











# MetaComNet: A random forest-based framework for making spatial predictions of plant–pollinator interactions

Markus A. K. Sydenham<sup>1</sup>  | Zander S. Venter<sup>1</sup>  | Trond Reitan<sup>2</sup>  |  
 Claus Rasmussen<sup>3</sup>  | Astrid B. Skrindo<sup>1</sup>  | Daniel I. J. Skoog<sup>4</sup> | Kaj-Andreas Hanevik<sup>4</sup> |  
 Stein Joar Hegland<sup>5</sup>  | Yoko L. Dupont<sup>6</sup>  | Anders Nielsen<sup>7,2</sup>  |  
 Joseph Chipperfield<sup>8</sup>  | Graciela M. Rusch<sup>9</sup> 

<sup>1</sup>The Norwegian Institute for Nature Research, Oslo, Norway; <sup>2</sup>Department of Biosciences, Centre for Ecological and Evolutionary Synthesis (CEES), University of Oslo, Oslo, Norway; <sup>3</sup>Department of Agroecology, Aarhus University, Foulum, Denmark; <sup>4</sup>Faculty of Environmental Sciences and Natural Resource Management, Norwegian University of Life Sciences, Ås, Norway; <sup>5</sup>Department of Environmental Sciences, Western University of Applied Sciences, Sogndal, Norway; <sup>6</sup>Department of Ecoscience, Aarhus University, Rønde, Denmark; <sup>7</sup>Department of Landscape and Biodiversity, Norwegian Institute of Bioeconomy Research (NIBIO), Ås, Norway; <sup>8</sup>The Norwegian Institute for Nature Research, Bergen, Norway and <sup>9</sup>The Norwegian Institute for Nature Research, Trondheim, Norway

## Correspondence

Markus A. K. Sydenham  
 Email: markus.sydenham@nina.no

## Funding information

Norges Forskningsråd, Grant/Award Number: 302692

Handling Editor: Luisa Carvalheiro

## Abstract

1. Predicting plant–pollinator interaction networks over space and time will improve our understanding of how environmental change is likely to impact the functioning of ecosystems. Here we propose a framework for producing spatially explicit predictions of the occurrence and number of pairwise plant–pollinator interactions and of the species richness, diversity and abundance of pollinators visiting flowers. We call the framework ‘MetaComNet’ because it aims to link metacommunity dynamics to the assembly of ecological networks.
2. To illustrate the MetaComNet functionality, we used a dataset on bee–flower networks sampled at 16 sites in southeast Norway along with random forest models to predict bee–flower interactions. We included variables associated with climatic conditions (elevation) and habitat availability within a 250 m radius of each site. Regional commonness, site-specific distance to conspecifics, social guild and floral preference were included as bee traits. Each plant species was assigned a score reflecting its site-specific abundance, and four scores reflecting the bee species that the plant family is known to attract. We used leave-one-out cross-validations to assess the models' ability to predict pairwise plant–bee interactions across the landscape.
3. The relationship between observed occurrence or absence of interactions and the predicted probability of interactions was nearly proportional ( $\text{GLM}_{\text{logistic regression slope}} = 1.09$ ), matching the data well ( $\text{AUC} = 0.88$ ), and explained 30% of the variation. Predicted probability of interactions was also correlated with the number of observed pairwise interactions ( $r = 0.32$ ). The sum

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society

of predicted probabilities of bee–flower interactions were positively correlated with observed species richness ( $r = 0.50$ ), diversity ( $r = 0.48$ ) and abundance ( $r = 0.42$ ) of wild bees interacting with plant species within sites.

- Our findings show that the MetaComNet framework can be a useful approach for making spatially explicit predictions and mapping plant–pollinator interactions. Such predictions have the potential to identify areas where the pollination potential for wild plants is particularly high, and where conservation action should be directed to preserve this ecosystem function.

#### KEYWORDS

interactions, network, plants, pollinators, predict, random forest

## 1 | INTRODUCTION

Nearly nine of every 10 species of flowering plants rely on interactions, mainly with insects, for cross-pollination (Ollerton et al., 2011). However, wild plants are experiencing a shortage of pollinators in both natural and managed terrestrial ecosystems (Bennett et al., 2020), which can considerably affect plant reproduction and persistence (e.g. Gomez et al., 2010; Ollerton, 2017; Thomann et al., 2013). Our ability to predict how plant communities will respond to environmental change partly relies on our ability to predict how plant–pollinator interactions vary in heterogeneous landscapes (Tylianakis & Morris, 2017).

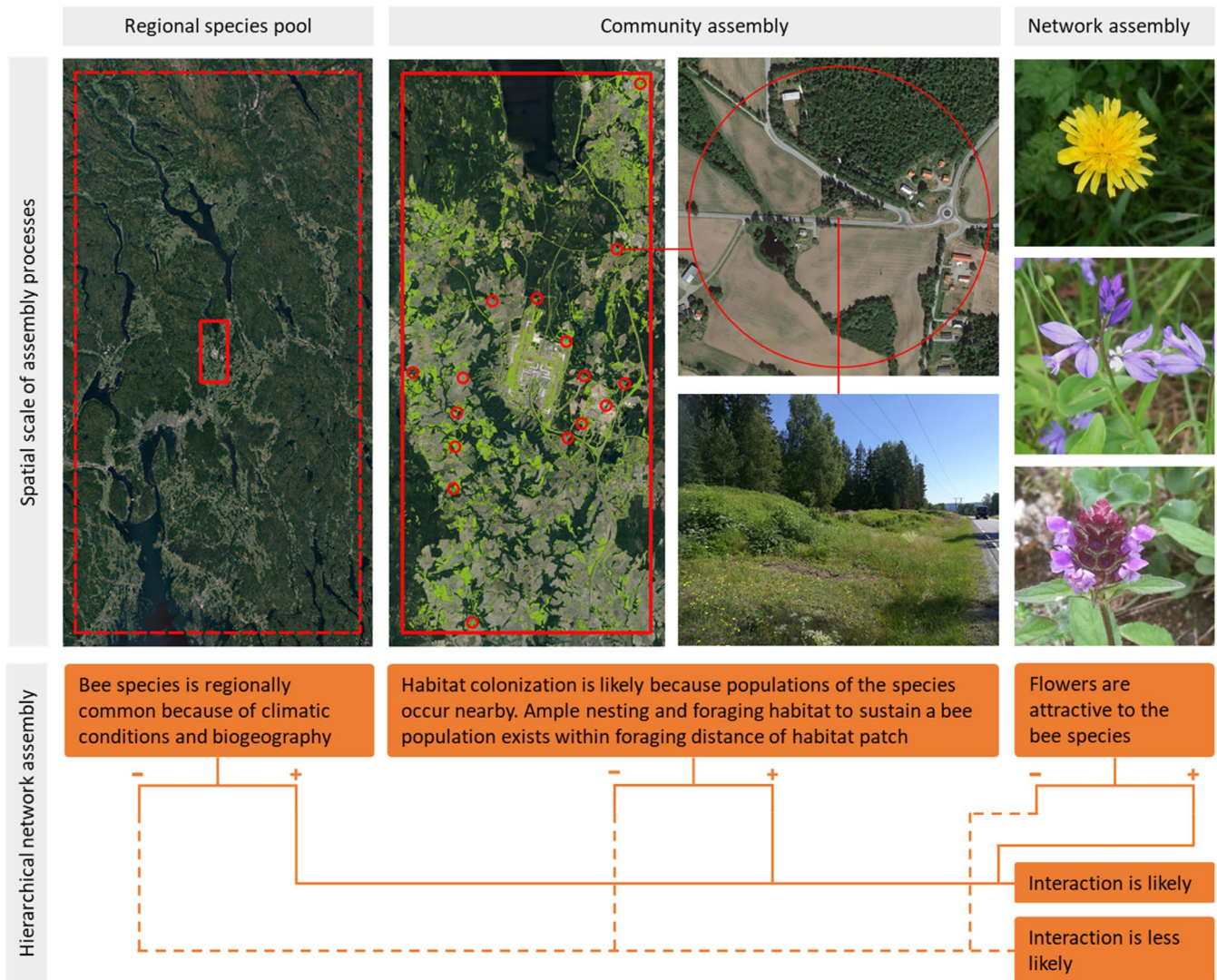
Most approaches to modelling plant–pollinator interactions in heterogeneous landscapes have focused on network properties, such as modularity (reviewed in Pellissier et al., 2018), but these properties may not fully encapsulate the underlying community assembly processes (Olito & Fox, 2015). An alternative approach is to model pairwise plant–pollinator interactions directly, and to derive network properties by aggregating model predictions (Graham & Weinstein, 2018), for instance by including plant and pollinator abundances, traits and phylogenies (e.g. Benadi et al., in press; Pichler et al., 2020; Stock et al., 2021). However, Benadi et al. (in press) found a drop in model performance when predicting into novel habitats, which may arise from differences in habitat environmental conditions influencing pollinator distributions (e.g. Hoiss et al., 2012). Hence, accounting for processes behind community assembly (as per Vellend, 2016) is central for making spatial predictions of plant–pollinator interactions. We therefore hypothesise that the number of pairwise interactions between plants and pollinators can be effectively modelled as a function of plant and pollinator affiliations *and* of variables underlying pollinator community assembly at different spatial scales. We propose a conceptual framework to make spatial predictions of pairwise interactions between plants and pollinators, which we refer to as ‘MetaComNet’ because it aims at linking meta-community structuring factors to the structure of ecological networks.

MetaComNet is pollinator-oriented in that it focuses on predicting the occurrence, or number, of plant–pollinator interactions by

modelling the distribution of wild bees across plant species while considering regional and landscape level factors. The response variables considered are the number, or presence/absence, of observed interactions between pollinator and plant species observed in specific localities. We focus on pollinators because they tend to display more pronounced spatial turnover in interaction networks than plants do (Trøjelsgaard et al., 2015). MetaComNet builds on Tylianakis and Morris (2017) who define the occurrence, or number, of interactions between plants and pollinators as the endproduct of processes and conditions that determine species composition at different spatial scales (Figure 1). We pose the following hypotheses relevant at different spatial scales:

*Regional scale:* If community, or network, assembly is ecologically neutral, the abundance of species within communities (Vellend, 2016) will be proportional to species commonness in the regional species pool, in turn dependent on climatic requirements and the biogeography of species (Cornell & Harrison, 2014). Accounting for differences in regional commonness is also important because species abundance is related to the level of random interactions (reviewed in Krishna et al., 2008; Tylianakis & Morris, 2017; Vázquez et al., 2007). In MetaComNet, the regional commonness of pollinators is included as a regional level predictor of pairwise interactions to account for neutral network assembly processes.

*Landscape scale:* Pollinator communities are assembled through dispersal processes (Hagen et al., 2012) and through mechanisms of species sorting determined by the suitability of an area as habitat for the species (environmental filtering). Dispersal rates depend in part on geographic distance and barriers (Carstensen et al., 2014; Trøjelsgaard et al., 2015); hence the likelihood of a species occurring in a habitat patch, and of it interacting with the plants therein, is expected to decrease with the distance to the nearest population in the region. In MetaComNet, we accounted for the influence of immigration rates by including site- and species-specific estimates of the distance to the nearest population. To support viable pollinator populations, a location must contain enough nesting and foraging resources within enough proximity (Westrich, 1996). The amount of semi-natural habitat or degree of landscape diversity within a 250 m radius can be used as a proxies for habitat amount since solitary bee



**FIGURE 1** Conceptualisation of the MetaComNet framework and the hierarchical assembly of plant-bee interaction networks. The likelihood of a bee interacting with a plant species growing in a habitat patch is a product of a series of scale-dependent conditions, that is, *regional* commonness of pollinators, *landscape* habitat suitability (nesting and foraging resources) and distance to neighbouring populations, and *local* composition of flower plants, their abundance and level of attractiveness

species richness increases with habitat area at this scale (Steffan-Dewenter et al., 2002). The distance to soil deposits with high sand concentrations can be used as a proxy for distance to high quality nesting substrates, because many species prefer such substrates for nest sites (Antoine & Forrest, 2021; Heneberg et al., 2013).

**Local scale:** At the scale of individual flowers within habitats, the occurrence and number of bee–plant interactions will depend on the attractiveness of the flower. Flower attractiveness depends in part on their relative abundance (Fowler et al., 2016; Stavert et al., 2019), even though visitation rates to flowers may saturate (Totland, 1994) or show unimodal responses (Benadi & Pauw, 2018). Other factors determining flower attractiveness to pollinators are, for example, particular morphological characteristics (e.g. Benadi et al., in press; Pichler et al., 2020; Stock et al., 2021), non-visual cues such as floral scent (Larue et al., 2016) and pollen toxicity that require adaptations to overcome (reviewed in Rivest & Forest, 2020). Trait-matching

reduces the frequency or even excludes some combinations of partners in plant–pollinator networks (Olesen et al., 2011). However, which floral traits select for specific bees is not always easy to predict and species may respond to different traits on the same plant (Rowe et al., 2020). Also, in bumble bees, pollen preferences can be more directly related to phylogenetic relationships than to proboscis length (Wood et al., 2021). In MetaComNet, trait-matching between plants and pollinators can be accounted for by assigning a set of functional and/or floral preference traits reflecting the expected plant–pollinator associations.

The MetaComNet model integrates data from the three geographic levels indicated above into a data frame illustrated in Table 1. The model parameters include response variables (number, or the presence or absence of interactions), grouping variables (pollinator species, plant species, site identity) and predictor variables such as pollinator traits, plant traits and site-specific environmental

**TABLE 1** Variables included in the MetaComNet network model. The data frame contains columns with response variables including: (i) number, or presence or absence, of observed interactions between a pollinator species and a plant species in a particular study site. Grouping variables include: pollinator species (Pol.), plant species (Plnt.) and site identity (Site). Predictor variables that can be linked to pollinator species, plant species or the site. Pollinator variables include the regional commonness (RC); the distance to the nearest known population of that species (Dst); and traits (Tr8s) related to environmental requirements or the floral preference of the species. Plant variables include the local commonness (LC), or abundance, and traits (TR8s) related to pollinator affiliations. Site-specific variables include geographic coordinates that are used for calculating distances to potential source populations and for extracting georeferenced environmental variables (Env) such as m a.s.l., or area of semi-natural habitat within 250 m radius in the surrounding landscape

Response variables		Grouping variables			Predictor variables		
Number	Presence	Pol.	Plnt.	Site	Pol.	Plant.	Site
0	0	Pol.1	Plnt.1	S.1	RC; Dst; Tr8 <sub>1-n</sub>	LC; Tr8 <sub>1-n</sub>	xy; Env <sub>1-n</sub>
5	1	Pol.2	Plnt.1	S.1	RC; Dst; Tr8 <sub>1-n</sub>	LC; Tr8 <sub>1-n</sub>	xy; Env <sub>1-n</sub>
2	1	Pol.1	Plnt.2	S.1	RC; Dst; Tr8 <sub>1-n</sub>	LC; Tr8 <sub>1-n</sub>	xy; Env <sub>1-n</sub>
10	1	Pol.2	Plnt.2	S.1	RC; Dst; Tr8 <sub>1-n</sub>	LC; Tr8 <sub>1-n</sub>	xy; Env <sub>1-n</sub>
0	0	Pol.1	Plnt.1	S.2	RC; Dst; Tr8 <sub>1-n</sub>	LC; Tr8 <sub>1-n</sub>	xy; Env <sub>1-n</sub>
0	0	Pol.2	Plnt.1	S.2	RC; Dst; Tr8 <sub>1-n</sub>	LC; Tr8 <sub>1-n</sub>	xy; Env <sub>1-n</sub>
2	1	Pol.1	Plnt.2	S.2	RC; Dst; Tr8 <sub>1-n</sub>	LC; Tr8 <sub>1-n</sub>	xy; Env <sub>1-n</sub>
0	0	Pol.2	Plnt.2	S.2	RC; Dst; Tr8 <sub>1-n</sub>	LC; Tr8 <sub>1-n</sub>	xy; Env <sub>1-n</sub>
...	...	...	...	...	...	...	...

conditions. Using empirical data from Norway, we illustrate the ability of MetaComNet to predict the occurrence and number of interactions between wild bees and plants.

MetaComNet uses a random forest modelling framework (Breiman, 2001) to model pollinator–plant associations because of its ability to define nonlinear interactions between predictor variables. Random forest models also offer an intuitive way to view the hierarchical assembly of pairwise interaction networks (Figure 1). Moreover, random forest models have been shown to outperform other machine learning and glm-based techniques when predicting pairwise plant–pollinator interactions in both simulated and empirical data (Pichler et al., 2020) and to perform as well as mechanistic likelihood-based models (Benadi et al., in press). We used three random-forest approaches to test if predictor variables (see Table 2) could predict spatial variation in:

- Occurrence and number of pairwise interactions between wild bee species and plants.
- Species richness, Shannon diversity and abundance of wild bee floral visitors.

## 2 | MATERIALS AND METHODS

### 2.1 | Bee–flower network sampling

We sampled bee–flower networks along 16 roadsides (sites) in Southeast Norway in 2017. Eight of the study sites were located on sandy sediments and the remaining eight were located on clay dominated sediments (Skoog, 2018). At each study site, flower-visiting bees were collected during 1 hr by two observers along a 50 m transect, once during early to mid-July and once during early

to mid-August in 2017. A total of 910 interactions between wild bees ( $n = 45$  species) and plants ( $n = 44$  species) were observed. We only included non-parasitic bees ( $n = 39$ ) in our analyses because parasitic bees do not visit plants for nectar and pollen. We estimated the site's abundance of flowering plant species by recording the number of occurrences (presence/absence) of each plant species in a grid of sixteen  $25 \times 25$  cm squares within six  $1 \text{ m}^2$  quadrats placed in a 2 by 30 m grid along the roadside, totalling 96 squares per site. No permissions were required to conduct the fieldwork.

The site-specific plant abundance was used as a proxy for the number of flowers from that species within a site, and was based on species which were flowering at the time of the inventory. Plants that did not occur within any of the six  $1 \text{ m}^2$  vegetation quadrats within a site, but with bee visitation records were given a plant cover abundance value of 0.05 to indicate that the plant was locally present.

We assigned variables to the recorded plants and bees that we expected would influence the number of plant–pollinator interactions. For plants we used the site abundance since we expected a high correlation of bee visitations with plant abundance. Because our aim was not to identify functional traits that determine bee–flower interactions, but to account for floral associations of bees in our models, we assembled a binary network of 207 bee species and 61 plant family interactions based on existing information from interaction records sampled at a greater temporal and spatial extent than our study area (Rasmussen et al., 2021; Wood et al., 2021). This approach provided a more inclusive measure of host plants of bees, closer to the fundamental niche, than what would be achieved from a single survey, such as ours. We used a detrended correspondence analysis (DCA) in the VEGAN package in R (Oksanen et al., 2018) to establish the four main axes of correspondence between plant families and non-parasitic bee species associations (Figure S1). We used the plant and bee DCA scores to account for plant–bee associations.

**TABLE 2** Description of variables included in the MetaComNet network model. The model data frame contained 9,594 rows defined by combinations of the grouping variables: bee species, plant species and site identity

Variable	Description
<b>Response variables</b>	
Occurrence of interactions	The presence or absence of interactions between the bee species and plant within a site. The variable was transformed into a two-level categorical variable for models using classification trees and left as a numeric variable (zero or one) for the models using regression trees
Number of interactions	Number of interactions between the bee species and plant within a site, used in regression trees
<b>Grouping variables</b>	
Bee species	Character string with 39 unique values, one per bee species
Plant species	Character string with 44 unique values, one per plant species
Site	Character string with 16 unique values, one per study site
<b>Predictor variables</b>	
Regional commonness	Number of 10 km grid cells within the region, occupied by the bee species
Distance to conspecifics	Geographic distance from the site to the nearest GBIF record of the bee species, ranging from 200 m to 45.5 km
Elevation	Elevation of the study site, random from 147 to 222 m a.s.l.
Grassland 250 m	Proportion of area classified as grassland within 250 m of the site, ranging from 5% to 70%
Landscape H 250 m	Landscape Shannon diversity (or heterogeneity) within 250 m of the site, ranging from 0.59 to 1.48
Distance to sandy soils	Geographic distance to soil deposits with high concentrations of sand, ranging from zero to 2,130 m
Plant abundance	Number of 25 × 25 cm squares, within which the plant species occurs, ranging from 0.05 to 64, out of 96 possible (i.e. occurring in all 16 25 × 25 cm squares in all 1 × 1 m plant quadrats)
Plant DCA1-4	Plant family scores along the first to fourth detrended correspondence axes, one column per axis
Bee DCA1-4	Bee species score along the first to fourth detrended correspondence axes, one column per axis
Bee sociality	Categorical variable indicating if the bee species belonged to the <i>Bombus</i> genus or not

In our model we distinguished between solitary and social wild bees (*Bombus* spp.) because solitary bee diversity responds more strongly to landscape conditions at local spatial scales than bumble bees (Steffan-Dewenter et al., 2002). We used records from GBIF.org (GBIF, 2021) to estimate the regional commonness of bee species and the distance to the nearest known occurrence of each species. We downloaded occurrence records covering our region (Figure 1,  $x_{\min} = 9.99$ ,  $x_{\max} = 12.19$ ,  $y_{\min} = 59.09$ ,  $y_{\max} = 61.31$ , projection: WGS84) and excluded records older than 20 years and with a coordinate uncertainty >100 m. To estimate regional commonness, we tallied the number of 10 km grid cells within which a species had been observed. For each site we calculated the geographic distance to the nearest GBIF record of each species.

For each site we obtained information on environmental conditions known to affect wild bee distributions. As a proxy for climatic conditions, we used elevation above sea level, obtained from a digital elevation model with a 50 m resolution (Norwegian Mapping Authority, 2016). As proxies for landscape conditions, we used the European ELC10 land cover map (Venter & Sydenham, 2021) to calculate: the proportion of grassland area and Shannon landscape

diversity within a 250 m radius. As a proxy for nesting conditions, we calculated the distance to sand-dominated geological deposits, that is, soils with a high permeability (Geological Survey of Norway, 2011).

## 2.2 | Modelling and predicting empirical bee-plant interactions

We assembled a data frame where each row was defined by a study site, a plant species found within the study site, and one of the bee species occurring in the 16 study sites (Tables 1 and 2). The data frame included (response) variables for the presence (1) or absence (0) as well as the number of visitations recorded of the bee species on the plant within the site. The predictor variables were the bee species-specific variables: regional commonness; site-specific distance to the nearest known occurrence; DCA scores; and *Bombus* versus non-*Bombus*. Plant species-specific variables were: DCA scores; and site-specific abundance. Site-specific variables were: Elevation; proportion of grassland and Shannon landscape diversity within a 250 m radius; and distance to geological deposits dominated by sand.

We used random forest (Breiman, 2001) with the *RANGER* package (Wright & Ziegler, 2017) via the *Caret* package (Kuhn, 2018) in *R* (R Core Team, 2020). We fitted three models to the data depending on the type of response variable: the presence versus absence of interactions, using classification trees; the presence (one) versus absence (zero) of interactions using regression trees; and number of interactions using regression trees. The three resulting model outputs were: predicted probability of interactions, that is, class probability; predicted frequency of interaction, that is, predicted proportion of presences; and predicted number of interactions. We used leave-one-out cross-validation, by iteratively training models on data from 15 sites and predicting onto the remaining site to allow assessing model performances across all 16 sites as well as their variability in terms of predicting pairwise interactions within sites. We also conducted leave-one-out cross-validations for each bee and plant species by iteratively training models on data (across sites) where the focal bee or plant species had been removed, and then attempted to predict pairwise interactions for the removed species. This approach allowed us to assess if models differed in their ability to predict plant interactions for bee species, and bee interactions for plant species, not encountered while training the models (Stock et al., 2021). To assess model fit during training, we used a fivefold cross-validation with accuracy, for classification trees, and root mean square error (RMSE), for regression trees. The hyperparameters in random forest models were tuned by cycling through a range of possible values, following default settings in *caret* (Kuhn, 2018), and selecting the combination of tuning parameters that resulted in the highest accuracy.

We tested if the predicted probability, frequency or number of interactions corresponded to the observed presence or absence of interactions. Predictions of interaction probabilities and frequencies were compared against actual occurrences of interactions by calculating the logistic GLM regression slopes, with the predicted probabilities logit-transformed so that regression slopes equal to one would indicate a 1:1 relationship between predicted probability, or frequency, of occurrence and observed proportion of occurrences. Because we did not expect a linear relationship (on the logit-scale) between occurrences and predicted number of interactions, we compared logistic GLMs with the number of interactions left un-transformed,  $\log(x + 1)$ , or square root transformed and selected the square root transformed model because it had the lowest Bayesian information criterion (BIC). For all three models we used Nagelkerkes log likelihood-based  $R^2$  from *R* package *MuMIn* (Barton, 2018), from the GLM models, and the area under the curve (AUC) from *R* package *pROC* (Robin et al., 2011) to assess model performance. We calculated the regression slopes for the first two models (with presence/absence as response variables), and the average  $R^2$  and AUC. To assess the models' power to predict the number of interactions, we calculated the Pearson correlations between the observed number of interactions and predicted probability, frequency and number of interactions. All validation metrics were calculated by (a) including predictions and observations from all 16 sites, (b) and by calculating the mean, and standard deviation from predictions for each site individually. We calculated the scaled importance of predictor variables from each of the 16 models in order

to assess if the variation in variable selection across models differed between the three modelling strategies.

### 2.3 | Predicting flower-visitor richness, diversity and abundance

We tested the level of correspondence between the sum of predicted pairwise bee–plant interactions and observed flower-visitor species richness, diversity and abundance within sites. For each of the three models, we calculated predicted flower-visitor species richness from the sum of predicted probabilities of interactions and from the sum of predicted frequencies of interactions for each plant species per site. We also calculated the predicted abundance of flower visitors as the sum of predicted number of interactions across bee species. We then calculated the Pearson correlation between predicted species richness or abundance and observed: flower-visitor species richness; species diversity; and species abundance.

### 2.4 | Mapping flower-visitor species richness

To illustrate how the predicted flower-visitor species richness can be mapped and thus used to identify areas where plants are most or least likely to be pollen limited, we re-fitted the random forest regression to the occurrence of interactions using all the data from the 16 sites and used this model to produce prediction maps. We created prediction maps of flower-visitor richness for plant species belonging to the *Compositae* and *Leguminosae* families. For each plant family and bee species combination, we created a data matrix with one row per cell. In addition to the environmental variables, each row contained the plant family specific DCA scores, the bee species-specific DCA scores, regional commonness, if the species belonged to the *Bombus* genus, and the distance of the centroid of the raster cell to the nearest occurrence of the bee species. We held the plant commonness constant at 16 or 50 indicating the number of 25 cm plots within the vegetation quadrats where the plant occurred, to illustrate how local floral abundance can affect flower-visitor diversity. To aid visual interpretations of the model predictions, we masked out areas with predicted species richness lower than the third quantile of predicted values.

## 3 | RESULTS

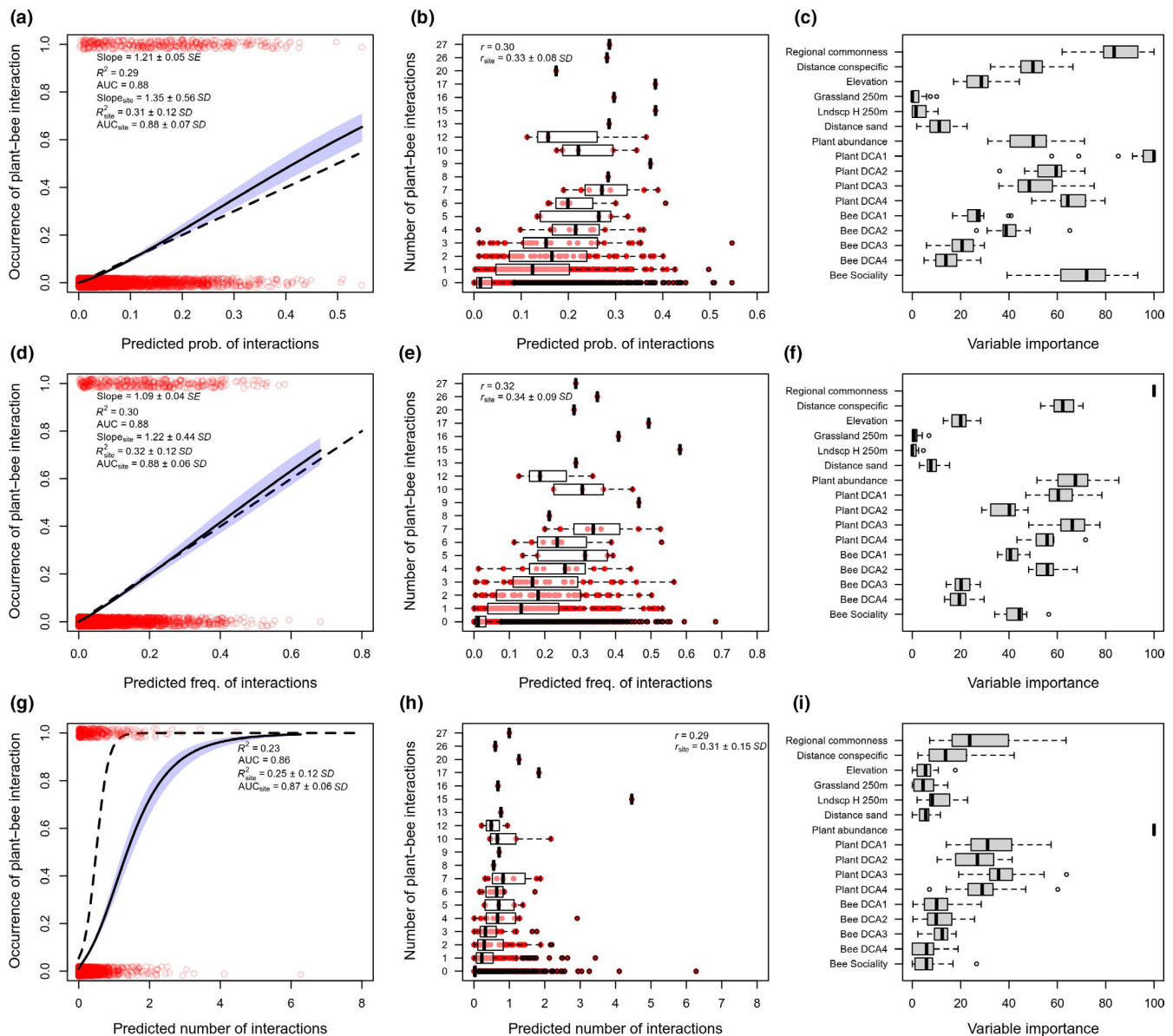
### 3.1 | Predicting pairwise interactions between bee and plant species

There was a positive relationship between observed and predicted occurrence and number of bee–plant interactions irrespective of the modelling strategy (Figure 2a–i). However, predictions from random forest classification trees (Figure 2a–c) and from random forest regressions (Figure 2d–f) explained more of the variation in observed occurrences of interactions than those from random forest regression

models on the number of interactions (Figure 2g–h). Compared to an ideal 1:1 relationship between observed and predicted occurrence or number of interactions, the random forest regression (i.e. with presences treated as one's and absences as zero's) was the most similar, with an average regression slope closest to one ( $1.09 \pm 0.04$  SE). For single site predictions, regression slopes varied less (lower standard deviation) for predictions from regression trees (Figure 2d) than those from classification trees (Figure 2a). Moreover, predictions from classification trees more often underestimated the occurrence of interactions, while predictions from the number of interactions overestimated

the occurrence of interactions. All model predictions were similarly correlated with the observed number of interactions (Figure 2b,e,h). Bee and plant species-based leave-one-out validations showed that classification trees and regression trees on occurrence or absence of interactions, performed equally well and outperformed regression trees based on number of interactions when attempting to predict pairwise interactions for bees and plant species (Figures S2 and S3).

Comparing predictor variable importance (Figure 2c,f,i) across the 16 models for each of the three modelling strategies also showed that models for interaction frequencies (Figure 2f) were generally



**FIGURE 2** Predictions of occurrences and number of pairwise interactions between wild bee and plant species. Results for models on: (a–c) predicted probability of interactions from classification trees; (d–f) predicted frequency of interactions, from regression trees on presences and absences; (g–i) and predicted number of interactions, from regression trees. Figures in the left panel (a, d, g) show logistic GLM regression curves for observed occurrences of interactions against the model predictions across the 16 study sites (networks). Ideal 1:1 relationships are shown as black dashed lines for reference. Figures in the mid panel (b, e, h) show data points in red together with boxplot summary statistics for the observed number of interactions against predicted values. Validation metrics are shown for validations across all 16 sites (Slope  $\pm$  SE,  $R^2$ , AUC and Person's  $r$ ) together with and summarised with means and standard deviations for within site validations (Slope<sub>site</sub>,  $R^2_{site}$ , AUC<sub>site</sub> and Person's  $r_{site}$ ). Figures in the right-hand panel (c, f, i) show the boxplot summary statistic for the relative importance of predictor variables across the 16 random forest models

more in agreement on the relative importance of variables, than those of interaction probabilities (Figure 2c) and number of interactions (Figure 2i). Regional commonness was the most important predictor of interaction frequencies, followed by the distance to conspecifics and variables associated with plant and pollinator associations (Figure 2f). For all three modelling strategies, elevation and landscape conditions were consistently among the least important variables in predicting pairwise interactions (Figure 2c,f,i).

### 3.2 | Predicting flower-visitor species richness, diversity and abundance

Predicted bee species richness and abundance were positively correlated with observed flower-visitor species richness, diversity and abundance. The Pearson correlation coefficient between observed flower-visitor species richness, diversity or abundance, and predicted species richness was similar for classification (Figure 3a–c) and regression random forest models (Figure 3d–e). In both cases median values for observed species richness, diversity and abundance, increased with predicted species richness but the rate of increase seemed to saturate at a predicted species richness at and above two. In comparison, correlations between predicted flower-visitor abundance and observed flower-visitor species richness, diversity and abundance were weaker (Figure 3g–i).

### 3.3 | Mapping flower-visitor species richness

The random forest regression models produced the most accurate predictions of observed pairwise interactions (Figure 2), and equal to or stronger relationships with the number of pairwise interactions, and species richness, diversity and abundance of flower visitors (Figure 3) than the other two models. It showed increasing flower abundance increased the predicted flower-visitor species richness to both Compositae (Figure 4b,c), and Leguminosae (Figure 4e,f). However, despite some overlap, there is a considerable difference between the two plant families in terms of where plants were predicted to receive the highest richness of floral visitors (Figure 4e–g). For Compositae, areas with predicted values at or above the third quantile were mainly concentrated on areas with sand dominated geological deposits (around the Gardermoen airport, Figure 4d; Figure S2). For Leguminosae, the area predicted with highest diversity of wild bees was found southwest of the Airport (Figure 4g), where the soil substrate is dominated by marine, clayish deposits.

## 4 | DISCUSSION

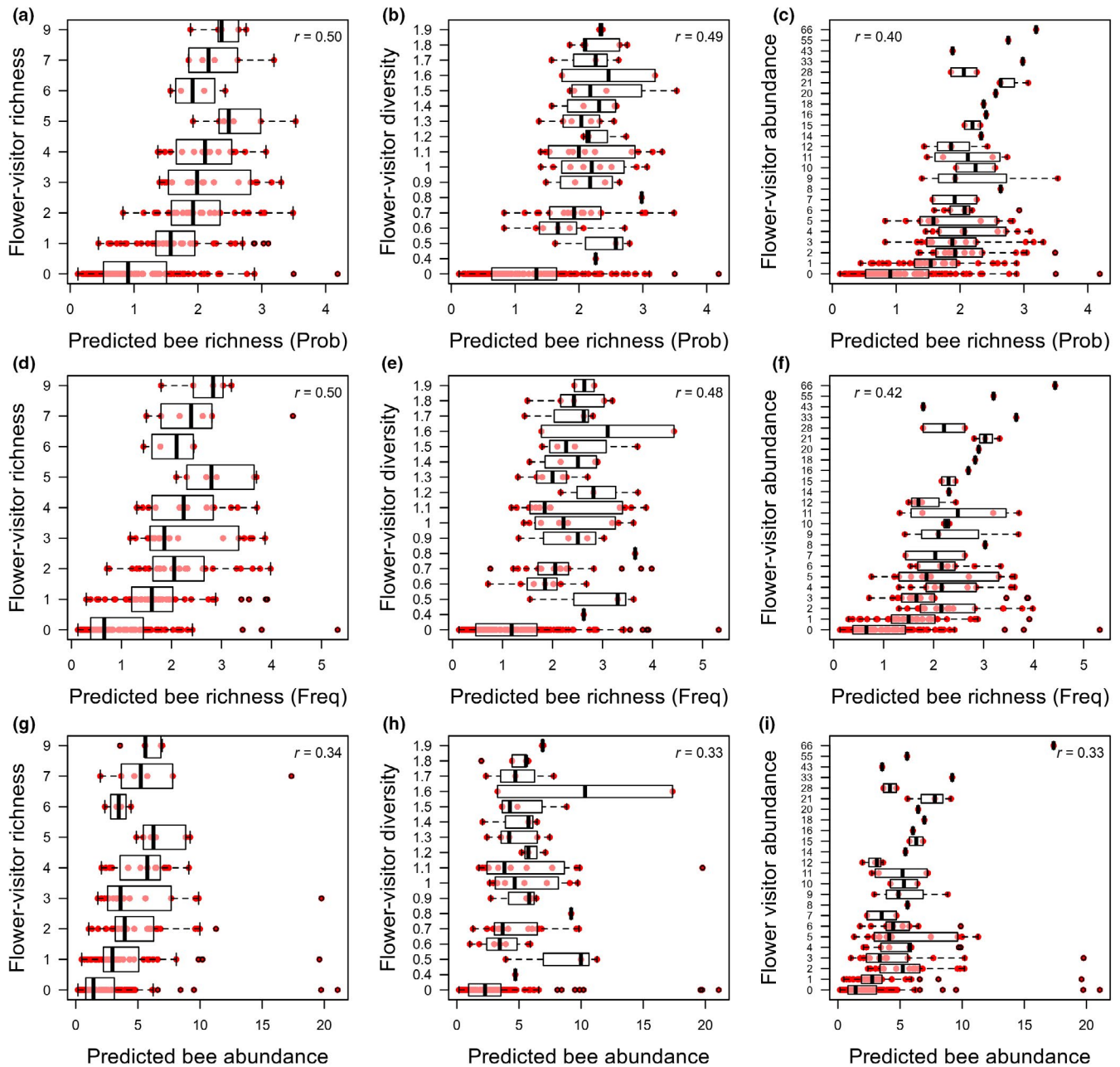
The aim of this study was to develop and test a framework for producing spatially explicit predictions of plant–pollinator networks. Despite the relatively low predictive importance of landscape level variables, there was a considerable spatial difference in the predicted

species richness of bees that visit plants belonging to Compositae and Leguminosae (Figure 4).

Approaches to modelling plant–pollinator networks can be classified according to the main strategies of species distribution modelling identified by D'Amen et al. (2017). In the 'assemble first and predict later' approach (sensu D'Amen et al., 2017), network indices are modelled as functions of environmental conditions (reviewed in Pellissier et al., 2018). The advantages of this strategy are that structural properties of entire networks are captured; and network indices have hypotheses affiliated to their drivers and relationships to ecosystem functioning. A problem with this strategy is that different species compositions can result in similar network properties (Olito & Fox, 2015), so that processes such as, for example, competition, trait-matching and neutrality can all theoretically give rise to similar degrees of modularity within ecological networks (reviewed in Dormann et al., 2017). An alternative strategy is the 'predict first assemble later' (sensu D'Amen et al., 2017), where distributions of flower visitors are modelled individually, and the resulting predictions are then aggregated to network properties. This approach is typically adopted in single plant species systems and, for example, used to predict pollinator abundances in crops (e.g. Gardner et al., 2020; Lonsdorf et al., 2009). A drawback of this approach is that because species are modelled individually, it is difficult to link predictions to community assembly processes (sensu Vellend, 2016) and meta-community ecological theory (sensu Leibold et al., 2004). In the final strategy, the 'assemble and predict together' strategy (sensu D'Amen et al., 2017), species interactions are modelled simultaneously for all pollinator species across plant species and communities. This strategy has been adopted in recent frameworks for predicting plant–pollinator interactions and allows linking network assembly to metacommunity ecological theory such as trait-based species sorting among communities (Leibold et al., 2004). For instance, Graham and Weinstein (2018) predicted plant hummingbird interactions simultaneously for hummingbird species along a gradient of elevation and outlined a strategy for how functional traits could be integrated into their modelling framework. In addition to trait-based pollinator species sorting, the role of pollinator abundances has previously been included into models of plant–pollinator interactions that identify linkage rules within networks (Bartomeus et al., 2016). Furthermore, in addition to trait-based filtering, and random (i.e. abundance-based) encounters, plant–pollinator interactions, for example, between mustards and wild bees, have been shown to depend on habitat isolation (Steffan-Dewenter & Tscharntke, 1999). By including the influence of dispersal limitation, in addition to those of trait-based species sorting and random encounters, MetaComNet extends existing 'assemble and predict together' frameworks for predicting plant–pollinator interactions.

A direct comparison of the predictive power of MetaComNet and other frameworks for predicting pairwise interactions (e.g. Pichler et al., 2020; Stock et al., 2021; Benadi et al., in press) is partly hindered by differences in study designs, and particularly that we in our models included spatial predictors in order to predict across networks of pairwise interactions (sites). For instance, Stock

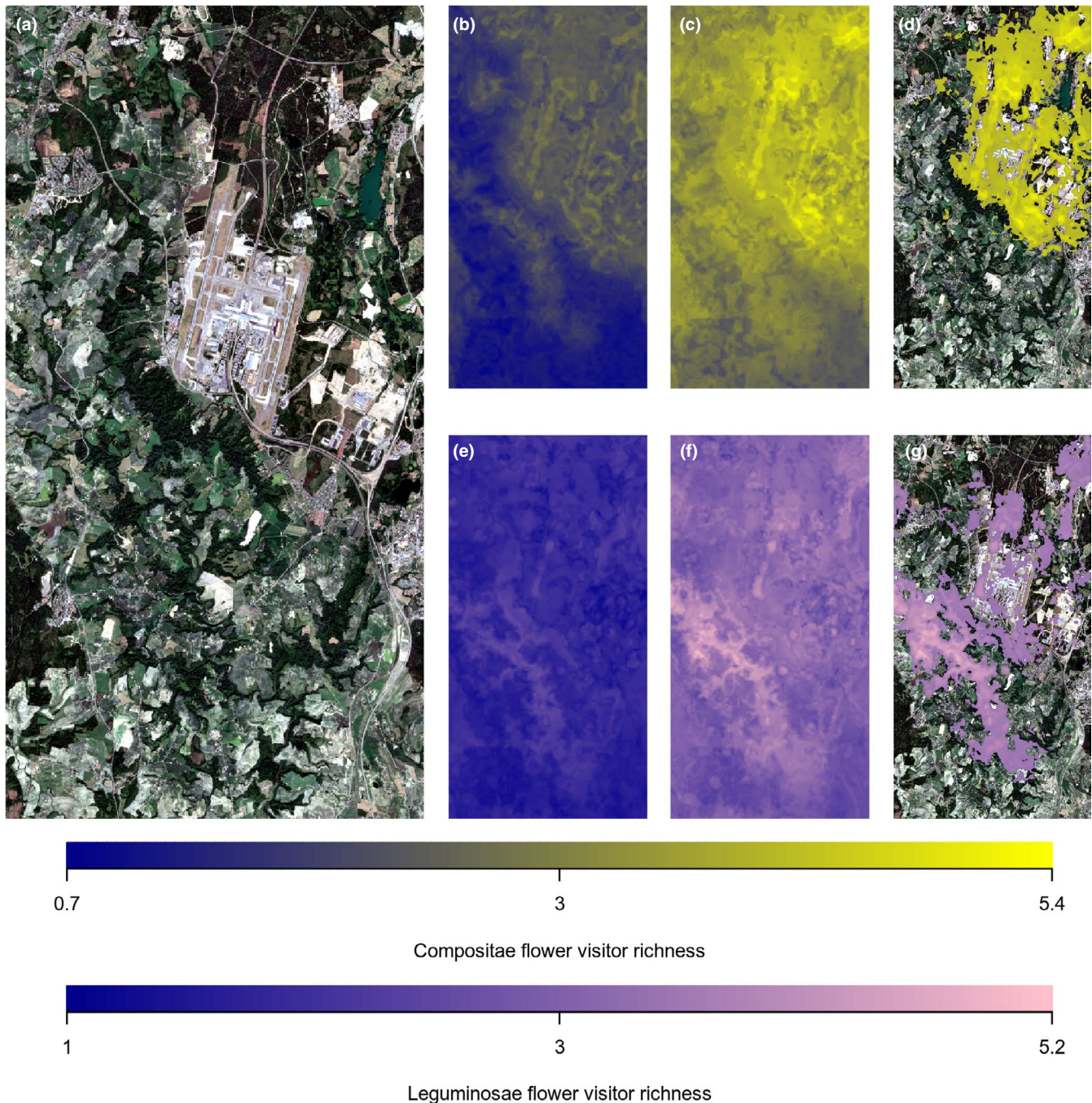




**FIGURE 3** Relationships between observed flower-visitor species richness, diversity or abundance and predictions flower-visitor species richness and abundance modelled with three random tree approaches: (i) predicted probability of interactions based on classification trees (a–c); (ii) predicted frequency of interactions from regression trees on presence/absence data (b–h); (iii) and predicted number of interactions from regression trees (g–i). The Pearson correlation coefficient between observed and predicted values is shown for each of the modelling approaches. Data points are plotted as red points together with their boxplot summary statistics

et al. (2021) devised a cross-validation strategy to assess within-network model prediction of: pairwise interactions; interactions for particular bee or plant species; or particular combinations of bees and plants. The most similar strategy to that we followed would be that of the ‘pairwise interactions’, however, in some instances particular plant species occurring in the validation data, would not have been found in the combination of sites used to train our models. Predicting across networks will therefore produce a mix of the cross-validation strategies proposed by Stock et al. (2021) and we only found marginal differences in model performances when predicting

pairwise interactions within and across sites (Figure 2) and when predicting interactions for individual species of bees (Figure S2) and plants (Figure S3). However, despite including spatial variables and cross-validating predictions into new sites which we would expect would reduce model performances, our models yielded AUC statistics for predictions of interaction occurrences that were comparable to those obtained using simulated and empirical data for within network predictions (Pichler et al., 2020; Stock et al., 2021). This may suggest that plant–pollinator trait-matching and neutral, abundance-based, processes, are likely to be the main structuring processes



**FIGURE 4** Spatial predictions of flower-visitor species richness in a representative landscape in the study area, the area surrounding Oslo airport Gardemoen (a). Predicted flower-visitor species richness to plant species in the Compositae (b–d) and Leguminosae (e–g) families depended on plant abundance, shown by predicting with plant abundances held constant at low (b, e) and high (c, f) levels. The predicted flower-visitor species richness differs spatially between the two plant taxa (d, g), illustrated by masking out areas with predicted values less than the 75th quantile of predicted values from (c) and (f) respectively. Satellite imagery from Copernicus Sentinel-2 data (2019)/processed by the Norwegian Mapping Authority

behind network assembly, although our models did identify spatial signals in plant–pollinator interactions (Figure 4).

We used detrended correspondence scores to reflect the floral associations of bees, and bee associations of plant families (Figure S2). Another option would be to use morphological and phenological traits instead (Pichler et al., 2020). A benefit of the latter approach is that novel interactions can be predicted, for example, for

invasive species, if one has information on their traits. Alternatively, or additionally, to traits one can use information on phylogenetic relatedness when modelling interactions (e.g. Benadi et al., in press; Stock et al., 2021) because plant–pollinator associations are often somewhat phylogenetically conserved (e.g. Wood et al., 2021). If information on host plants of some pollinator species is not known, and one wishes to apply the approach using DCA scores adopted in

this study, a potential would be to use the average DCA score values from the closest relatives for which one has information, or assume that the species will be restricted to the same host-plant families as in its native range (Vaudo et al., 2020). Using floral preference traits, inferred from, for example, DCA scores, should therefore not necessarily hinder the prediction of novel interactions. However, a limitation of using floral association (e.g. DCA) scores is that they do not provide tests of how specific plant–pollinator trait-combinations influence pairwise interaction probabilities. Thus if an aim is to identify linkage rules in ecological networks then, hypothesised, functional traits should be used in the MetaComNet framework instead of floral preference scores.

In our models, we treated absences of interactions with the same degree of confidence as presences. However, because one is unlikely to detect all interactions when sampling plant–pollinator interaction networks (Chacoff et al., 2012), data on multiple interaction networks are likely to include many false absences, resulting in models underestimating pairwise interactions. A sophisticated approach to handling false absences would be to incorporate species-specific detectability's into the model to weight the absence values in the data (e.g. Graham & Weinstein, 2018). However, obtaining enough data to estimate detectability requires extensive, and repeated, surveys with the cost of a decreased sample size of environmental conditions (study sites) which may reduce the ability to parameterise the effects of environmental filtering on network assembly. An alternative is to remove plausible but unobserved pairwise interactions from the training data and thereby retain only the most credible absence values (Liu et al., 2015). However, this approach requires some degree of subjectivity in terms of which traits are included when estimating and setting thresholds for interaction credibility. We are also unsure how one would weight the credibility of an interaction in a spatial setting, where some pairwise interactions might be considered credible, but were unobserved because of environmental conditions at larger spatial scales than within the network. Removing such plausible but unobserved pairwise interactions from the data would likely result in models underestimating, for example, the landscape context effects on interaction occurrences. A final approach is to remove all absence values and use the frequency of trait-pair interactions (e.g. flower-visitor trait by plant trait, or flower-visitor trait by environmental filter combinations) to estimate the likelihood of species interactions within networks (e.g. Bartomeus et al., 2016). However, the frequency of recorded interactions per species (and trait group) is vulnerable to the same biases as the absence values. Moreover, the same biases are likely to occur in both training and validation data, unless these were sampled independently using different protocols. It may therefore be difficult to detect such biases, in terms of whether the model over- or under-predicts interactions in the validation data. Still, in our models, models using classification trees underestimated occurrences of pairwise interactions (Figure 2a, slope = 1.21) more frequently than models from regression trees (Figure 2d, slope = 1.09), suggesting that the latter approach may be less vulnerable to biases that would lead to underestimations. If the detectability of interactions does not depend

on the predictor variables used in the MetaComNet models, and occurrences of pairwise interactions vary according to the predictor variables, then the occurrence of pairwise interactions should be proportional to the predicted probabilities of occurrences (Elkan & Noto, 2008). We therefore suggest that absence values, and the information they contain, are included in models aimed at predicting plant–pollinator interactions.

Despite a reasonable fit to the validation data, a considerable amount of variation in pairwise bee–plant interactions was left unaccounted for by our models (Figure 2a,d,g). While some of this unexplained variation is likely attributable to random error, it seems fair to assume that a large fraction of it was due to unmeasured predictor variables. Negative biotic interactions, such as interspecific competition which can suppress bee–flower visitations (Wignall et al., 2020), were not accounted for in our models. However, we would expect that negative biotic interactions would result in our models overestimating local interaction probabilities, which was not the case. By contrast, our models tended to underestimate the occurrence of pairwise interactions (i.e. Figure 2a,d, regression slopes >1) which may suggest that there were elements related to habitat conditions, such as habitat continuity (Morandin & Kremen, 2013), that could result in higher-than-expected occurrences of bees. A potential solution for future implementations of MetaComNet would be to use annually updatable land cover maps (e.g. Venter & Sydenham, 2021) to estimate the continuity of habitat patches. It could also be that bee species were found more frequently than predicted because the distance to source population variable was too conservative, that is, because species occurrences in the GBIF record did not reflect the actual distribution of species. Ideally one would have more information on the location of potential source populations, than from potentially scattered, and potentially biased, species observation records. Another important contributor to the unexplained variation in our models was the small number of study sites available (15 for building models, 1 for validating). In order to accurately estimate the contributions of landscape level variables, we therefore suspect that increasing the sample size of study sites will enable stronger predictions of bee–flower interactions than what was possible given the limited data available to us. Despite these potential limitations, our models did produce prediction maps that correspond well with how we would expect floral visitation patterns to be distributed (Figure 4), that is, with Compositae-visiting solitary bees being concentrated near sandy soils, and Leguminosae-visiting bumble bees being less concentrated on sandy soils (Figure S2).

## 5 | CONCLUSIONS

Spatial models of pollinator diversity should theoretically allow for identifying areas (a) where plant populations would benefit from pollinator enhancement schemes, and (b) where plant populations are likely to have a higher genetic diversity because of high levels of pollination. We believe that the modelling framework presented

here provides a promising avenue for producing spatially explicit predictions of plant–pollinator interactions. If possible, future studies adopting the framework should attempt to assess the degree of pollen limitation within plant populations to empirically test if predicting interaction partner diversity allows to also predict the degree of pollen limitation experienced by plant species.

## ACKNOWLEDGEMENTS

We thank the Norwegian public roads administration for facilitating our field work along road sides. The associate editor, and Dr Gita Benadi and one anonymous reviewer provided thoughtful and constructive reviews of previous versions of the manuscript. This research was funded by The Research Council of Norway [Proj. no. 302692] as part of the project 'MetaComNet: Linking meta-community dynamics to the structure of ecological networks and ecosystem functioning', awarded to Dr Sydenham. A CC BY or equivalent licence is applied to any Author Accepted Manuscript (AAM) version arising from this submission, in accordance with the grant's open access conditions.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## AUTHORS' CONTRIBUTIONS

M.A.K.S. conceived the ideas and designed methodology; M.A.K.S., Z.S.V., C.R., A.B.S., D.I.J.S. and K.-A.H. collected and collated the data; M.A.K.S., Z.S.V. and T.R. analysed the data; M.A.K.S. led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.









## PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/2041-210X.13762>.

## DATA AVAILABILITY STATEMENT

R code for reproducing analyses can be found at <https://doi.org/10.5281/zenodo.5644742> (Sydenham et al., 2021a). Data for running the R code are available through the Dryad Digital Repository <https://doi.org/10.5061/dryad.n02v6wwzn> (Sydenham et al., 2021b).

## ORCID

Markus A. K. Sydenham  <https://orcid.org/0000-0002-7711-2399>  
 Zander S. Venter  <https://orcid.org/0000-0003-2638-7162>  
 Trond Reitan  <https://orcid.org/0000-0003-0793-8731>  
 Claus Rasmussen  <https://orcid.org/0000-0003-1529-6548>  
 Astrid B. Skringo  <https://orcid.org/0000-0003-0065-3434>  
 Stein Joar Hegland  <https://orcid.org/0000-0001-7039-5966>  
 Yoko L. Dupont  <https://orcid.org/0000-0002-8811-2773>  
 Anders Nielsen  <https://orcid.org/0000-0002-3294-6234>  
 Joseph Chipperfield  <https://orcid.org/0000-0003-3314-0846>  
 Graciela M. Rusch  <https://orcid.org/0000-0003-3769-8345>

## REFERENCES

- Antoine, C. M., & Forrest, J. R. (2021). Nesting habitat of ground-nesting bees: A review. *Ecological Entomology*, 46(2), 143–159. <https://doi.org/10.1111/een.12986>
- Bartomeus, I., Gravel, D., Tyljanakis, J. M., Aizen, M. A., Dickie, I. A., & Bernard-Verdier, M. (2016). A common framework for identifying linkage rules across different types of interactions. *Functional Ecology*, 30(12), 1894–1903. <https://doi.org/10.1111/1365-2435.12666>
- Barton, K. (2018). *MuMIn: Multi-model inference*. R package version 1.42.1. Retrieved from <https://CRAN.R-project.org/package=MUMIn>
- Benadi, G., Dormann, C., Fründ, J., Stephan, R., & Vázquez, D. P. (in press). Quantitative prediction of interactions in bipartite networks based on traits, abundances, and phylogeny. *The American Naturalist*. <https://doi.org/10.1086/714420>
- Benadi, G., & Pauw, A. (2018). Frequency dependence of pollinator visitation rates suggests that pollination niches can allow plant species coexistence. *Journal of Ecology*, 106(5), 1892–1901. <https://doi.org/10.1111/1365-2745.13025>
- Bennett, J. M., Steets, J. A., Burns, J. H., Burkle, L. A., Vamosi, J. C., Wolowski, M., Arceo-Gómez, G., Burd, M., Durka, W., Ellis, A. G., Freitas, L., Li, J., Rodger, J. G., Ştefan, V., Xia, J., Knight, T. M., & Ashman, T.-L. (2020). Land use and pollinator dependency drives global patterns of pollen limitation in the Anthropocene. *Nature Communications*, 11(1), 1–6. <https://doi.org/10.1038/s41467-020-17751-y>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Carstensen, D. W., Sabatino, M., Trøjelsgaard, K., & Morellato, L. P. C. (2014). Beta diversity of plant–pollinator networks and the spatial turnover of pairwise interactions. *PLoS One*, 9(11), e112903. <https://doi.org/10.1371/journal.pone.0112903>
- Chacoff, N. P., Vázquez, D. P., Lomáscolo, S. B., Stevani, E. L., Dorado, J., & Padrón, B. (2012). Evaluating sampling completeness in a desert plant–pollinator network. *Journal of Animal Ecology*, 81(1), 190–200. <https://doi.org/10.1111/j.1365-2656.2011.01883.x>
- Cornell, H. V., & Harrison, S. P. (2014). What are species pools and when are they important? *Annual Review of Ecology, Evolution, and Systematics*, 45, 45–67. <https://doi.org/10.1146/annurev-ecolsys-120213-091759>
- D'Amen, M., Rahbek, C., Zimmermann, N. E., & Guisan, A. (2017). Spatial predictions at the community level: From current approaches to future frameworks. *Biological Reviews*, 92(1), 169–187. <https://doi.org/10.1111/brv.12222>
- Dormann, C. F., Fründ, J., & Schaefer, H. M. (2017). Identifying causes of patterns in ecological networks: Opportunities and limitations. *Annual Review of Ecology, Evolution, and Systematics*, 48, 559–584. <https://doi.org/10.1146/annurev-ecolsys-110316-022928>
- Elkan, C., & Noto, K. (2008). Learning classifiers from only positive and unlabeled data. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 213–220. <https://doi.org/10.1145/1401890.1401920>
- Fowler, R. E., Rotheray, E. L., & Goulson, D. (2016). Floral abundance and resource quality influence pollinator choice. *Insect Conservation and Diversity*, 9(6), 481–494. <https://doi.org/10.1111/icad.12197>
- Gardner, E., Breeze, T. D., Clough, Y., Smith, H. G., Baldock, K. C. R., Campbell, A., Garratt, M. P. D., Gillespie, M. A. K., Kunin, W. E., Mc Kerchar, M., Memmott, J., Potts, S. G., Senapathi, D., Stone, G. N., Wäckers, F., Westbury, D. B., Wilby, A., & Oliver, T. H. (2020). Reliably predicting pollinator abundance: Challenges of calibrating process-based ecological models. *Methods in Ecology and Evolution*, 11(12), 1673–1689. <https://doi.org/10.1111/2041-210X.13483>
- GBIF. (2021). Filtered export of GBIF occurrence data. <https://doi.org/10.15468/dd.tv79q4>
- Geological Survey of Norway. (2011). *Løsmasser WMS*. Retrieved from <https://kartkatalog.geonorge.no/metadata/loesmasser/3de4ddf6-d6b8-4398-8222-f5c47791a757>

- Gomez, J. M., Abdelaziz, M., Lorite, J., Jesús Muñoz-Pajares, A., & Perfectti, F. (2010). Changes in pollinator fauna cause spatial variation in pollen limitation. *Journal of Ecology*, 98(5), 1243–1252. <https://doi.org/10.1111/j.1365-2745.2010.01691.x>
- Graham, C. H., & Weinstein, B. G. (2018). Towards a predictive model of species interaction beta diversity. *Ecology Letters*, 21(9), 1299–1310. <https://doi.org/10.1111/ele.13084>
- Hagen, M., Kissling, W. D., Rasmussen, C., De Aguiar, M. A. M., Brown, L. E., Carstensen, D. W., Alves-Dos-Santos, I., Dupont, Y. L., Edwards, F. K., Genini, J., Guimarães, P. R., Jenkins, G. B., Jordano, P., Kaiser-Bunbury, C. N., Ledger, M. E., Maia, K. P., Marquitti, F. M. D., McLaughlin, Ó., Morellato, L. P. C., ... Olesen, J. M. (2012). 2 - Biodiversity, species interactions and ecological networks in a fragmented world. In U. Jacob & G. Woodward (Eds.), *Advances in ecological research* (Vol. 46, pp. 89–210). Academic Press. <https://doi.org/10.1016/B978-0-12-396992-7.00002-2>
- Heneberg, P., Bogusch, P., & Řehounek, J. (2013). Sandpits provide critical refuge for bees and wasps (Hymenoptera: Apocrita). *Journal of Insect Conservation*, 17(3), 473–490. <https://doi.org/10.1007/s10841-012-9529-5>
- Hoiss, B., Krauss, J., Potts, S. G., Roberts, S., & Steffan-Dewenter, I. (2012). Altitude acts as an environmental filter on phylogenetic composition, traits and diversity in bee communities. *Proceedings of the Royal Society B: Biological Sciences*, 279(1746), 4447–4456. <https://doi.org/10.1098/rspb.2012.1581>
- Krishna, A., Guimaraes, P. R. Jr, Jordano, P., & Bascompte, J. (2008). A neutral-niche theory of nestedness in mutualistic networks. *Oikos*, 117(11), 1609–1618. <https://doi.org/10.1111/j.1600-0706.2008.16540.x>
- Kuhn, M. with contributions from Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., the R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., & Hunt, T. (2018). caret: Classification and regression training. R package version 6.0-81. Retrieved from <https://cran.r-project.org/package=caret>
- Larue, A. A. C., Raguso, R. A., & Junker, R. R. (2016). Experimental manipulation of floral scent bouquets restructures flower–visitor interactions in the field. *Journal of Animal Ecology*, 85(2), 396–408. <https://doi.org/10.1111/1365-2656.12441>
- Leibold, M. A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F., Holt, R. D., Shurin, J. B., Law, R., Tilman, D., Loreau, M., & Gonzalez, A. (2004). The metacommunity concept: A framework for multi-scale community ecology. *Ecology Letters*, 7(7), 601–613. <https://doi.org/10.1111/j.1461-0248.2004.00608.x>
- Liu, H., Sun, J., Guan, J., Zheng, J., & Zhou, S. (2015). Improving compound-protein interaction prediction by building up highly credible negative samples. *Bioinformatics*, 31, i221–i229. <https://doi.org/10.1093/bioinformatics/btv256>
- Lonsdorf, E., Kremen, C., Ricketts, T., Winfree, R., Williams, N., & Greenleaf, S. (2009). Modelling pollination services across agricultural landscapes. *Annals of Botany*, 103(9), 1589–1600. <https://doi.org/10.1093/aob/mcp069>
- Morandin, L. A., & Kremen, C. (2013). Hedgerow restoration promotes pollinator populations and exports native bees to adjacent fields. *Ecological Applications*, 23(4), 829–839. <https://doi.org/10.1890/12-1051.1>
- Norwegian Mapping Authority. (2016). Høyde DTM 50 (UTM33) WMS. Retrieved from <https://kartkatalog.geonorge.no/metadata/dtm-50/e25d0104-0858-4d06-bba8-d154514c11d2>
- Oksanen, J., Guillaume Blanchet, F., Friendly, M., Roeland, K., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2018). *vegan: Community ecology package*. R package version 2.5-3. Retrieved from <https://cran.r-project.org/package=vegan>
- Olesen, J. M., Bascompte, J., Dupont, Y. L., Elberling, H., Rasmussen, C., & Jordano, P. (2011). Missing and forbidden links in mutualistic networks. *Proceedings of the Royal Society B: Biological Sciences*, 278(1706), 725–732. <https://doi.org/10.1098/rspb.2010.1371>
- Olito, C., & Fox, J. W. (2015). Species traits and abundances predict metrics of plant–pollinator network structure, but not pairwise interactions. *Oikos*, 124(4), 428–436. <https://doi.org/10.1111/oik.01439>
- Ollerton, J. (2017). Pollinator diversity: Distribution, ecological function, and conservation. *Annual Review of Ecology, Evolution, and Systematics*, 48, 353–376. <https://doi.org/10.1146/annurev-ecolsys-110316-022919>
- Ollerton, J., Winfree, R., & Tarrant, S. (2011). How many flowering plants are pollinated by animals? *Oikos*, 120(3), 321–326. <https://doi.org/10.1111/j.1600-0706.2010.18644.x>
- Pellissier, L., Albouy, C., Bascompte, J., Farwig, N., Graham, C., Loreau, M., Maglianesi, M. A., Melián, C. J., Pitteloud, C., Roslin, T., & Rohr, R. (2018). Comparing species interaction networks along environmental gradients. *Biological Reviews*, 93(2), 785–800. <https://doi.org/10.1111/brv.12366>
- Pichler, M., Boreux, V., Klein, A. M., Schleuning, M., & Hartig, F. (2020). Machine learning algorithms to infer trait-matching and predict species interactions in ecological networks. *Methods in Ecology and Evolution*, 11(2), 281–293. <https://doi.org/10.1111/2041-210X.13329>
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Rasmussen, C., Dupont, Y. L., Madsen, H. B., Bogusch, P., Goulson, D., Herbertsson, L., Maia, K. P., Nielsen, A., Olesen, J. M., Potts, S. G., Roberts, S. P. M., Sydenham, M. A. K., & Kryger, P. (2021). Evaluating competition for forage plants between honey bees and wild bees in Denmark. *PLoS One*, 16(4), e0250056. <https://doi.org/10.1371/journal.pone.0250056>
- Rivest, S., & Forrest, J. R. (2020). Defence compounds in pollen: Why do they occur and how do they affect the ecology and evolution of bees? *New Phytologist*, 225(3), 1053–1064. <https://doi.org/10.1111/nph.16230>
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12(1), 1–8. <https://doi.org/10.1186/1471-2105-12-77>
- Rowe, L., Gibson, D., Bahlai, C. A., Gibbs, J., Landis, D. A., & Isaacs, R. (2020). Flower traits associated with the visitation patterns of bees. *Oecologia*, 193(2), 511–522. <https://doi.org/10.1007/s00442-020-04674-0>
- Skoog, D. I. J. (2018). *The influence of nesting resources on bee-flower interactions, revealed through functional traits, network structure and geology*. Master's theses (MINA). Norwegian University of Life Sciences. <http://hdl.handle.net/11250/2566841>
- Stavert, J. R., Bartomeus, I., Beggs, J. R., Gaskett, A. C., & Pattemore, D. E. (2019). Plant species dominance increases pollination complementarity and plant reproductive function. *Ecology*, 100(9), e02749. <https://doi.org/10.1002/ecy.2749>
- Steffan-Dewenter, I., Münzenberg, U., Bürger, C., Thies, C., & Tschardtke, T. (2002). Scale-dependent effects of landscape context on three pollinator guilds. *Ecology*, 83(5), 1421–1432. [https://doi.org/10.1890/0012-9658\(2002\)083\[1421:SDEOLC\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2002)083[1421:SDEOLC]2.0.CO;2)
- Steffan-Dewenter, I., & Tschardtke, T. (1999). Effects of habitat isolation on pollinator communities and seed set. *Oecologia*, 121(3), 432–440. <https://doi.org/10.1007/s004420050949>
- Stock, M., Piot, N., Vanbesien, S., Meys, J., Smagghe, G., & De Baets, B. (2021). Pairwise learning for predicting pollination interactions based on traits and phylogeny. *Ecological Modelling*, 451, 109508. <https://doi.org/10.1016/j.ecolmodel.2021.109508>
- Sydenham, M. A. K., Venter, Z. S., Reitan, T., Rasmussen, C., Skringo, A. B., Skoog, D. J., Hanevik, K.-A., Hegland, S. J., Dupont, Y. L., Nielsen, A., Chipperfield, J., & Rusch, G. M. (2021a). R code for: MetaComNet: A random forest-based framework for making spatial prediction of plant–pollinator interactions. *Methods in Ecology and Evolution*, <https://doi.org/10.5281/zenodo.5644742>

- Sydenham, M. A. K., Venter, Z. S., Reitan, T., Rasmussen, C., Skringo, A. B., Skoog, D. J., Hanevik, K.-A., Hegland, S. J., Dupont, Y. L., Nielsen, A., Chipperfield, J., & Rusch, G. M. (2021b). Data from: MetaComNet: A random forest-based framework for making spatial prediction of plant-pollinator interactions. *Dryad Digital Repository*, <https://doi.org/10.5061/dryad.n02v6wwzn>
- Thomann, M., Imbert, E., Devaux, C., & Cheptou, P. O. (2013). Flowering plants under global pollinator decline. *Trends in Plant Science*, *18*(7), 353–359. <https://doi.org/10.1016/j.tplants.2013.04.002>
- Totland, Ø. (1994). Influence of climate, time of day and season, and flower density on insect flower visitation in alpine Norway. *Arctic and Alpine Research*, *26*(1), 66–71. <https://doi.org/10.2307/1551879>
- Trøjelsgaard, K., Jordano, P., Carstensen, D. W., & Olesen, J. M. (2015). Geographical variation in mutualistic networks: Similarity, turnover and partner fidelity. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1802), 20142925. <https://doi.org/10.1098/rspb.2014.2925>
- Tylianakis, J. M., & Morris, R. J. (2017). Ecological networks across environmental gradients. *Annual Review of Ecology, Evolution, and Systematics*, *48*(1), 25–48. <https://doi.org/10.1146/annurev-ecolsys-110316-022821>
- Vaudo, A. D., Biddinger, D. J., Sickel, W., Keller, A., & López-Urbe, M. M. (2020). Introduced bees (*Osmia cornifrons*) collect pollen from both coevolved and novel host-plant species within their family-level phylogenetic preferences. *Royal Society Open Science*, *7*(7), 200225. <https://doi.org/10.1098/rsos.200225>
- Vázquez, D. P., Melián, C. J., Williams, N. M., Blüthgen, N., Krasnov, B. R., & Poulin, R. (2007). Species abundance and asymmetric interaction strength in ecological networks. *Oikos*, *116*(7), 1120–1127. <https://doi.org/10.1111/j.0030-1299.2007.15828.x>
- Vellend, M. (2016). *The theory of ecological communities*. Monographs in Population Biology (Vol. 57, pp. 1–248). Princeton University Press.
- Venter, Z. S., & Sydenham, M. A. K. (2021). Continental-scale land cover mapping at 10 m resolution over Europe (ELC10). *Remote Sensing*, *13*(12), 2301. <https://doi.org/10.3390/rs13122301>
- Westrich, P. (1996). Habitat requirements of central European bees and the problems of partial habitats. In A. Matheson, S. L. Buchmann, C. O'toole, P. Westrich, & I. H. Williams (Eds.), *The conservation of bees*. Linnean Society symposium series 18 (pp. 1–16). Academic Press Limited.
- Wignall, V. R., Brolly, M., Uthoff, C., Norton, K. E., Chipperfield, H. M., Balfour, N. J., & Ratnieks, F. L. (2020). Exploitative competition and displacement mediated by eusocial bees: Experimental evidence in a wild pollinator community. *Behavioral Ecology and Sociobiology*, *74*(12), 1–15. <https://doi.org/10.1007/s00265-020-02924-y>
- Wood, T. J., Ghisbain, G., Rasmont, P., Kleijn, D., Raemakers, I., Praz, C., Killewald, M., Gibbs, J., Bobiwash, K., Boustani, M., Martinet, B., & Michez, D. (2021). Global patterns in bumble bee pollen collection show phylogenetic conservation of diet. *Journal of Animal Ecology*, *90*(10), 2421–2430. <https://doi.org/10.1111/1365-2656.13553>
- Wright, M. N., & Ziegler, A. (2017). ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, *77*(1), 1–17. <https://doi.org/10.18637/jss.v077.i01>

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Sydenham, M. A. K., Venter, Z. S., Reitan, T., Rasmussen, C., Skringo, A. B., Skoog, D. I. J., Hanevik, K.-A., Hegland, S. J., Dupont, Y. L., Nielsen, A., Chipperfield, J., & Rusch, G. M. (2021). MetaComNet: A random forest-based framework for making spatial predictions of plant–pollinator interactions. *Methods in Ecology and Evolution*, 00, 1–14. <https://doi.org/10.1111/2041-210X.13762>