Supplementary materials for ‘Openness and computational reproducibility in plant pathology: where do we stand and a way forward’

Adam H. Sparks1,2,✉, Emerson M. Del Ponte3, Kaique S. Alves3, Zachary S. L. Foster4, and Niklaus J. Grünwald4

1 Department of Primary Industries and Regional Development, Farming Systems Innovation, Perth WA 6000, Australia  
2 University of Southern Queensland, Centre for Crop Health, Toowoomba Qld 4350, Australia  
3 Departmento de Fitopatologia, Universidade Federal de Viçosa, Brazil  
4 Horticultural Crops Research Unit, USDA Agricultural Research Service, Corvallis OR 97330, USA

✉ Corresponding author: [Adam H. Sparks <[Adam.Sparks@dpird.wa.gov.au](mailto:Adam.Sparks@dpird.wa.gov.au)>](mailto:Adam.Sparks@dpird.wa.gov.au)

# Supplementary Materials

## Supplementary Tables

Table 1: Full description of the model that was fit to scoring data, which were used to evaluate the affect of journal title on code availability for 450 papers published in 21 plant pathology journals or plant pathology focused articles from other specialized journals. Scoring for 'Code Availability' was scored 0--3 where, '0' was 'Not available or not mentioned in the publication'; '1' was 'Available upon request to the author; '2' was 'Online, but inconvenient or non-permanent (e.g., login needed, paywall, FTP server, personal lab website that may disappear, or may have already disappeared)'; and '3' was 'Freely available online to anonymous users for foreseeable future (e.g., archived using Zenodo, dataverse or university library or some other proper archiving system)'; 'NA' indicates that no code was created to conduct the work that was detectable. We fit a Bayesian logistic mixed model (estimated using MCMC sampling with 4 chains of 10000 iterations and a warmup of 5000) to predict comp\_mthds\_avail with abbreviation (formula: comp\_mthds\_avail ~ abbreviation). The model included assignee as random effect (formula: ~1 | assignee). Priors over parameters were set as normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00) and student\_t (location = 0.00, scale = 2.50) distributions.

| Parameter | Median | CI | CI Low | CI High | pd | Rhat | ESS | Effect Size |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AustralasPlantPath | -0.12 | 0.95 | -2.03 | 1.65 | 0.55 | 1.00 | 30,384.50 | very small |
| CanJPlantPathol | -0.18 | 0.95 | -2.06 | 1.57 | 0.58 | 1.00 | 31,898.52 | very small |
| CropProt | -0.22 | 0.95 | -2.02 | 1.44 | 0.60 | 1.00 | 30,761.34 | very small |
| EurJPlantPathol | -0.17 | 0.95 | -2.02 | 1.54 | 0.57 | 1.00 | 28,146.82 | very small |
| ForestPathol | -0.20 | 0.95 | -2.00 | 1.50 | 0.58 | 1.00 | 31,905.00 | very small |
| JPhytopathol | -0.33 | 0.95 | -2.08 | 1.30 | 0.65 | 1.00 | 31,231.12 | very small |
| JPlantPathol | -0.21 | 0.95 | -2.05 | 1.48 | 0.59 | 1.00 | 29,063.96 | very small |
| MolPlantMicroIn | 0.49 | 0.95 | -1.24 | 2.06 | 0.72 | 1.00 | 30,218.34 | very small |
| MolPlantPathol | -0.26 | 0.95 | -2.07 | 1.39 | 0.61 | 1.00 | 31,148.37 | very small |
| Nematology | -0.20 | 0.95 | -2.05 | 1.50 | 0.59 | 1.00 | 31,651.84 | very small |
| PhysiolMolPlantP | -0.21 | 0.95 | -2.03 | 1.46 | 0.60 | 1.00 | 30,455.31 | very small |
| Phytoparasitica | -0.24 | 0.95 | -2.10 | 1.44 | 0.61 | 1.00 | 30,266.81 | very small |
| PhytopatholMediterr | -0.19 | 0.95 | -2.10 | 1.54 | 0.58 | 1.00 | 29,854.90 | very small |
| PlantDis | -0.17 | 0.95 | -2.05 | 1.54 | 0.57 | 1.00 | 26,753.21 | very small |
| PlantHealthProgress | -0.15 | 0.95 | -2.02 | 1.61 | 0.56 | 1.00 | 30,382.01 | very small |
| PlantPathol | -0.24 | 0.95 | -2.08 | 1.43 | 0.61 | 1.00 | 32,634.40 | very small |
| RevMexFitopatol | -0.22 | 0.95 | -2.08 | 1.42 | 0.60 | 1.00 | 29,507.33 | very small |
| TropPlantPathol | 0.66 | 0.95 | -1.16 | 2.34 | 0.77 | 1.00 | 29,671.83 | small |
| VirolJ | -0.13 | 0.95 | -1.98 | 1.60 | 0.56 | 1.00 | 31,229.60 | very small |

Table 2: Full description of the model that was fit to scoring data, which were used to evaluate the affect of journal title on data availability for 450 papers published in 21 plant pathology journals or plant pathology focused articles from other specialized journals. Scoring for 'Data Availability' was scored 0--3 where, '0' was 'Not available or not mentioned in the publication'; '1' was 'Available upon request to the author; '2' was 'Online, but inconvenient or non-permanent (e.g., login needed, paywall, FTP server, personal lab website that may disappear, or may have already disappeared)'; and '3' was 'Freely available online to anonymous users for foreseeable future (e.g., archived using Zenodo, dataverse or university library or some other proper archiving system)'; 'NA' indicates that no data were generated, e.g., a methods paper. We fit a Bayesian logistic mixed model (estimated using MCMC sampling with 4 chains of 10000 iterations and a warmup of 5000) to predict comp\_mthds\_avail with abbreviation (formula: comp\_mthds\_avail ~ abbreviation). The model included assignee as random effect (formula: ~1 | assignee). Priors over parameters were set as normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00) and student\_t (location = 0.00, scale = 2.50) distributions.

| Parameter | Median | CI | CI Low | CI High | pd | Rhat | ESS | Effect Size |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AustralasPlantPath | 0.62 | 0.95 | -0.71 | 1.84 | 0.83 | 1.00 | 22,512.63 | small |
| CanJPlantPathol | 0.30 | 0.95 | -0.83 | 1.32 | 0.70 | 1.00 | 20,869.52 | very small |
| CropProt | -1.23 | 0.95 | -2.73 | 0.02 | 0.97 | 1.00 | 20,199.84 | small |
| EurJPlantPathol | 0.08 | 0.95 | -1.11 | 1.15 | 0.55 | 1.00 | 21,647.41 | very small |
| ForestPathol | -0.09 | 0.95 | -1.26 | 0.95 | 0.57 | 1.00 | 21,567.86 | very small |
| JPhytopathol | -0.25 | 0.95 | -1.22 | 0.62 | 0.71 | 1.00 | 18,003.18 | very small |
| JPlantPathol | 0.15 | 0.95 | -0.94 | 1.14 | 0.62 | 1.00 | 20,336.25 | very small |
| MolPlantMicroIn | 0.54 | 0.95 | -0.43 | 1.47 | 0.87 | 1.00 | 19,561.11 | small |
| MolPlantPathol | 0.94 | 0.95 | 0.10 | 1.75 | 0.99 | 1.00 | 16,326.91 | small |
| Nematology | -0.19 | 0.95 | -1.51 | 0.95 | 0.62 | 1.00 | 22,360.61 | very small |
| PhysiolMolPlantP | 0.59 | 0.95 | -0.34 | 1.47 | 0.89 | 1.00 | 17,765.30 | small |
| Phytoparasitica | -0.47 | 0.95 | -1.72 | 0.62 | 0.79 | 1.00 | 21,343.22 | very small |
| PhytopatholMediterr | 1.68 | 0.95 | 0.78 | 2.57 | 1.00 | 1.00 | 17,609.88 | medium |
| PlantDis | -1.26 | 0.95 | -2.74 | -0.02 | 0.98 | 1.00 | 22,282.14 | medium |
| PlantHealthProgress | -0.59 | 0.95 | -2.01 | 0.64 | 0.82 | 1.00 | 21,371.07 | small |
| PlantPathol | -0.08 | 0.95 | -1.08 | 0.83 | 0.57 | 1.00 | 19,017.40 | very small |
| RevMexFitopatol | -1.15 | 0.95 | -2.64 | 0.12 | 0.96 | 1.00 | 24,184.64 | small |
| TropPlantPathol | 0.35 | 0.95 | -0.76 | 1.39 | 0.74 | 1.00 | 20,363.83 | very small |
| VirolJ | 0.87 | 0.95 | -0.03 | 1.73 | 0.97 | 1.00 | 17,099.91 | small |

Table 3: Full description of the model that was fit to scoring data, which were used to evaluate the affect of the year of publication on code availability for 450 papers published in 21 plant pathology journals or plant pathology focused articles from other specialized journals. Scoring for 'Code Availability' was scored 0--3 where, '0' was 'Not available or not mentioned in the publication'; '1' was 'Available upon request to the author; '2' was 'Online, but inconvenient or non-permanent (e.g., login needed, paywall, FTP server, personal lab website that may disappear, or may have already disappeared)'; and '3' was 'Freely available online to anonymous users for foreseeable future (e.g., archived using Zenodo, dataverse or university library or some other proper archiving system)'; 'NA' indicates that no code was created to conduct the work that was detectable. We fit a Bayesian logistic mixed model (estimated using MCMC sampling with 4 chains of 10000 iterations and a warmup of 5000) to predict comp\_mthds\_avail with abbreviation (formula: comp\_mthds\_avail ~ abbreviation). We fit a Bayesian logistic mixed model (estimated using MCMC sampling with 4 chains of 10000 iterations and a warmup of 5000) to predict comp\_mthds\_avail with year (formula: comp\_mthds\_avail ~ year). The model included abbreviation and assignee as random effects (formula: list(~1 | abbreviation, ~1 | assignee)). Priors over parameters were set as normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), student\_t (location = 0.00, scale = 2.50) and student\_t (location = 0.00, scale = 2.50) distributions.

| Parameter | Median | CI | CI Low | CI High | pd | Rhat | ESS | Effect Size |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| year | 0.20 | 0.95 | -0.13 | 0.58 | 0.88 | 1.00 | 19,642.86 | very small |

Table 4: Full description of the model that was fit to scoring data, which were used to evaluate the affect of the year of publication on data availability for 450 papers published in 21 plant pathology journals or plant pathology focused articles from other specialized journals. Scoring for 'Data Availability' was scored 0--3 where, '0' was 'Not available or not mentioned in the publication'; '1' was 'Available upon request to the author; '2' was 'Online, but inconvenient or non-permanent (e.g., login needed, paywall, FTP server, personal lab website that may disappear, or may have already disappeared)'; and '3' was 'Freely available online to anonymous users for foreseeable future (e.g., archived using Zenodo, dataverse or university library or some other proper archiving system)'; 'NA' indicates that no data were generated, e.g., a methods paper. We fit a Bayesian logistic mixed model (estimated using MCMC sampling with 4 chains of 10000 iterations and a warmup of 5000) to predict comp\_mthds\_avail with abbreviation (formula: comp\_mthds\_avail ~ abbreviation). We fit a Bayesian logistic mixed model (estimated using MCMC sampling with 4 chains of 10000 iterations and a warmup of 5000) to predict data\_avail with year (formula: data\_avail ~ year). The model included abbreviation and assignee as random effects (formula: list(~1 | abbreviation, ~1 | assignee)). Priors over parameters were set as normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), student\_t (location = 0.00, scale = 2.50) and student\_t (location = 0.00, scale = 2.50) distributions.

| Parameter | Median | CI | CI Low | CI High | pd | Rhat | ESS | Effect Size |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| year | 0.08 | 0.95 | -0.01 | 0.18 | 0.96 | 1.00 | 21,691.79 | very small |

Table 5: Full description of the model that was fit to scoring data, which were used to evaluate the affect of five-year impact factor on code availability for 450 papers published in 21 plant pathology journals or plant pathology focused articles from other specialized journals. Scoring for 'Code Availability' was scored 0--3 where, '0' was 'Not available or not mentioned in the publication'; '1' was 'Available upon request to the author; '2' was 'Online, but inconvenient or non-permanent (e.g., login needed, paywall, FTP server, personal lab website that may disappear, or may have already disappeared)'; and '3' was 'Freely available online to anonymous users for foreseeable future (e.g., archived using Zenodo, dataverse or university library or some other proper archiving system)'; 'NA' indicates that no code was created to conduct the work that was detectable. We fit a Bayesian logistic mixed model (estimated using MCMC sampling with 4 chains of 10000 iterations and a warmup of 5000) to predict comp\_mthds\_avail with year (formula: comp\_mthds\_avail ~ year). The model included abbreviation and assignee as random effects (formula: list(~1 | abbreviation, ~1 | assignee)). Priors over parameters were set as normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), student\_t (location = 0.00, scale = 2.50) and student\_t (location = 0.00, scale = 2.50) distributions.

| Parameter | Median | CI | CI Low | CI High | pd | Rhat | ESS | Effect Size |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| IF\_5year | 0.45 | 0.95 | -0.05 | 1.04 | 0.96 | 1.00 | 13,935.20 | small |

Table 6: Full description of model fit for the affect of five-year impact factor on data availability. Scoring for 'Data Availability' was scored 0--3 where, '0' was 'Not available or not mentioned in the publication'; '1' was 'Available upon request to the author; '2' was 'Online, but inconvenient or non-permanent (e.g., login needed, paywall, FTP server, personal lab website that may disappear, or may have already disappeared)'; and '3' was 'Freely available online to anonymous users for foreseeable future (e.g., archived using Zenodo, dataverse or university library or some other proper archiving system)'; 'NA' indicates that no data were generated, e.g., a methods paper. We fit a Bayesian logistic mixed model (estimated using MCMC sampling with 4 chains of 10000 iterations and a warmup of 5000) to predict comp\_mthds\_avail with year (formula: comp\_mthds\_avail ~ year). The model included abbreviation and assignee as random effects (formula: list(~1 | abbreviation, ~1 | assignee)). Priors over parameters were set as normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), normal (mean = 0.00, SD = 1.00), student\_t (location = 0.00, scale = 2.50) and student\_t (location = 0.00, scale = 2.50) distributions.

| Parameter | Median | CI | CI Low | CI High | pd | Rhat | ESS | Effect Size |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| year | 0.08 | 0.95 | -0.01 | 0.18 | 0.96 | 1.00 | 21,691.79 | very small |

## Supplementary Figure

![Figure 1: Criteria scores for 450 articles computational materials and data availability for each of the five evaluators. Each article was evaluated on a 0 to 3 scale for computational materials (Code) and raw data availability (Data) by one of five evaluators. Scoring for ‘Code Availability’ was scored 0–3 where, ‘0’ was ‘Not available or not mentioned in the publication’; ‘1’ was ‘Available upon request to the author; ’2’ was ‘Online, but inconvenient or non-permanent (e.g., login needed, paywall, FTP server, personal lab website that may disappear, or may have already disappeared)’; and ‘3’ was ‘Freely available online to anonymous users for foreseeable future (e.g., archived using Zenodo, dataverse or university library or some other proper archiving system)’; ‘NA’ indicates that no code was created to conduct the work that was detectable. And the scoring for ‘Data Availability’ was scored 0–3 where, ‘0’ was ‘Not available or not mentioned in the publication’; ‘1’ was ‘Available upon request to the author; ’2’ was ‘Online, but inconvenient or non-permanent (e.g., login needed, paywall, FTP server, personal lab website that may disappear, or may have already disappeared)’; and ‘3’ was ‘Freely available online to anonymous users for foreseeable future (e.g., archived using Zenodo, dataverse or university library or some other proper archiving system)’; ‘NA’ indicates that no data were generated, e.g., a methods paper.](data:application/eps;base64,)

Figure 1: Criteria scores for 450 articles computational materials and data availability for each of the five evaluators. Each article was evaluated on a 0 to 3 scale for computational materials (Code) and raw data availability (Data) by one of five evaluators. Scoring for ‘Code Availability’ was scored 0–3 where, ‘0’ was ‘Not available or not mentioned in the publication’; ‘1’ was ‘Available upon request to the author; ’2’ was ‘Online, but inconvenient or non-permanent (e.g., login needed, paywall, FTP server, personal lab website that may disappear, or may have already disappeared)’; and ‘3’ was ‘Freely available online to anonymous users for foreseeable future (e.g., archived using Zenodo, dataverse or university library or some other proper archiving system)’; ‘NA’ indicates that no code was created to conduct the work that was detectable. And the scoring for ‘Data Availability’ was scored 0–3 where, ‘0’ was ‘Not available or not mentioned in the publication’; ‘1’ was ‘Available upon request to the author; ’2’ was ‘Online, but inconvenient or non-permanent (e.g., login needed, paywall, FTP server, personal lab website that may disappear, or may have already disappeared)’; and ‘3’ was ‘Freely available online to anonymous users for foreseeable future (e.g., archived using Zenodo, dataverse or university library or some other proper archiving system)’; ‘NA’ indicates that no data were generated, e.g., a methods paper.