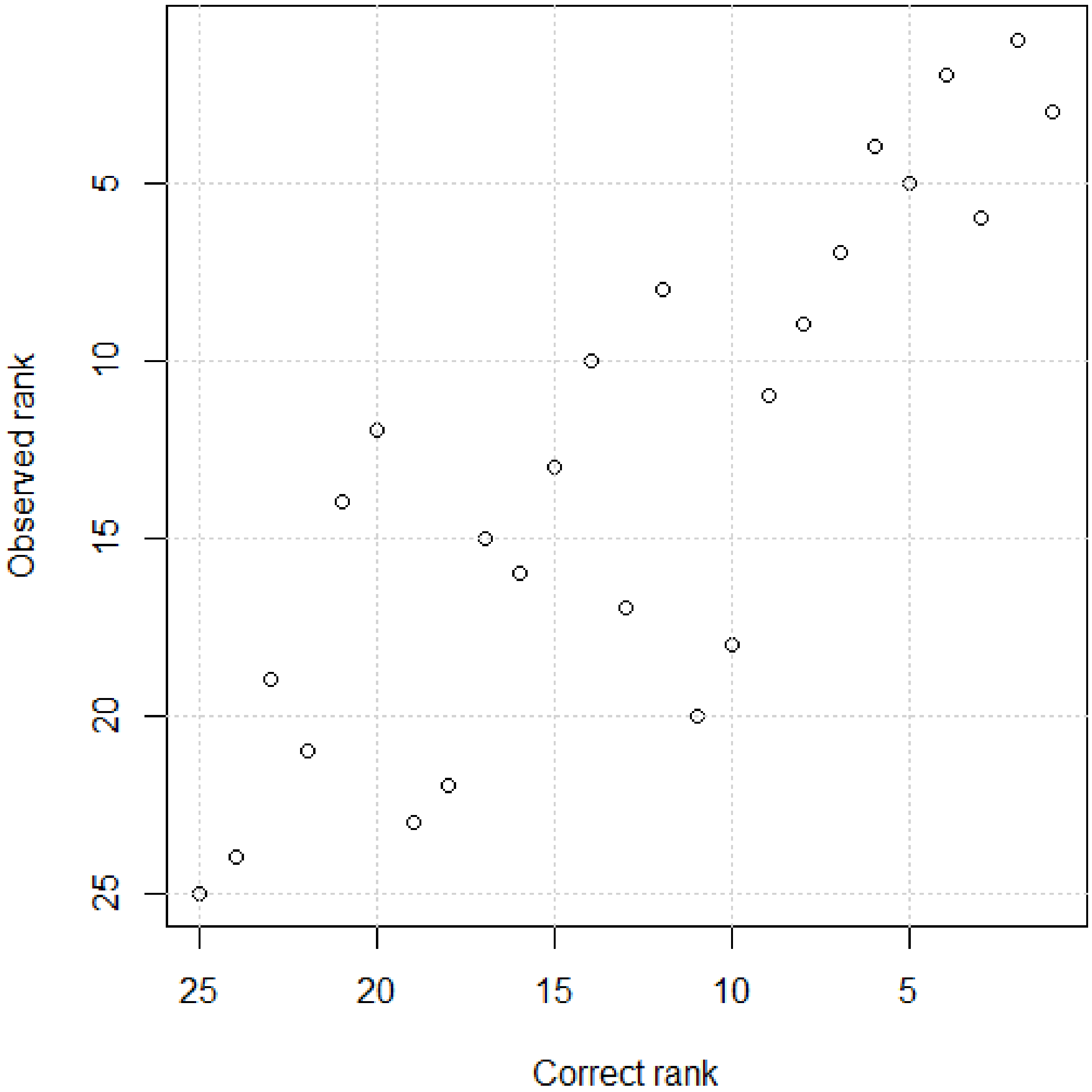


J. Forkman

Rank correlation 0.80

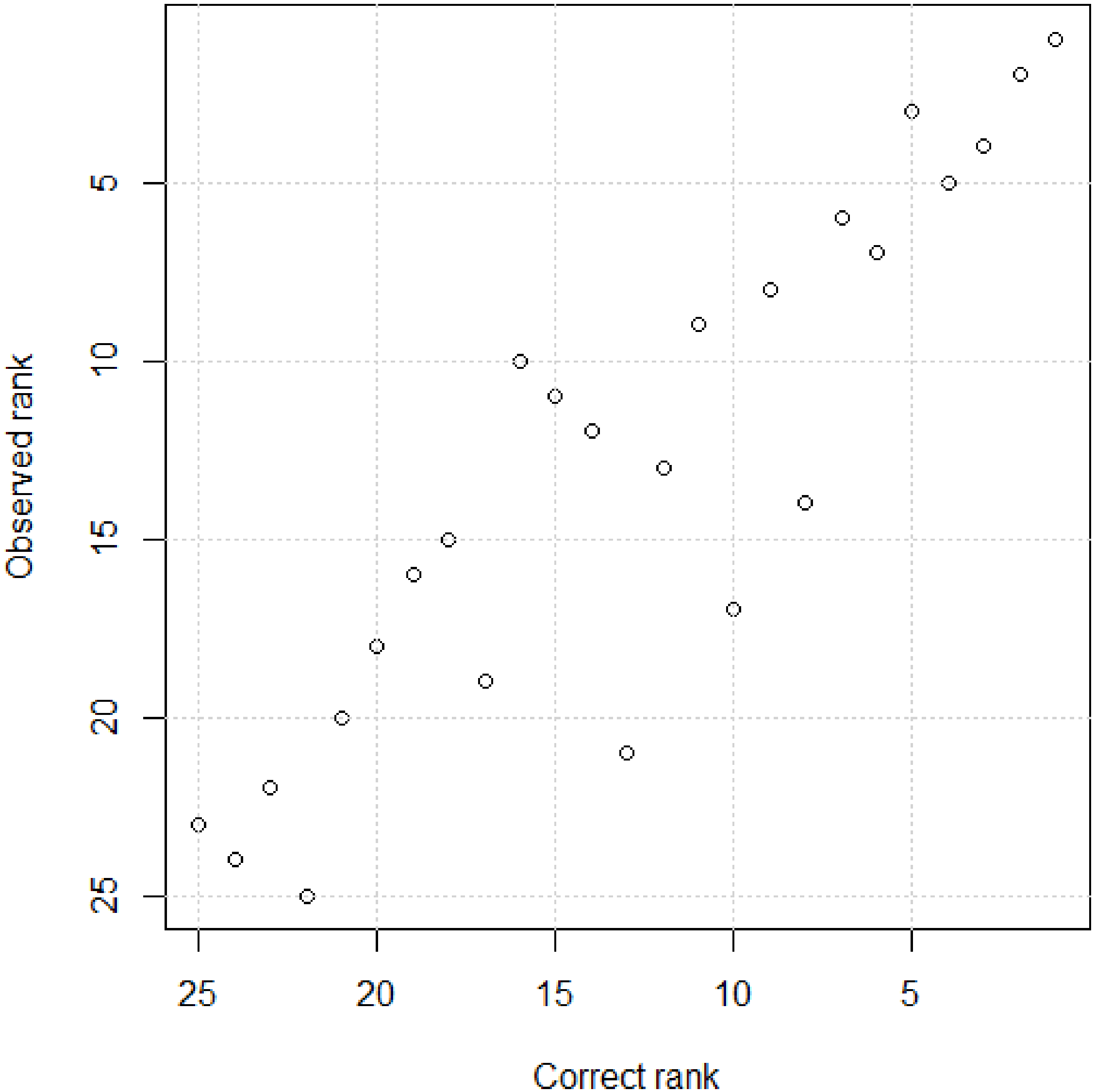


Rank correlation 0.85



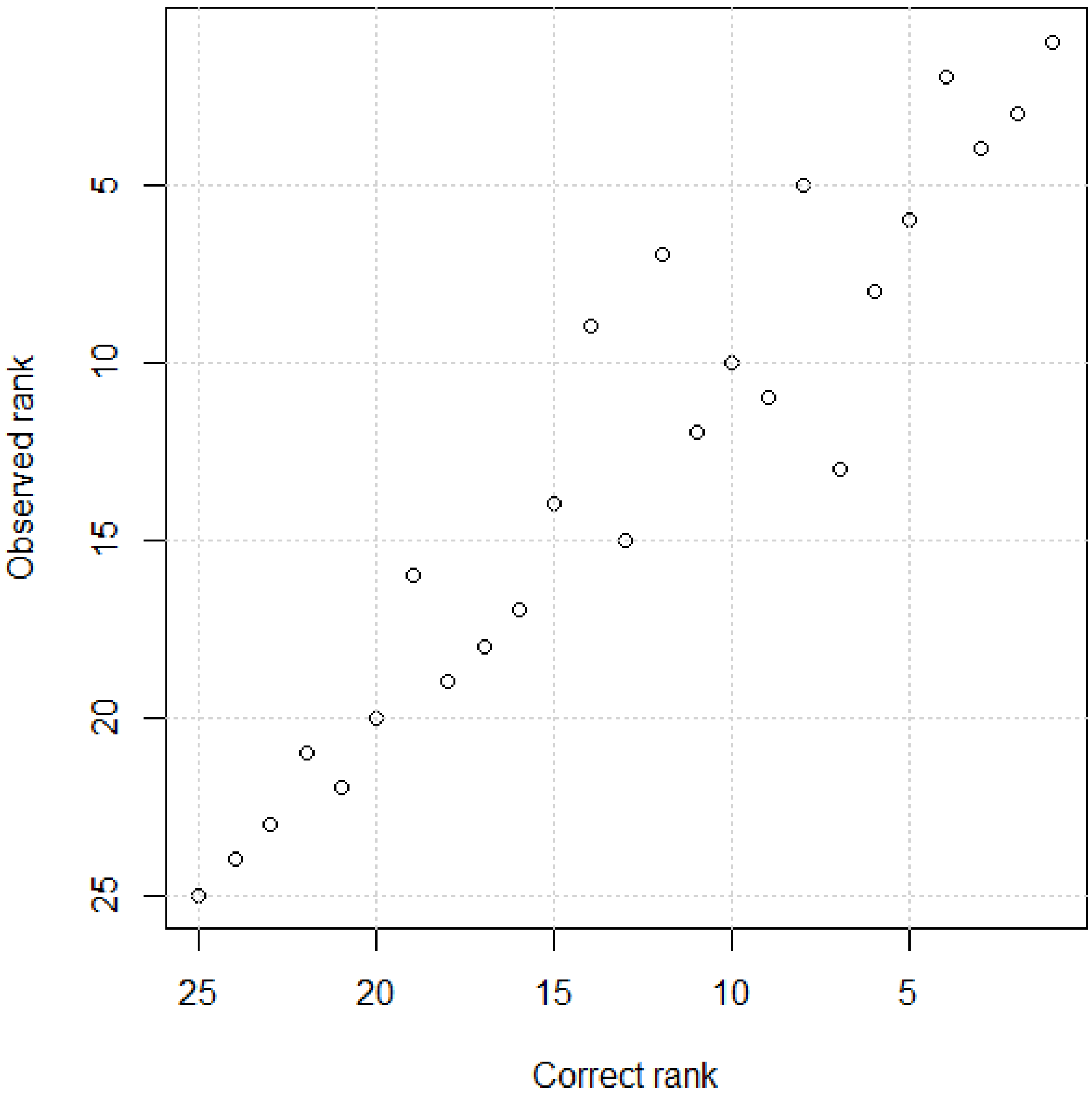
J. Forkman

Rank correlation 0.90



J. Forkman

Rank correlation 0.95



J. Forkman

How many trials are needed?

	Spring barley	Winter wheat
One-year series	✓	✓
Five-year series	✓	✓

Number of Swedish variety trials in **spring barley** and number of varieties per trial

Year	Number of trials	Number of varieties
2015	26	33
2014	26	29
2013	26	12
2012	23	11
2011	25	23

Number of Swedish variety trials in **winter wheat** and number of varieties per trial

Harvest year	Number of trials	Number of varieties
2015	23	29
2014	24	37
2013	14	29
2012	18	32
2011	17	31

Unweighted two-stage procedure

Stage 1: Analyses of each trial seperately

- Fixed effects of varieties
- Random effects of replicates and incomplete blocks

Stage 2: Analysis of means from the first stage

- Using the following model...

Random-effects model for one-year series

Yield — Variety

$$y_{ij} = \mu + t_i + v_j + e_{ij}$$

Trial

$$t_i \sim N(0, \sigma_T^2), \quad v_j \sim N(0, \sigma_V^2), \quad e_{ij} \sim N(0, \sigma_E^2)$$

Parameter estimates (spring barley)

	2015	2014	2013	2012	2011	Mean
μ	7912	7462	7561	7620	6560	7423
σ_V^2	141 632	76 009	18 187	54 707	49 534	68 014
σ_T^2	926 594	2 623 146	1 729 659	1 599 108	1 378 978	1 651 497
σ_E^2	149 648	113 014	104 356	161 065	117 829	129 183

Chosen settings for simulation:

$$\mu = 7400$$

$$\sigma_V^2 = 68\,000$$

$$\sigma_E^2 = 130\,000$$

Simulation procedure

Series with $N = 1, 3, 5, 7, 10, 12, 15, 20, 25, 30, 50,$ and 100 trials were simulated.

For each value of N, 1000 series were generated with M varieties.

$M = 25$ and 30 for spring barley and winter wheat, respectively.

Simulation procedure

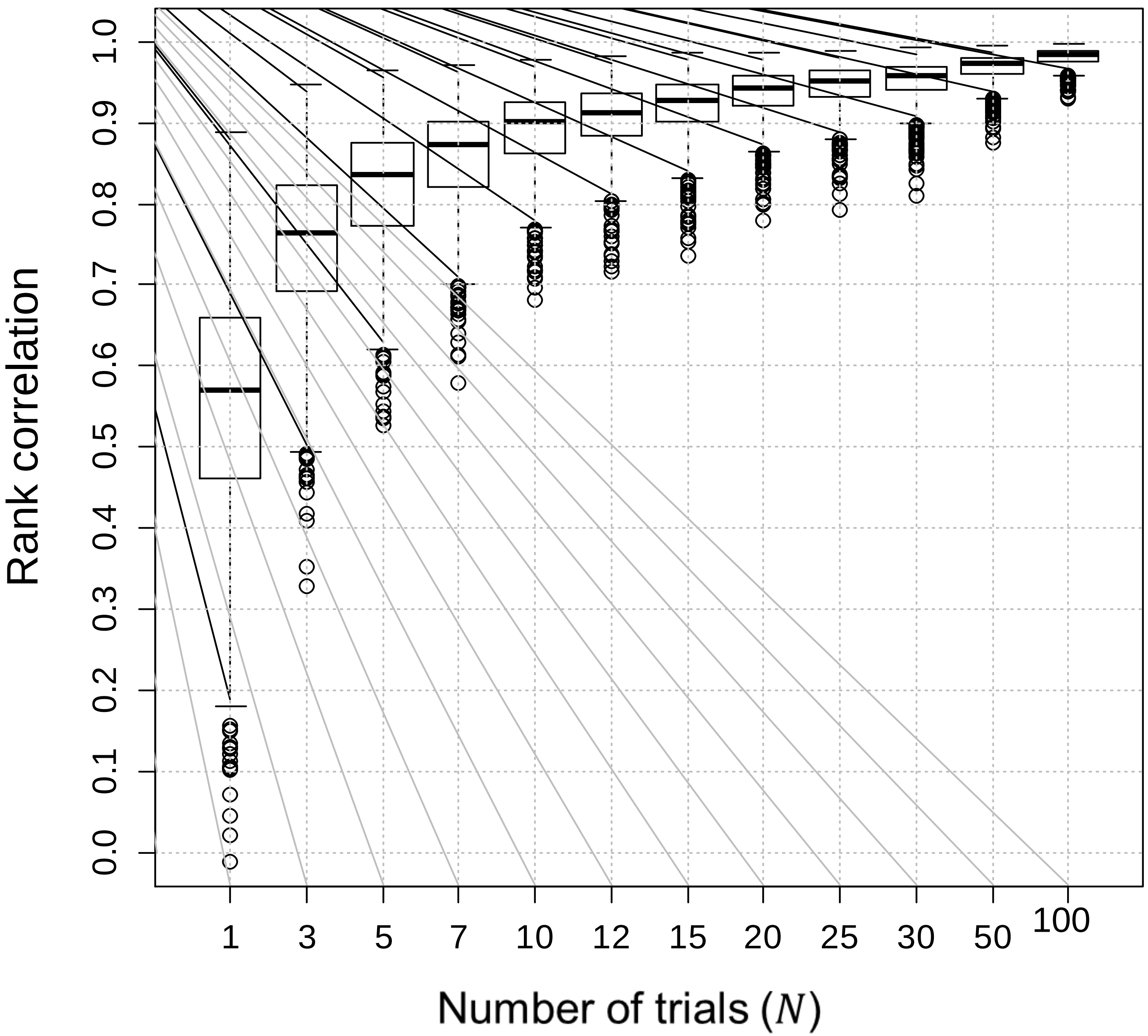
Correct means: $\mu_i \sim N(\mu, \sigma_V^2)$

Estimated means: $\hat{\mu}_i \sim N(\mu_i, \sigma_E^2/N)$

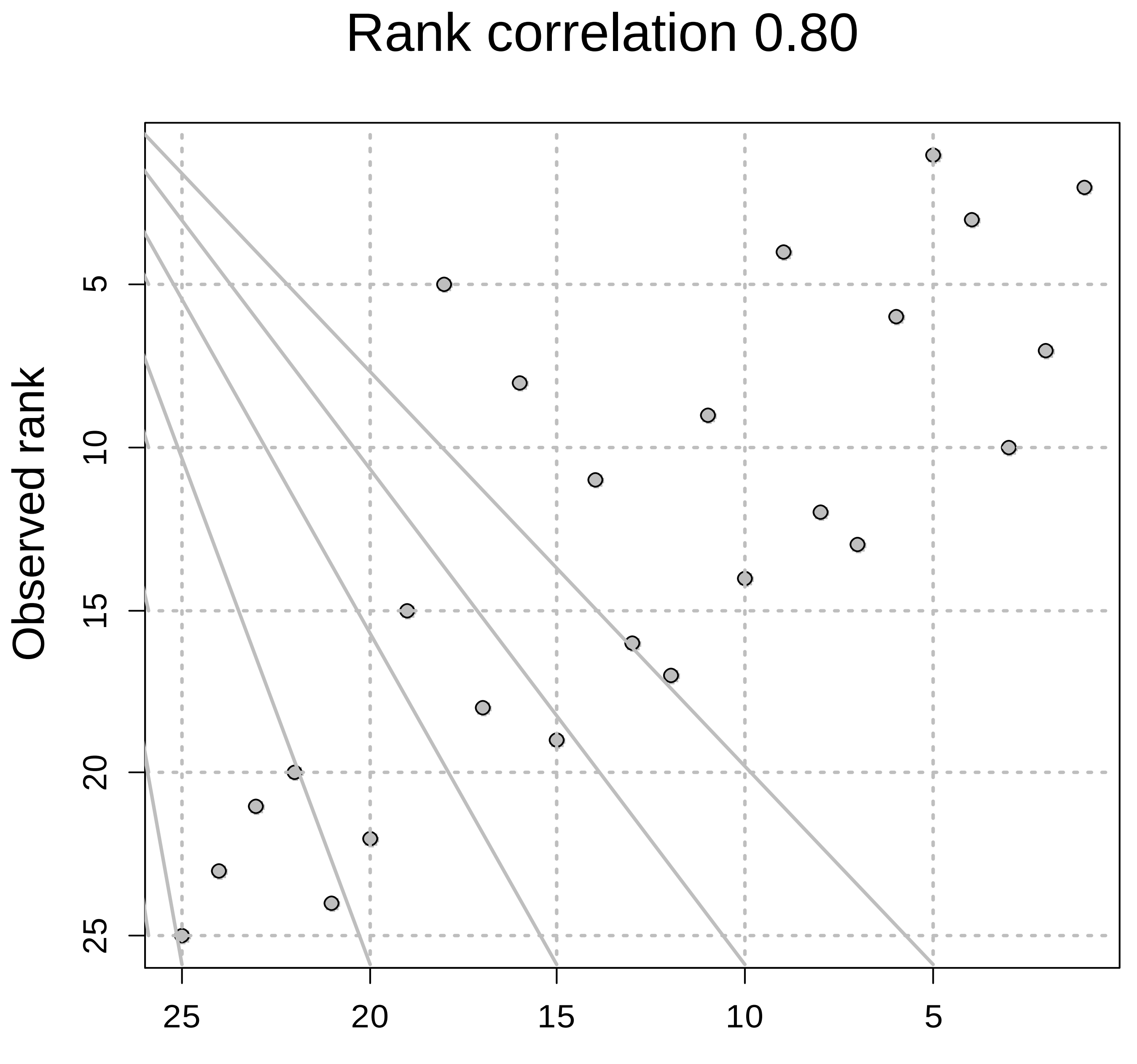
$$i = 1, 2, \ldots, M$$

Spearman's rank correlation between $\mu_1, \mu_2, \ldots, \mu_M$ and $\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_M$ was computed.

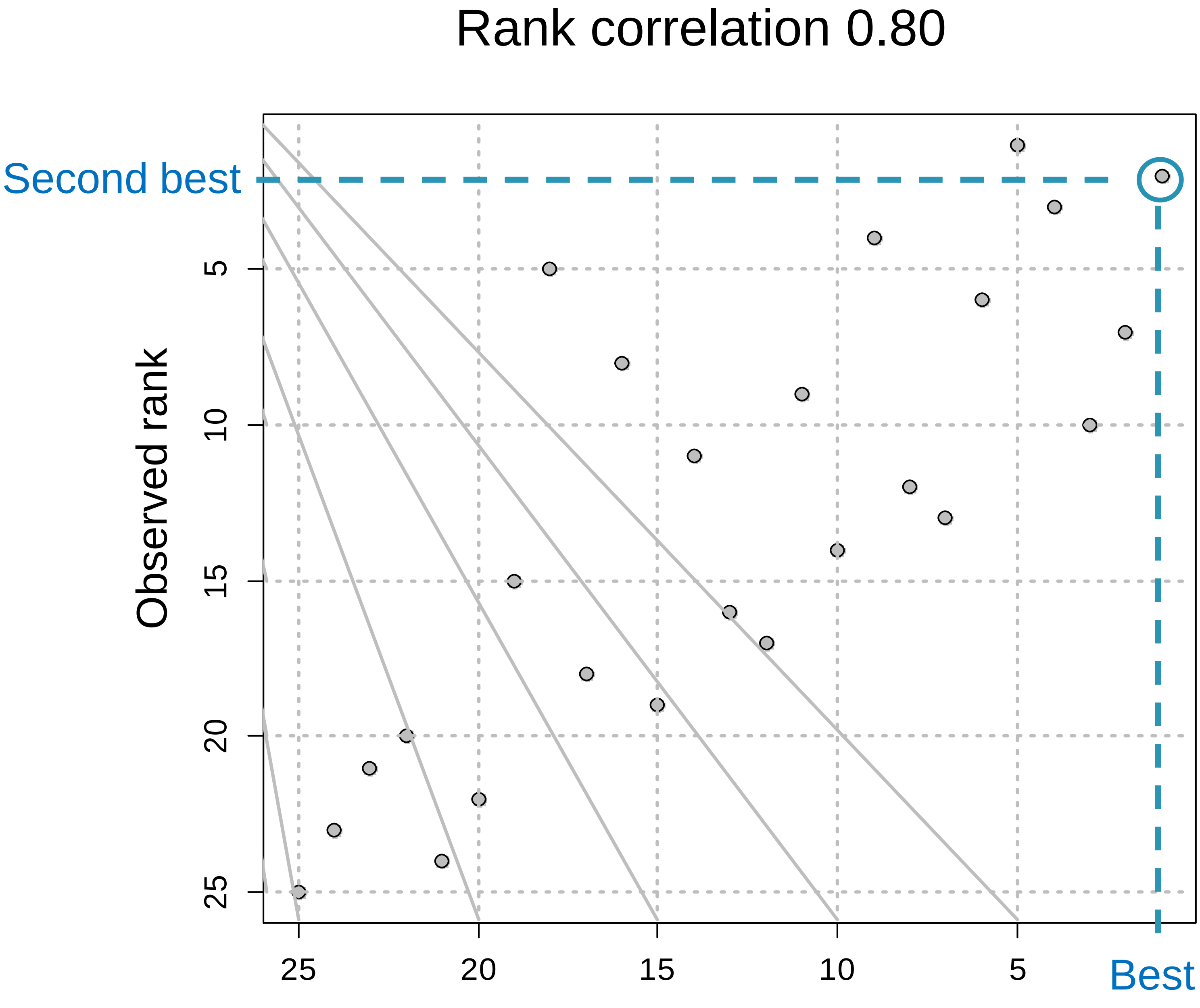
For each value of N, 1000 rank correlations were obtained.



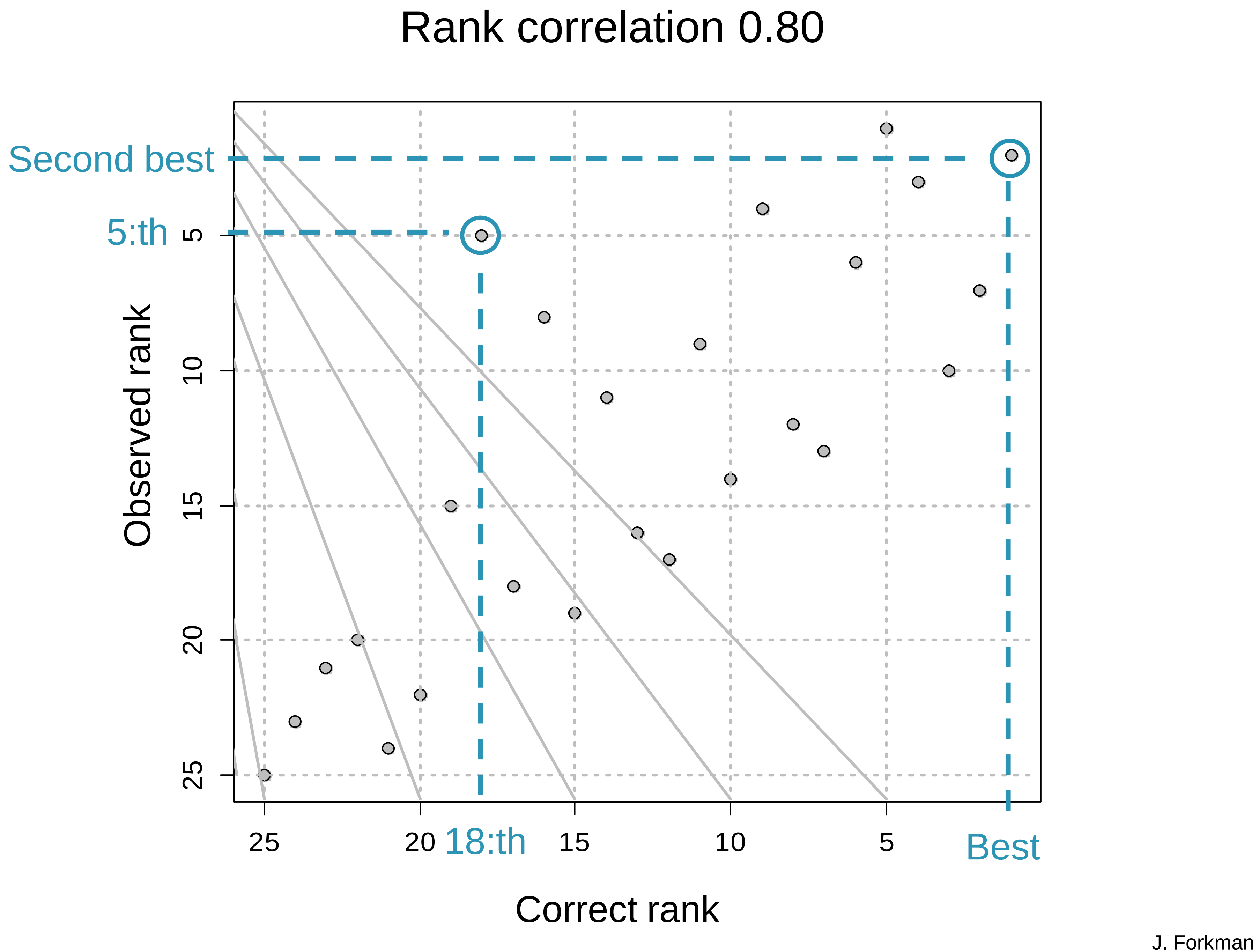
Spring barley, one-year series


Rank correlation 0.80
Observed rank
5
10
15
20
25
25
20
15
10
5
Correct rank

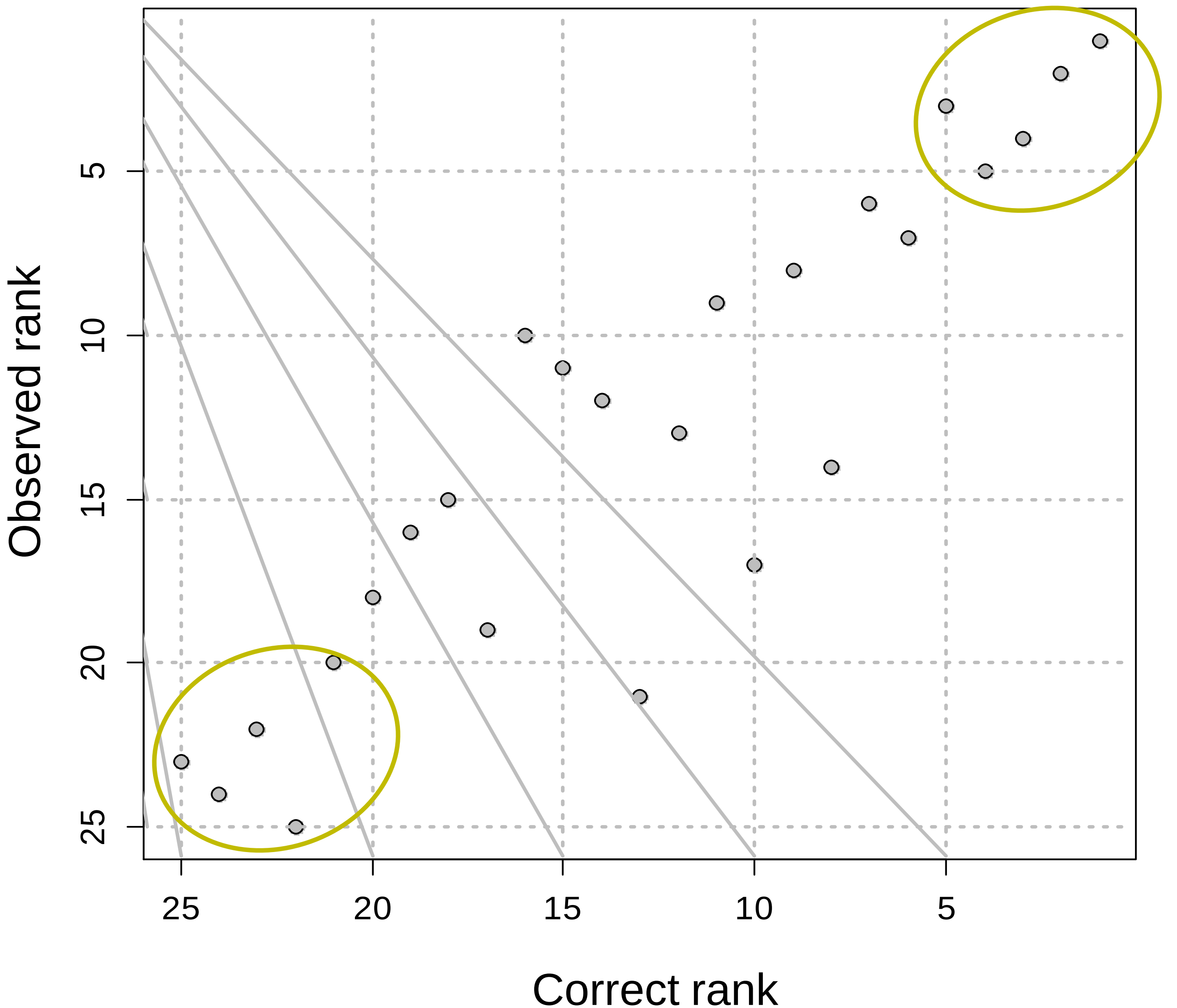
J. Forkman


Rank correlation 0.80
Second best
Best
Observed rank
Correct rank
5
10
15
20
25
25
20
15
10
5

J. Forkman


Rank correlation 0.80
Second best
5:th
Observed rank
5
10
15
20
25
25
20
18:th
15
10
5
Best
Correct rank
J. Forkman

Rank correlation 0.90



J. Forkman

Probability is 95% that rank correlation becomes larger than…

Number of trials	Rank correlation
1	0.28
3	0.56
5	0.68
7	0.74
10	0.80
12	0.82
15	0.85
20	0.87
25	0.89
30	0.91
50	0.93
100	0.96

J. Forkman

Rank correlation (one-year series, spring barley)

		Probability (%)					
Number of trials	95	90	75	50	25	10	5
1	0.28	0.36	0.46	0.57	0.66	0.74	0.77
3	0.56	0.61	0.69	0.76	0.82	0.86	0.88
5	0.68	0.71	0.77	0.84	0.88	0.90	0.92
7	0.74	0.77	0.82	0.87	0.90	0.93	0.94
10	0.80	0.83	0.86	0.90	0.93	0.95	0.95
12	0.82	0.85	0.88	0.91	0.94	0.95	0.96
15	0.85	0.87	0.90	0.93	0.95	0.96	0.97
20	0.87	0.89	0.92	0.94	0.96	0.96	0.97
25	0.89	0.91	0.93	0.95	0.96	0.97	0.98
30	0.91	0.92	0.94	0.96	0.97	0.98	0.98
50	0.93	0.95	0.96	0.97	0.98	0.99	0.99
100	0.96	0.97	0.98	0.98	0.99	0.99	0.99

Conclusions for one-year series

$$\Pr(r_s > 0{,}80) = 0{,}95 \Rightarrow \begin{cases} 10 \text{ trials in spring barley} \\ 8 - 9 \text{ trials in winter wheat} \end{cases}$$

$$\Pr(r_s > 0{,}90) = 0{,}95 \Rightarrow \begin{cases} 25 - 30 \text{ trials in spring barley} \\ 20 \text{ trials in winter wheat} \end{cases}$$

$$\Pr(r_s > 0{,}85) = 0{,}90 \Rightarrow \begin{cases} 12 \text{ trials in spring barley} \\ 10 \text{ trials in winter wheat} \end{cases}$$

Random-effects model for five-year series

Yield — Trial

$$y_{ijk} = \mu + s_i + t_{ij} + v_k + (sv)_{ik} + e_{ijk}$$

Year — Variety

$$s_i \sim N(0, \sigma_S^2), \quad t_{ij} \sim N(0, \sigma_T^2), \quad v_k \sim N(0, \sigma_V^2),$$
$$(sv)_{ik} \sim N(0, \sigma_{SV}^2), \quad e_{ij} \sim N(0, \sigma_E^2)$$

Parameter estimates (spring barley)

	Estimate
μ	7457
σ_V^2	74 826
σ_S^2	140 860
σ_{SV}^2	7658
σ_T^2	1 655 329
σ_E^2	131 380

Chosen settings for simulation:

$$\mu = 7500$$

$$\sigma_V^2 = 75\,000$$

$$\sigma_S^2 = 141\,000$$

$$\sigma_{SV}^2 = 7\,600$$

$$\sigma_T^2 = 1\,165\,000$$

$$\sigma_E^2 = 131\,000$$

Simulation procedure

Series with $N = 1, 3, 5, 7, 10, 12, 15, 20, 25, 30, 50$, and 100 trials were simulated.

For each value of N, 1000 series were generated with M varieties.

$M = 25$ and 30 for spring barley and winter wheat, respectively.

Missing-value pattern

(spring barley)

	25 Varieties																								
Year 1	x	x	x	x	x	x	x																		
Year 2	x	x	x	x	x	x	x	x	x	x	x	x													
Year 3	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x								
Year 4	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Year 5	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

Simulation procedure

Correct means: $\phi_m \sim N(\mu, \sigma_V^2)$

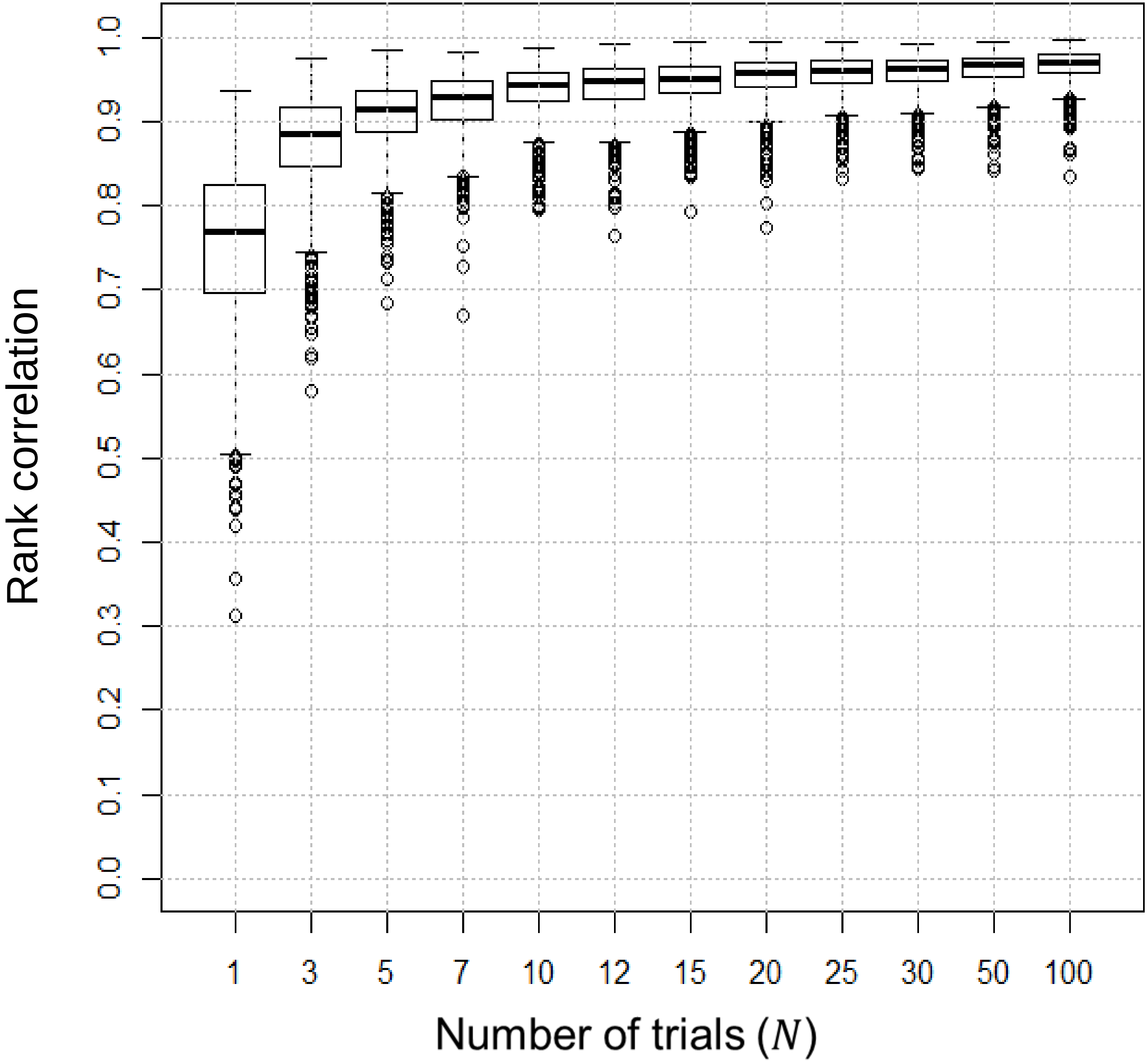
Estimated means: $\hat{\phi}_m$

$$y_{ijk} = \mu + s_i + t_{ij} + \phi_k + (sv)_{ik} + e_{ijk}$$

$$m = 1, 2, \ldots, M$$

Spearman's rank correlation between $\phi_1, \phi_2, \ldots, \phi_M$ and $\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_M$ was computed.

For each value of N, 1000 rank correlations were obtained.



Spring barley, five-year series

J. Forkman

Rank correlation (five-year series, spring barley)

		Probability (%)					
Number of trials	95	90	75	50	25	10	5
1	0.57	0.62	0.70	0.77	0.83	0.87	0.89
3	0.77	0.80	0.85	0.89	0.92	0.94	0.94
5	0.83	0.85	0.89	0.91	0.94	0.95	0.96
7	0.86	0.88	0.90	0.93	0.95	0.96	0.97
10	0.88	0.90	0.93	0.94	0.96	0.97	0.98
12	0.88	0.90	0.93	0.95	0.96	0.97	0.98
15	0.89	0.91	0.93	0.95	0.97	0.98	0.98
20	0.91	0.92	0.94	0.96	0.97	0.98	0.98
25	0.91	0.93	0.95	0.96	0.97	0.98	0.98
30	0.92	0.93	0.95	0.96	0.97	0.98	0.98
50	0.92	0.94	0.95	0.97	0.98	0.98	0.99
100	0.93	0.94	0.96	0.97	0.98	0.98	0.99

Final remarks

- Results from small series can be very misleading
- Ten trials per year suffices for ranking in five-year series
- Less trials are needed in winter wheat than in spring barley
- Results are dependent on assumed variances
- Some years, variety distributions were negatively skewed
- Simulation gives many answers, but also many questions

CBCS
HTTP://AGROBIOL.SGGW.WAW.PL/CBCS

COMMUNICATIONS IN BIOMETRY AND CROP SCIENCE
Vol. 11, No. 2, 2016, pp. 164–180

Forkman, J. (2016). How many trials are needed in crop variety testing? *Communications in Biometry and Crop Science* 11, 164-180.