Assessing the flower visitation to soybean (*Glycine max*) and different sampling methods in an intensive agricultural system of the Argentinian Pampas

by

Ingvild Fonn Asmervik

Master Thesis

Department of Biosciences

Faculty of Mathematics and Natural Sciences

UNIVERSITY OF OSLO

December 2017

Assessing the flower visitation to soybean (*Glycine max*) and different sampling methods in an intensive agricultural system of the Argentinian Pampas

2017

Assessing the flower visitation to soybean (*Glycine max*) and different sampling methods in an intensive agricultural system of the Argentinian Pampas

Ingvild Fonn Asmervik

http://www.duo.uio.no/

Print: Reprosentralen, Universitetet i Oslo

This project was carried out in collaboration with the Faculty of Agronomy at the University of Buenos Aires (UBA).

ACKNOWLEDGEMENTS

In the process of making this thesis, I have received outstanding help and massive support from many people. First of all, I want to thank my brilliant supervisors Anders Nielsen (UiO), Anne Krag Brysting (UiO), Trond Reitan (UiO) and Mariano Devoto (UBA) – you have made up the most excellent supporting team. Anders, your insight and interest in so many aspects of pollination ecology, combined with your never-ending motivational words, have been invaluable. I'm amazed by all the time and patience you've devoted to work through countless thesis drafts of varying quality, and thankful for your intelligent, and sometimes witty, feedbacks. Anne, your warming character and infectious laugh have lifted my spirits again and again, and your ingenious feedbacks, including 'småpirk', have been received with great gratitude. Trond, your endless statistical knowledge never seizes to amaze me, and your enthusiastic involvement has been essential for the outcome. Mariano, your clever inputs and wise feedbacks along the way, also after my return from Argentina, have greatly contributed to improve the thesis.

I also want to express huge thanks to all the people that made my stay in Argentina such an enjoyable experience – *Gracias a Mariano, Pablo y Toño por hacerme sentir tan bienvenida en la FAUBA desde el primer día. Mil gracias a Vicky y Cris por su ayuda, y también a Toño, Aye, Ana, Ade y Marcos, por su buena compañía durante el trabajo de campo – nunca voy a olvidar sus bombachas gauchas divinas o cómo les encantó el queso marrón. Vicky, mi querida Pippi, quiero agradecerte a vos especialmente por abrir tu casa y por compartir tantas cosas conmigo. Tõno, una gracia adicional para vos también por compartir tus datos, por tu ayuda después de mi regreso y por siempre estar de buen humor (¡chido!). También quiero expresar muchas gracias a Mati, Maisa y Isis por hospedarme durante las primeras semanas en Buenos Aires, y por su linda compañía durante navidad.*

My beautiful friends and family, I can't even imagine how hard this process would have been if it wasn't for you. Lisa, your company and inputs during the last two years have been appreciated more than you know. Thank you for showing a genuine interest, and for reading the entire thesis before submission. Julie and Silje, my role models, thank you for making the summer at Blindern such a happy experience, and,

to Magnus and Gunvor as well, for the best company during autumn. Malin, your regular pop-ins to the study room and EndNote-help have been much appreciated. Thanks to Hilde and Inger for our friendship and your support – you are exceptional. Catharina, thank you for helping with R and tables, and for your encouraging words. Many thanks go to everybody else who have made the days at Blindern fun and gratifying. I also want to thank Julie, Anette and my brother Jonas for your support and understanding. Thanks to Pappa for your admiration and encouragement. Mamma, thank you for being the best listener and the sweetest mum. The writing skills that both of you possess have inspired me to always try to perfect my own. Finally, special thanks go to my biggest support and favourite person, Knut.

ABSTRACT

The agricultural sector is facing the challenge to provide food for a growing human population, while at the same time reducing its global impact on natural ecosystems and the services they provide. Drivers that have been identified to cause declines in pollinator populations worldwide include climate change and habitat loss. Among the most widespread and commercially important crops is the soybean (*Glycine max*). The demand for soybean is increasing, and field expansions are destructing forests and grasslands, mainly in South America. Conserving these habitats should be facilitated by finding ways to augment soybean yields. While some studies indicate that this is achieved by increasing pollination by bees, few have assessed the actual pollinator activity, or the factors that may affect pollinator populations, in modern soybean fields. Furthermore, when assessing pollinator abundance and declines, it is important to use methods that are feasible for the specific objectives to be addressed.

With this thesis, I aimed to assess 1) how the frequency of potential soybean pollinators is affected by environmental conditions, 2) the feasibility of different sampling methods and 3) the visitation frequency to soybean flowers in a highly intensive agricultural system. Fieldwork was carried out by collecting flower visitors in nine monoculture soybean fields, equally distributed among three blocks forming a latitudinal gradient of ~200 km, in the Argentinian Pampas. I recorded ambient temperature, relative humidity and wind speed to assess the effects of environmental conditions. The feasibility of different sampling methods was assessed by using both transect walks and plot samplings when collecting the data. To assess the visitation frequency, I estimated the probability for a flower in the focal system to receive at least one visit during its life span.

I found that environmental conditions had a significant impact on the flower visitor frequency. Plot samplings were shown to be suitable for studies that aim to estimate visitor or visitation frequencies, while transect walks appeared to be the best method when the aim is to obtain as many observations as possible. Finally, the visitation frequency to soybean was revealed to be extremely low, as the flowers were estimated to have less than a 6% chance to receive a visit while being open.

CONTENTS

1 INTRODUCTION

Achieving food security for a growing human population is a current concern; at the same time, modern agriculture is a dominant force behind many threats to biodiversity. Habitat loss, pesticide use and climate change cause declines in both managed and wild populations of pollinators that provide a vital ecosystem service to agriculture itself. Investigating the status of pollinator populations in modern agricultural systems, and how they may respond to environmental factors associated with climate change, is therefore important. Furthermore, to address questions related to pollinator abundance and declines, the sampling methodology needs to be tailored to yield data of the best possible quality, depending on the specific questions. These considerations are the motivators for this thesis.

1.1 Agriculture's footprint

Since *Homo sapiens* originated about 150,000 years ago (Clark et al., 2003; McDougall et al., 2005), humankind has spread out of Africa and evolved to become the most influential species on Earth. Today, the globe is inhabiting more than 7.5 billion people that have colonized almost every territory of its surface (UN, 2017). This enormous population is still rapidly increasing (UN, 2017), and has a tremendous impact on the environment; climate change, habitat degradation, and introductions of alien species are just a few examples of anthropogenic drivers that are currently threatening global biodiversity. At the same time, many raise the question of how food security can be achieved for all these people (e.g. Godfray et al., 2010). While the United Nations (UN) (2017) predict the human population to exceed 11 billion by 2100, more than a tenth of the current world population is already suffering from undernourishment (FAO, 2016). Tilman (1999) addressed the possibility of repeating the doubling of the world food production[1] that occurred during 1961-1996, an achievement made possible by agricultural advancements. Unfortunately, increased pesticide use, the development of artificially fertilized monocultures, and expansions of cultivated land areas have collectively contributed to a significant simplification and homogenization of the world's landscapes. This

[1] Includes cereals, coarse grains and root crops.

global degradation of ecosystems involves the loss of multiple services that provide substantial benefits to humans and, paradoxically, to agriculture itself (Foley et al., 2011; Palmer et al., 2004; Tilman, 1999). Natural areas lost to cultivation are, for instance, home to hundreds of thousands of animal species that instinctively visit flowers (Nabhan & Buchmann, 1997), seeking resources (e.g. nectar and pollen) or shelter (Willmer, 2011). As this behaviour usually entails transporting pollen between flowers, they are essential contributors to ensure the plants' reproduction. Out of 352,000 wild flowering plants worldwide, Ollerton et al. (2011) estimated that 87.5% are pollinated by animals. Comparatively, about 70% of the crops that are directly consumed by humans benefit from animal pollination (Klein et al., 2007). While keeping domesticated insects (mostly honeybees) to enhance crop yield is common among farmers (Potts et al., 2010b), the efficient and free of charge service provided by wild pollinators is often ignored or underestimated (e.g. Breeze et al., 2011; Garibaldi et al., 2013). Simultaneously, there have been reports of declines in both managed honeybees (e.g. National Research Council, 2007; Potts et al., 2010b; vanEngelsdorp et al., 2008) and wild pollinators (e.g. Freitas et al., 2009; Goulson et al., 2008; Keil et al., 2011). Biesmeijer et al. (2006) showed corresponding declines in bee species richness and insect-pollinated wild plants in Britain and the Netherlands. Several possible drivers for the global pollinator declines have been thoroughly reviewed by Potts et al. (2010a) and Goulson et al. (2015). In addition to habitat loss, they identified climate change to be among the most important drivers.

Climate change has already been shown to affect butterfly distributions in Europe (Hickling et al., 2006; Parmesan et al., 1999) and bumblebee distributions across Europe and North America (Kerr et al., 2015), and is predicted to have even more severe impacts on pollinators in the future (e.g. on bees: Dormann et al., 2008; and butterflies: Settele et al., 2008). Climate change may also have indirect effects on plant-pollinator relationships such as causing temporal or spatial mismatches (Hegland et al., 2009; Schweiger et al., 2008). Habitat loss is however the factor considered to have the strongest impact on bee declines (Brown & Paxton, 2009). The world's croplands have steadily expanded since the 1960s (FAO, 2016), especially in the tropics (Foley et al., 2011). Today, about 38% of Earth's terrestrial surface is used for agriculture (Foley et al., 2011). In 2014, the harvested areas of wheat (*Triticum aestivum* L.), maize (*Zea mays* L.), rice (*Oryza sativa* L. & *O. glaberrima*

Steud.) and soybean (*Glycine max* (L.) Merr.) were more than 100 million ha each, and equalled a total of 685 million ha (FAO, 2016). Wheat, maize and rice belong to the Poaceae family and are wind-pollinated. This leaves the soybean as the most widespread non-cereal crop, presumably depending on insect pollinators for maximum yield.

1.1.1 Soybean industry

Soybean is currently cultivated in 105 countries spread out on all continents except Antarctica (FAO, 2016). Although the Americas are the main providers of soybeans today, the plant was domesticated in north-eastern China around 3000 years ago (Hymowitz, 1970). It was then grown for centuries in the Orient, mainly for the seeds that provided the communities with food, fertilizer and animal feed (Probst & Judd, 1973). In the Americas, the crop was first introduced to the United States (US), around 1800, primarily as a forage crop (Probst & Judd, 1973). During the 19th century, a multitude of additional introductions were made from several oriental countries (Probst & Judd, 1973 and references therein). Scientific breeding programmes made it possible to make the transition from hay to grain production (Burton, 1997), and by 1950 the US became the world's largest producer of soybean (Hymowitz, 1970). Since then, intensive breeding programmes have led to increases in seed yield potential, incorporation of genes for resistance to diseases and pests, and further adaptations to even wider geographical areas (Burton, 1997). The latter allowed the production to increase dramatically in other parts of the Americas, and during the last decades the largest production growth worldwide has been in Brazil and Argentina (FAO, 2016). Although the US was still the world's leading soybean producer in 2014 (FAO, 2016), Argentina is predicted to take over the position by 2030 (Masuda & Goldsmith, 2009).

A major technological breakthrough came in the late 1980s, when *Agrobacterium*-mediated DNA transfer made it possible to insert various beneficial genes into several crops (e.g. rapeseed: Fry et al., 1987; tomato: McCormick et al., 1986; and cotton: Umbeck et al., 1987), including soybean (Hinchee et al., 1988). This allowed for the development of soybean varieties that are resistant to glyphosate, the most commonly used herbicide worldwide (e.g. Duke & Powles, 2008). The introduction of

genetically modified (GM) crops has contributed to increasing yields (e.g. Qaim & Zilberman, 2003) and reducing pesticide use (e.g. Phipps & Park, 2002). At the same time, it has raised significant concerns regarding environmental and health risks, adverse social implications, monopolization of seed markets, and exploitation of smallholder farmers (Qaim, 2009 and references therein). Nevertheless, GM crops covered 182.8 million ha in 2016, whereof half was occupied by herbicide tolerant soybean (ISAAA, 2016).

Multiple factors within agronomy have thus made the soybean among the most renowned crops in the world, and the attention it has received is well deserved. The soybean is often referred to as 'the wonder bean', because of the great variety of products that can be derived from the protein-rich oilseed. Soybeans can be consumed directly as a whole seed or incorporated in food like tofu, but the vast majority are crushed into meal (coarse-ground flour) and oil (HighQuest Partners & Soyatech, 2008). A crushed seed produces about 79% meal which is primarily used as protein supplement in animal feed (HighQuest Partners & Soyatech, 2008). Only 2.5% of the crushed seed is wasted, while 18.5% becomes oil which is used in human foods, biodiesel production, and industrial applications (HighQuest Partners & Soyatech, 2008).

As the per capita income and consumption is increasing (especially in developing countries), so is the request for animal protein. Simultaneously, the augmenting global biodiesel production is leading to a higher demand for vegetable oil. According to a forecasting model made by Masuda & Goldsmith (2009), the global demand for soybean will be 371.3 million t in 2030, approximately 1.4 times as high as it was in 2012-14 (FAO, 2016). Argentina is projected to be the main provider of soybean by this time, supplying 29.2% of the world production (Masuda & Goldsmith, 2009).

Soybean production in Argentina

Argentina is already the world's largest exporter of both soybean meal and oil (2014 data from FAO, 2016; HighQuest Partners & Soyatech, 2008). This is because of the low demand for these products nationally due to the low population density and the tradition of free-range cattle feeding (HighQuest Partners & Soyatech, 2008). The export of soybean products represents a significant income source for the country

($17.34 billion in 2016 according to the OEC), and played a particularly important role in the economic recovery following the 2002 financial crisis (Grau et al., 2005). The increasing demand for soybean products internationally thus leads to expansions of soybean fields in Argentina, often at the expense of specific habitat types such as temperate and sub-tropical forests and grasslands (Grau et al., 2005). Habitat loss as a consequence of the soybean industry has been reported in neighbouring countries as well (e.g. Fearnside, 2001; Kaimowitz & Smith, 2001).

The key to keep augmenting the production without having to expand the fields is to increase yields (i.e. production per ha). To accomplish this, the bulk of research has focused on how to further improve genetic characteristics like increasing stress tolerance and seed growth (e.g. Ge et al., 2016; Komatsu et al., 2015; Sinclair et al., 2007; Sinclair et al., 2004; Valliyodan et al., 2016). However, some studies have addressed the potential importance of insect pollination as means to improve yields (e.g. Chiari et al., 2005b; Milfont et al., 2013; Monasterolo et al., 2015).

Pollination of soybean

The soybean is a self-fertilizing plant, but pollinating insects, especially bees (Hymenoptera: Apoidea: Apiformes), are regularly found to visit soybean flowers (e.g. Jaycox, 1970; McGregor, 1976). The flower morphology involves nectaries and nectar guides, suggesting that flower visitations are rewarding for both the plant (reproduction) and the visiting insect (food) (Palmer et al., 2009). Based on several studies (e.g. Chiari et al., 2005b), Klein et al. (2007) classified the soybean crop as clearly benefiting from insect pollination (10-40% decline in production without pollinators). Several later experimental studies have supported this conclusion (e.g. Blettler et al., 2017; Milfont et al., 2013; Monasterolo et al., 2015). The quality of these studies however, varies greatly. For instance, both Chiari et al. (2005b) (reviewed in Klein et al., 2007) and Milfont et al. (2013) used large enclosures (12 m^2 and 18 m^2, respectively) to exclude flower visitors, and compared the production yield between enclosed plants and plants receiving natural pollination. However, none of them controlled for potential impact of shading on yield, and both studies also present some degree of pseudoreplication.

Studies of poor quality are found in any discipline, mostly because of the methods employed. As methodological flaws lead to unreliable results, improving sampling methods in both experimental and observational studies is important, and needs to be considered continuously.

1.2 Sampling of pollination data

A great variety of methods are used to sample data in biological observational studies (Greenwood & Robinson, 2006; Southwood & Henderson, 2000), also within the field of pollination ecology (Kearns & Inouye, 1993; Nielsen et al., 2011; Westphal et al., 2008). Among the most widely used methods to assess pollination frequencies, pollinator diversity and abundance, and even entire plant-pollinator communities are transect walks, plot samplings, and pan traps (Nielsen et al., 2011; Westphal et al., 2008). Insect collection with pan traps is an example of a passive sampling method, while transect walks and plot samplings involve active counting of pollinators or pollination events by field workers. A disadvantage of the latter is that varying experience between the field workers may lead to observer or collector bias (Westphal et al., 2008). Active methods are also considered to be more time consuming than passive methods (but see Popic et al., 2013), yet they are essential in studies that aim to investigate pollination as an ecosystem service, i.e. plant-pollinator interactions, often referred to as *flower visits*. Furthermore, both active and passive methods are associated with taxonomical biases; pan traps seem to catch certain pollinator species less or more frequently than expected by their actual abundance (Cane et al., 2000; Wilson et al., 2008), and field workers may be more likely to collect the more conspicuous pollinator species over the smaller and more cryptic (Ausden & Drake, 2006).

Evidently, all methods have both benefits and drawbacks, and there is no method that is consistently superior for all purposes. The best choice of sampling methods thus depends on the objectives of the particular study. For detecting bee species richness, Westphal et al. (2008) found pan traps to be significantly more efficient than active sampling in both agricultural and semi-natural systems in Poland, England and Germany. Among the active methods, transect walks were significantly

more efficient than plot samplings. Their study design was repeated in Greece by Nielsen et al. (2011), who obtained similar results.

To assess pollination frequencies (e.g. number of flower visits or visitors per flower), however, pan trap sampling is inadequate as the insects' floral relations cannot be assessed. Calculating flower visitation frequencies also requires an estimate of the *flower exposure*, i.e. *the number of flowers that visits are counted over at any given time* (Reitan & Nielsen, 2016). It is therefore important to measure the exposure as correctly as possible, as this has a direct effect on the accuracy of the frequency estimate. Accomplishing this is challenging when using transect walks as compared to plot samplings. A relatively feasible technique to estimate the exposure is to regard it as the total number of flowers within the whole transect or plot area. This will, however, result in overestimations of the exposure when using transect walks, as the observer cannot view all the flowers within the transect simultaneously (Figure 1.1a). Conversely, with plot samplings the observer is continuously counting

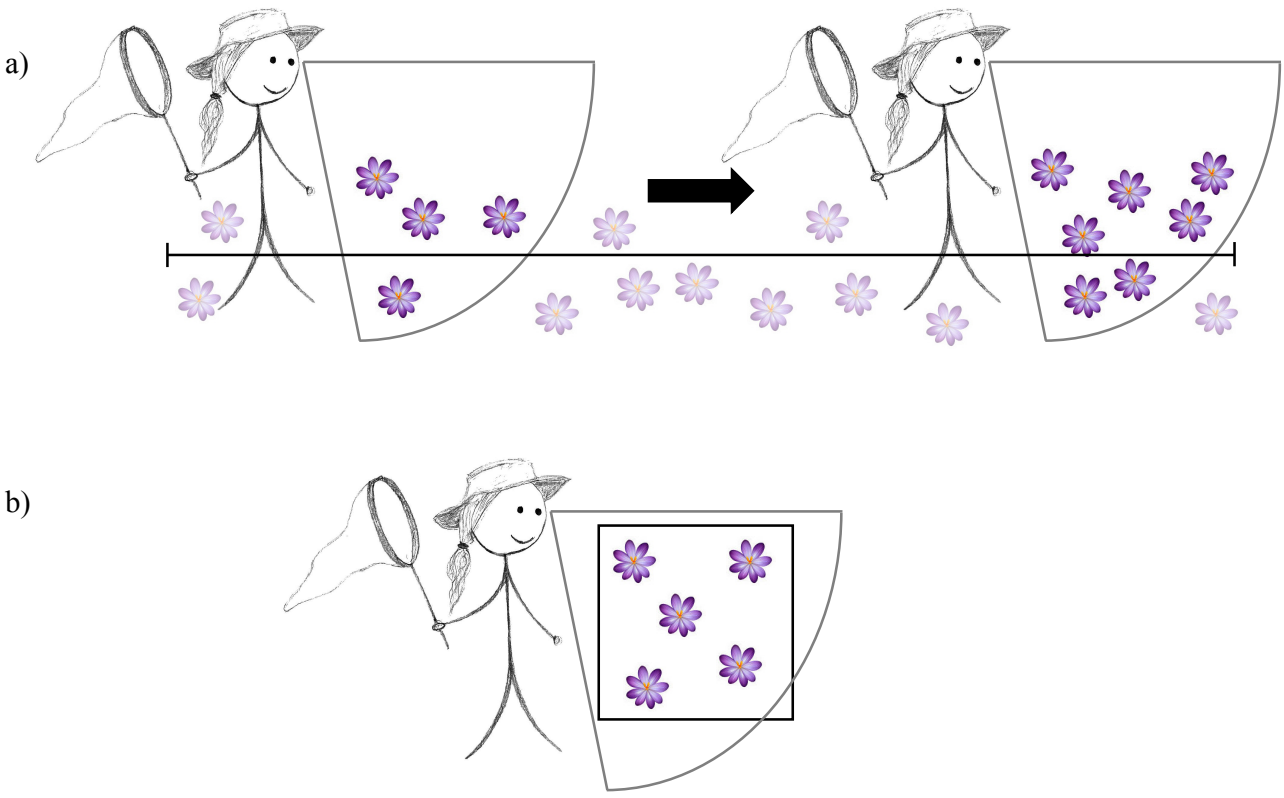


Figure 1.1: a) The number of observed flowers change continuously as the observer moves through the transect. If a flower visit occurs within the transect limit behind the observer, it will not be recorded. **b)** The observer is in complete control over all the flowers within the plot area throughout the sampling event, and is able to record all occurring flower visits.

visits over the exact same flowers for a given time period (assuming the plot is small enough for the observer to view all the flowers within it simultaneously). An accurate estimate of the exposure can then be achieved simply by counting the flowers inside the plot area (Figure 1.1b).

A drawback with plot samplings compared to transect walks, however, is that they often require a larger sampling effort (time spent observing) to obtain the same amount of flower visit observations, at least in heterogeneous plant communities (Gibson et al., 2011). It is not unreasonable to assume that this is also true in a presumably more homogenous system like a crop field, because bees tend to systematically visit neighbouring flowers before flying further away (M. Devoto, *pers. comm.*). This behaviour would make it more probable to discover visits when moving through larger areas during the sampling period. Obtaining enough observations is especially a problem in systems with low visitation frequencies. For example, Fijen & Kleijn (2017) have shown that when the actual visitation frequency is low, more time per plot sampling is needed to obtain accurate visitation frequency estimates.

Some soybean fields seem to have extremely low pollinator densities. In Chaco, Argentina, using linear sampling plots of 0.5 m in length along rows of soybean plants, Monasterolo et al. (2015) recorded less than 0.0007 flower visits per flower per hour of observation. Comparatively, Musicante (2013) (in Monasterolo et al., 2015) recorded more than 0.006 flower visits per flower per hour to forest plants nearby.

1.3 Objectives and hypotheses

Because of the potential impact that pollination services have on soybean yields, this topic deserves more research. Some studies have already aimed to investigate yield increments in soybean when visited by insects (e.g. Blettler et al., 2017; Chiari et al., 2005b; Milfont et al., 2013). Yet, few have focused on detecting the actual pollinator activity, or the factors that may affect pollinator populations, in modern soybean fields. The overarching aim of this thesis was therefore to assess the flower visitor and visitation frequencies in a highly intensive agricultural system, while simultaneously identifying the feasibility of different sampling methods. To

accomplish this, I carried out field surveys, using both transect walks and plot samplings, in monoculture fields of GM soybean in the Argentinian Pampean Region; an area where 90% of the original grasslands have been converted into fields used for agriculture or cattle-raising (Medan et al., 2011). The specific objectives and hypotheses are listed below.

Objective I: Assess how the frequency of potential soybean pollinators is affected by environmental conditions.

> **H1:** The frequency of bee visitors to soybean flowers is affected by ambient temperature, humidity and/or wind speed.

Objective II: Assess the feasibility of different sampling methods regarding **a)** estimation of flower visitor frequency and **b)** sampling effort.

> **H2-a:** Plot samplings yield higher estimates of the flower visitor frequency than transect walks, due to more accurate estimations of the exposure.

> **H2-b:** Transect walks yield higher counts of observed flower visitors per unit time than plot samplings.

Objective III: Quantify the visitation frequency to soybean in a highly intensive agricultural system with extremely few flower visitors.

> **H3:** In the system of the study, i.e. an intensively managed monoculture crop surrounded by few wild plants, the probability for a soybean flower to be visited once by a bee while open is low.

2 METHODS

2.1 Study area

The fieldwork took place in the counties Carlos Casares and Bolívar, located in the centre of the Province of Buenos Aires, Argentina. This area is part of the humid Pampas, which, together with the semi-arid Pampas, constitutes the most heavily populated region of the country (Dellafiore, a). The fieldwork was conducted around Estancia San Claudio (35°57' S, 61°12' W), a farm owned by the University of Buenos Aires (UBA). The climate in the area is sub-humid and the nearest city, Bolívar, has mean daily temperatures of 15-30°C in January and 2-14°C in July (Servicio Meteorológico Nacional). Changes in El Niño Southern Oscillation (ENSO) cause high inter-annual rainfall variability (Messina et al., 1999 and references therein); during 1979-1992 the annual precipitation varied from ~620 to ~1,300 mm, with an average of 911.5 mm (Omacini et al., 1995). The original biome of the ecoregion is grasslands dominated by *Stipa* spp. (Burkart et al., 1999), which are excellent food for grazers. The climatic conditions and geochemical characteristics make the soil highly suitable for vegetative growth (Burkart et al., 1999). Therefore, most of the natural vegetation has been converted to agricultural lands characterized by intensive crop farming and cattle grazing (Medan et al., 2011). The remaining semi-natural habitats grow in small patches along roadsides and in some abandoned fields, and are still under threat from conversion to agricultural land and degradation through over-grazing (Dellafiore, b). Because of this, the ecoregion is categorized as *Endangered* by the World Wildlife Fund (WWF), and is regarded as a high priority conservation area at the regional scale (Dellafiore, b).

The most common crops in the area are soybean, maize, winter wheat and winter barley (*Hordeum vulgare* L.). To prevent the build-up of soil-borne diseases, the sowing of the crops is often rotated in the course of two years. For a given lot, maize and soybean are sown every other summer in turn. Every second winter, wheat or barley is sown, and the next winter the lot is left fallow. Each year, the soybean is sown in two cohorts; the earliest in October (after maize) in a lot left fallow the previous winter, and the latest in December in a different lot after the winter cereal

is harvested. This gives two soybean flowering periods each lasting about a month, from late December and the beginning of February, respectively.

2.2 Study species

Soybean is a bushy herbaceous annual in the Fabaceae family, and in the genus *Glycine*. The cultivated form is *Glycine max*, and has most likely derived from the wild soybean *G. max* subsp. *soja* (Siebold & Zucc.) H. Ohashi (Burton, 1997). The crop usually has an erect growth habit and can reach a height of 1.8 m, but is highly diverse both genetically and morphologically (Williams, 1950). Figure 2.1 shows the GM soybean variety DM3070 (Rindes y Cultivos DAS, 2012), which was the cultivar planted at Estancia San Claudio. The legume has perfect flowers (i.e. flowers that contain both female and male reproductive organs) that are 3-8 mm in diameter when opened (Williams, 1950). The zygomorphic flowers are either white or some shade of purple. Their structure is typical for the subfamily Faboideae (= Papilionoideae), having the corolla formed by one posterior banner petal, two lateral wing petals, and two anterior keel petals. The androecium is located between the two keel petals, and consists of nine stamens that are fused by the filaments and one that is free (= diadelphous androecium). Surrounded by the androecium is a single unicarpellate pistil that turns into a legume (fruit) after fertilization. The style curves backward so the stigma points towards the free posterior stamen. The other stamens form a ring around the base of the stigma. Shortly before anthesis, the stamens start to elongate so that the ring of anthers is positioned around the stigma at the time of

Figure 2.1: Variety Don Mario 3070 has purple flowers, reaches a height of 92 cm and has an indeterminate growth habit. The variety is resistant to glyphosate. Photo: I. F. Asmervik

pollination (Williams, 1950). The pollen will then shed directly onto the stigma, and the flower has self-pollinated before it has even opened. According to Williams (1950), the stigma is receptive to pollen approximately 24 hours before, and remains so 48 hours after, anthesis. However, the time in which the flower remains open may vary among varieties and local environmental conditions (Gazzoni, 2016). Reports of observed blooming duration of individual soybean flowers have been from about one (Severson & Erickson Jr., 1984) to three (Chiari et al., 2005a) days.

Even though the soybean flowers self-pollinate, they have cup-shaped nectaries about 0.2-0.4 mm in height, located inside the staminal sheath (Erickson & Garment, 1979). Depending on the cultivar, the average nectar volume per flower may vary between 0.022 and 0.127 μL, as observed by Severson & Erickson Jr. (1984) on 17 soybean varieties. The total nectar carbohydrate content varied between 301 and 1,354 $\frac{\mu g}{\mu L}$ (Severson & Erickson Jr., 1984). Both the nectar and pollen of soybean flowers have been shown to be important resources to some beehives (Gazzoni, 2016 and references therein; Santos et al., 2013). Bees are considered to be the most important pollinators of other legumes (e.g. Bohart, 1960), and they are also the most often observed flower visitors to soybean. Jaycox (1970) reported to have observed two bumblebee species (*Bombus* spp.) and nine genera of solitary bees on soybean in the US. During two years, also in the US, Rust et al. (1980) collected 29 species of wild bees from the families Apidae, Anthoporidae, Megachilidae and Halictidae from 13 soybean varieties. Monasterolo et al. (2015) observed four Halictidae species to visit soybean flowers in Argentina. The honeybee is, however, the most common visitor to soybean flowers (e.g. Milfont et al., 2013; Monasterolo et al., 2015; Santos et al., 2013). The Western honeybee (*Apis mellifera*, Figure 2.2) is highly eusocial, and is a super-generalist pollinator. It was domesticated centuries ago, and is now the standard managed pollinator for most entomophilous crops. It is the most widely distributed bee species worldwide, and has

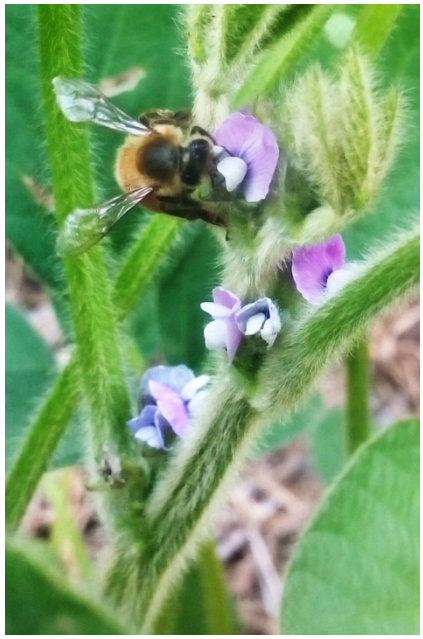


Figure 2.2: A honeybee visiting a soybean flower near San Claudio in 2016. Photo: I. F. Asmervik

evolved into several subspecies that have adapted to different climatic regions (Kjøhl et al., 2011). Keeping bees for crop pollination is not common practice in the study area, as the soybean is autogamous and the other most common crops are wind-pollinated. However, the area comprises some managed hives kept for honey production, and some feral populations.

2.3 Spatial outline of the study design

The fieldwork was carried out along a latitudinal gradient spanning ~200 km. I established three study locations about 100 km apart: 'north' (~35°06' S), 'mid' (~36°13' S), and 'south' (~36°28' S). The mid location was situated near the field station at San Claudio. Within each study location, I selected three study sites (Figure 2.3). Each site was defined as a soybean field, and located at least 1 km away

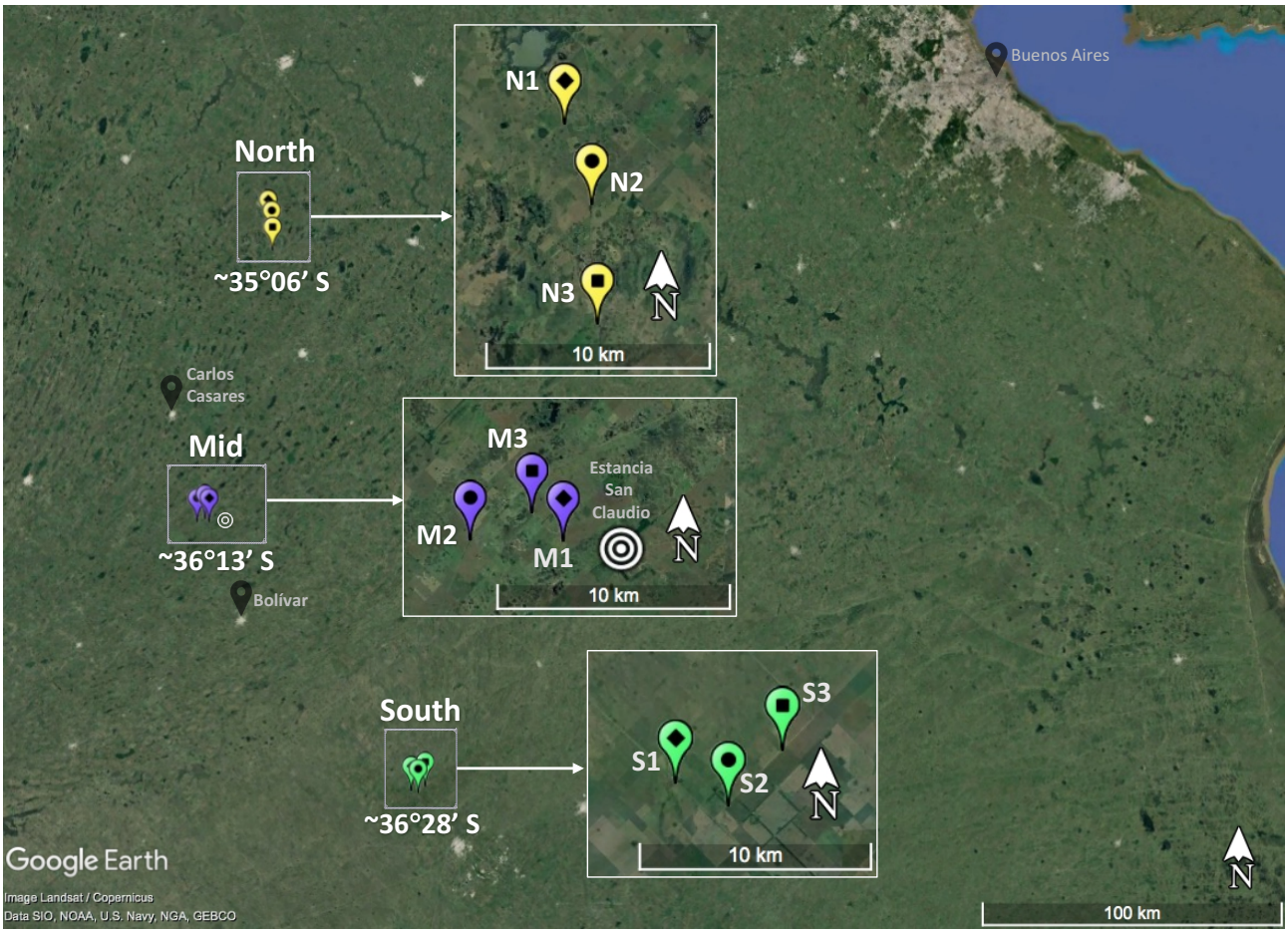


Figure 2.3: Overview of the spatial outline of the study design in the Buenos Aires province of Argentina. Nine sites (N1, N2, N3, M1, M2, M3, S1, S2, S3) were equally distributed among three locations (north, mid, south). The mid location was established near the field station at San Claudio, located between the county capitals of Carlos Casares and Bolívar.

from the other two within the particular location. The sites were selected by first driving approximately 100 km north and south from the field station, and then looking for soybean fields that were at least 4 ha (200 m × 200 m) and separated by at least 1 km.

In the south location, only one of the study sites (S3) had flowering plants by the beginning of the fieldwork, and the morphology of the crop differed somewhat from that in the other two study locations (Appendix A). Sampling from S1 and S2 began as soon as the plants were in bloom (17 January 2016, while the other sites were observed from 7 January 2016).

Environmental variables

To record environmental variables, I placed a weather logger (iButton – Hydrochron from Maxime Integrated) in each of the nine study sites. The loggers, mounted onto wooden poles 25 cm above ground (Figure 2.4), recorded ambient temperature and relative humidity every hour throughout the study period. Four plastic plates attached at the top of the pole protected the loggers from rain and direct sunlight (Figure 2.4). In addition, I measured temperature, humidity and wind speed with a handheld weather recorder (WeatherHawk: SM-28 Skymaster) in relation to each sampling event.

Figure 2.4: One of the weather loggers placed in each of the study sites. Photo: I. F. Asmervik

2.4 Data collection

I carried out the main data collection in 2016 with two field assistants from the UBA. The sampling was done during the flowering of the first soybean cohort between 7 and 25 January. Each sampling event lasted 20 min, and involved collecting all observed flower-visiting insects within a defined sampling unit. In accordance with the study objectives (I & II), the sampling intended to relate variation in flower visitor frequency to both environmental variables and sampling methods. By carrying out the sampling along a latitudinal gradient, I expected to obtain environmental variation, both among and within each of the three study

locations (Figure 2.3). To compare the effect of different sampling methods, I defined four distinct sampling units of contrasting size and outline; one transect (based on previous sampling conducted by López-Carretero et al., 2017), and three smaller sampling plots of different sizes. The transect was 100 m long and 2 m wide (200 m^2) and originally used to establish plant-pollinator interactions for network analyses. Based on previous studies in the area, I expected the frequency of flower visitors to be low (M. Devoto, *pers. comm.*). Therefore, I defined the smallest plot as a square of 16 m^2. The other plots were a square of 36 m^2 and a rectangle of 20 m × 4 m (80 m^2). Hereafter, the different sampling units will be referred to as 100×2, 4×4, 6×6 and 20×4.

Environmental variation

Thirteen days of the fieldwork were dedicated to assess the environmental and latitudinal variation (among the study locations) in flower visitor frequency (Objective I). In each study site, two observers carried out sampling simultaneously in two different sections of the focal field (i.e. parallel sampling, taking into account spatial variation within each site). A new sampling unit was established every day, but always within the respective sections of the fields. In total, we carried out 30 sampling events in north, and 24 sampling events in each of mid and south (78 in total). Most of these (68) were 6×6 plot samplings, as I initially expected a smaller sampling unit to give the most accurate estimate of the real flower visitor frequency. The remaining 10 sampling events were done with the 20×4 sampling unit.

Different sampling methods

Five days of the fieldwork were dedicated to assess the effect of different sampling methods on the perceived flower visitor frequency (Objective II-a) and count (Objective II-b). To maximize the number of sampling events, they were all carried out close to the field station (mid location). In total, we carried out 72 sampling events in site M2, and nine sampling events in each of the sites M1 and M3 (90 in total). In each site, three observers carried out sampling events within different sampling units simultaneously in different sections of the focal field (i.e. parallel sampling). A new sampling unit was established every day, but always within the

respective sections of the fields. We alternated sampling in different sampling units and field sections to be able to account for observer bias. An overview of the number of sampling events conducted with the different sampling units is shown in Table 2.1.

Table 2.1: The number of sampling events carried out with different sampling units for comparing the effect of different sampling methods on the observed frequency and count of flower visitors.

Sampling unit		Number of sampling events
Transect walks	100×2	30
Plot samplings	20×4	6
	6×6	30
	4×4	24

Figure 2.5: A 6×6 sampling unit defined by a flag in each corner of the square. Photo: I. F. Asmervik

The sampling events

Before we started a set of parallel sampling events, I measured the temperature, humidity and wind speed with the handheld weather recorder. I held the recorder with a straight arm at chest height, while facing the wind. The weather was also categorized as 'cloudy', 'partly cloudy' or 'sunny' (we did not sample during rain).

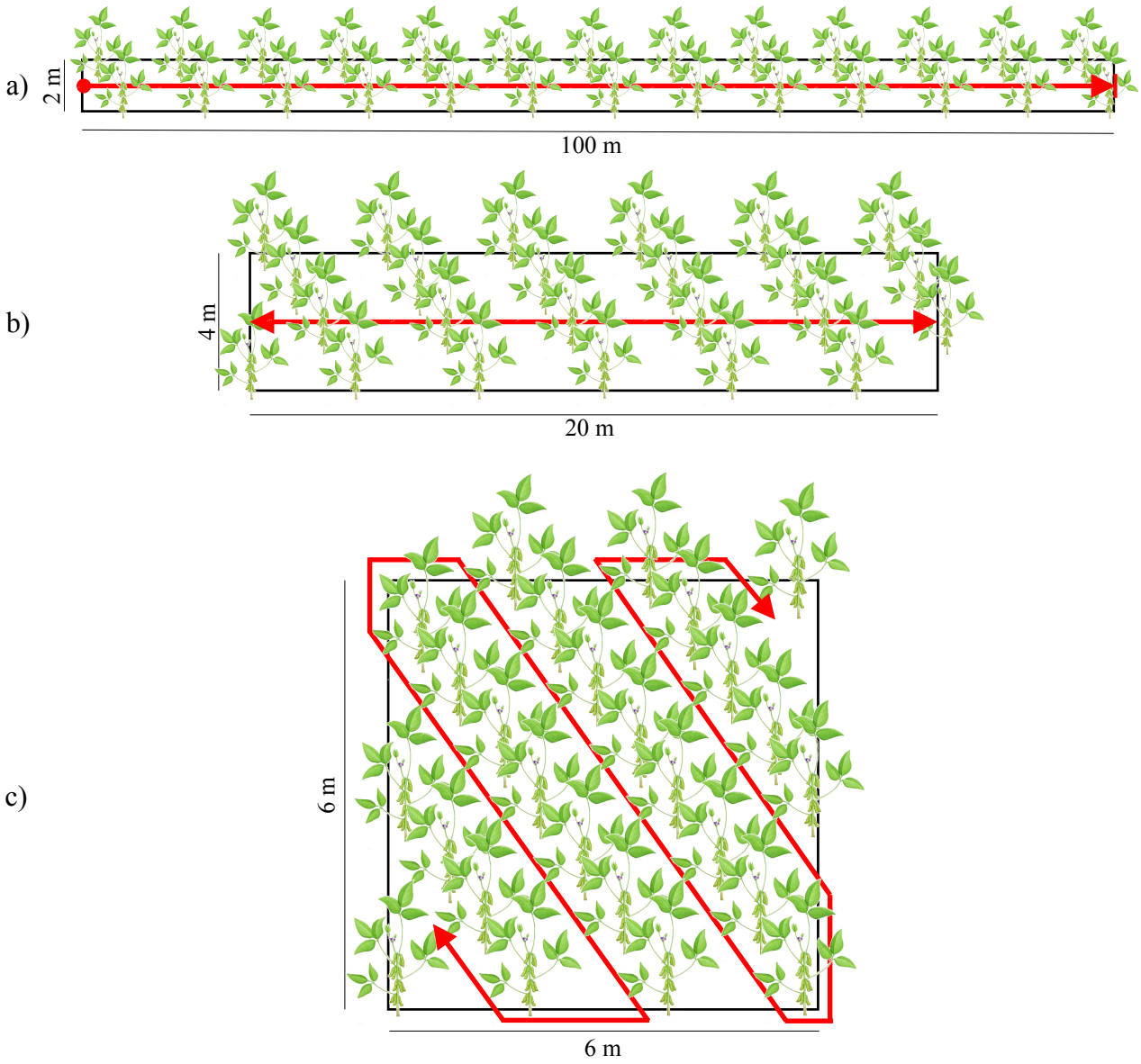


Figure 2.6: Illustrations (not reflecting real proportions to each other) of the observer's movements during a sampling event in different sampling units. **a)** In the 100×2 sampling unit, the observer moved through the unit once during 20 min. **b)** In the 20×4 sampling unit, the observer moved back and forth within the unit until the 20-min sampling period was over. **c)** In the 6×6 (and 4×4) sampling unit, the observer moved between the soybean rows within the unit until the 20-min sampling period was over.

Prior to each individual sampling event, we defined the sampling unit with flags (Figure 2.5). We carried out the 100×2 transect walks by moving through the transect once during the 20-min sampling period (Figure 2.6a). In the different plots, we moved at a similar pace (~5 $\frac{\text{m}}{\text{min}}$) within the defined sampling unit during the 20-min sampling period as illustrated in Figure 2.6b-c. As the vegetative parts of the soybean is quite dense (Figure 2.5), we used both sight and hearing to detect flower visitors.

The sampling itself was identical regardless of the sampling unit used. During the 20-min sampling period, all the insects observed to visit a soybean flower were put onto a vial and killed. If an insect could not be caught, it was noted as 'visiting individual', with a description of what kind of insect it was. In all these cases, the insect was specified well enough to be included in the analysis (6 honeybees, 1 *Megachile* and 1 non-*Apis* Apidae). The distance from the edge of the field to the sampling unit (30-120 m) was noted, and included in the analysis to account for possible edge effects. After each sampling event, we estimated the flower density by throwing 10 0.25 m^2 squares randomly throughout the sampling unit. We counted and noted the total number of flowers within the squares. After the fieldwork period, the flower density $\left(\frac{\#flowers}{m^2}\right)$ was calculated, and the total number of flowers in the sampling unit (exposure) could be estimated as $flower\ density \times unit\ area$. We also classified the captured and noted visiting individuals to the lowest possible taxonomical level (family, genus or species).

2.5 Datasets

Although the data collection was conducted using different sampling units, I could include all flower visitor recordings in the same dataset, as the flower density (exposure) and number of flower visitors for each sampling event had been recorded. I used this *primary* dataset to assess which factors, including environmental variables and sampling methods, explained most of the variation in the observed frequency of flower visitors to soybean (Objective I & II-a).

Based on the primary dataset, I made an *extended* dataset by adding data collected by my collaborators at the UBA in 2016. This data was sampled using 100×2

transect walks, and did not include any recorded environmental variables. I used the extended dataset to further assess how the different sampling methods affected the perceived flower visitor frequency (Objective II-a).

The measure of pollination in these two datasets were the number of flower-visiting individuals, not flower visits that could be used to assess visitation frequencies. In order to address Objective III, I therefore needed to assess the visitation frequency of bee individuals. To do this, I obtained a *revisits* dataset containing individual visitation data collected by my collaborators at the UBA during January 2017. The collection was done as follows: In order to obtain an adequate number of observations, the field workers looked for bees anywhere in the field, but at least 10 m from the field edge. When detecting a bee, the field worker counted the number of flowers that the bee visited before flying away more than two metres from the initiation point. The total number of visited flowers were recorded and included in the dataset.

For a total overview of the three datasets, see Appendix B.

2.6 Statistical analyses and data treatment

I made all statistical analyses, plots and calculations using the R programming environment, version 3.3.2 (R Core Team, 2016). All the presented models were generated using the 'glmer' function in the R package 'lme4', version 1.1.12 (Bates et al., 2015). The generation of some of the appendices (C, E & G.1) was facilitated by the use of R Markdown (RStudio Inc., 2016).

2.6.1 Assessment of effects on flower visitor frequency

To assess how the frequency of potential soybean pollinators varied with both environmental- (Objective I) and sampling method- (Objective II-a) related factors, I used the primary dataset to make a model. As bees are considered the most efficient pollinators of typical Faboideae flowers (e.g. Bohart, 1960), I selected the number of observed flower-visiting bees (b) compared to the number of flowers (f) to represent the pollinator frequency. A basic and relatively common approach to assess such a quantity statistically is linear modelling (LM), using the flower visitor (or visitation

when visits have been recorded) frequency $\left(\frac{b}{f}\right)$ as the response variable. However, this approach entails several problems. For example, the fitted values may become negative (i.e. not making sense), and heterogeneity (non-constant variance) is likely (i.e. violating one of the LM assumptions) (Zuur et al., 2009). Moreover, valuable information is lost by not including the bee and flower counts as separate entities in the model (Reitan & Nielsen, 2016). The use of LM also requires that the data is independent, which is not necessarily the case when collected repeatedly across different sites as in the present study (Figure 2.3). To account for the difficulties explained above, I used generalized linear mixed modelling (GLMM) (Bolker et al., 2009), with b as the response variable and f as an offset variable (Reitan & Nielsen, 2016). Using GLMM compared to LM allows for 1) assessing response variables that are not normally distributed[2] (de Jong & Heller, 2008), 2) assessing non-linear relationships and 3) detecting random effects (e.g. variation between sites or field sections) in addition to exact (i.e. fixed) effects (Bolker et al., 2009). As b is a count variable, I generated the model with a Poisson error distribution.

Explanatory variables and model selection

To assess which variables best explained my bee visitor data, I used a model selection function written by Trond Reitan (Appendix C). The function takes a list of covariates potentially explaining some variation in the response variable, generates and compares a great number of models, and returns the *best model* based on an information criterion (IC). It is desirable to obtain an IC value as low as possible, i.e. the model with the lowest IC value is the best model. I ran the function through the Abel computing cluster (UiO, 2012), as the process is computer intensive when including many covariates. A more detailed description of how the function works can be found in Appendix C.

I used the Bayesian information criterion (BIC), as this is recommended when handling a lot of explanatory variables because it penalizes complexity in a higher degree than other ICs. The list of potential covariates included 73 terms

[2] GL(M)Ms can be generated for response variables following any probability distribution from the exponential family, e.g. the Poisson and Bernoulli distributions.

(Appendix D.1). The covariates of most interest (with respect to the hypotheses to be tested) were the environmental, latitudinal and sampling method related variables. The environmental variables included temperature and humidity from both the weather loggers and the handheld weather recorder, in addition to wind speed (from the handheld recorder only). The latitudinal gradient was quantified both as a numerical variable of the exact latitude and a factor variable describing the spatial configuration of the study locations (north, mid and south, Figure 2.3). I included three sampling method variables: a factor variable distinguishing between transect walks and plot samplings, a factor variable of the different sampling units (100×2, 4×4, 6×6 and 20×4), and a numerical variable of the unit areas (200, 16, 36 and 80 m^2). Several interactions, quadratic terms and logarithmic terms were also included in the set of potential covariates (Appendix D.1).

The covariate list also included several random effects. The variables 'study site' (9 levels; Figure 2.3) and 'field section' (26 levels) were included to test for spatial variation on different scales. I also included a variable distinguishing between the observers (3 levels) to test for observer bias, and a variable representing each of the sampling events (n levels) to account for possible overdispersion (i.e. unexplained variation). Other random terms, including random slopes, are listed in Appendix D.1.

2.6.2 Further assessment of sampling method effects

Number of observations per flower per sampling event (frequency)

I made an additional analysis of how the sampling methods may affect the perceived flower visitor frequency (Objective II-a) by generating a second *best model*, based on the extended dataset. This was done to see if records from more sampling events (n_{ex} = 324, n_{pr} = 168), despite missing environmental variables, would result in a superior model. To generate the model, I used a similar approach as explained above. However, in this analysis I included the number of observed flower-visiting hoverflies (h) and used $b + h$ as the response variable, as the efficiency of potential pollinators was not important for the hypothesis to be tested. I did not include other flower visitors such as beetles (Coleoptera), because it varied considerably between observers whether they were recorded or not when observed. As $b + h$ is also a count

variable, I used Poisson as the error distribution. f was also here included as an offset variable. I found the best model using the model selection function (Appendix C). As the extended dataset consisted of fewer variables than the primary dataset, the list of potential covariates used in this analysis included only 40 terms (Appendix D.2). The covariates of most interest (with respect to the hypothesis to be tested) were the sampling method related variables, but the latitudinal variables ('study location' and 'latitude') were also included. Appendix D.2 provides the full list of the tested fixed effects, including quadratic and logarithmic terms. Included in the list were also random variables testing for unsystematic spatial variation ('study site' – 9 levels, 'field section' – 87 levels), observer bias ('observer' – 7 levels) and overdispersion ('unexplained variation'). Additionally, I included a variable distinguishing between the two groups of observers working together during the fieldwork. See Appendix D.2 for other random terms, including random slopes.

For simplicity, the two *best models* of observed flower visitor frequencies will hereafter be referred to as 'including environmental variation' ('IE'), and 'excluding environmental variation' ('EE'), respectively (Table 2.2).

Table 2.2: Overview of the two GLMMs generated through the model selection procedure using different datasets.

Model	Response	Offset	Dataset	Principal use
Flower visitor frequency, including environmental variation (IE)	Bees (b)	Number of flowers within sampling unit (f)	Primary	Objective I & II-a
Flower visitor frequency, excluding environmental variation (EE)	Bees and hoverflies ($b + h$)	Number of flowers within sampling unit (f)	Extended	Objective II-a

Number of observations per sampling event (count)

To assess how different sampling methods may affect the sampling effort needed to obtain a given number of observations (Objective II-b), I generated a model of the flower visitor count (per sampling event). This was done by reproducing the model in Table 2.2 that turned out to best explain the variation in the perceived flower visitor

frequency, while excluding the offset variable. The outcome of the count model was thus the expected number of observed flower visitors per sampling event, regardless of how many flowers were counted over.

2.6.3 Assessment of the flower visitation frequency

To quantify the visitation frequency to soybean in the study system (Objective III), I calculated two estimates of the *flower visitation probability*. I define this as '*the probability that a flower will be visited by a bee at least once during the time it is open and the stigma is receptive to pollen*'. One of the estimates was based on the raw data from the primary dataset, and the other was based on the best model of flower visitor frequency IE.

To obtain the probability estimates, I first calculated the mean number of bee visitors per flower per sampling event ($bee\_visitors$). I used Equation 1 for the model estimate, and Equation 2 for the raw data estimate.

$$bee\_visitors_{mod} = \text{mean}(e^{y_i}) \quad \text{(Equation 1)}$$

y_i = expectation of the best model for sampling event i

$$bee\_visitors_{raw} = \text{mean}\left(\frac{b_i}{f_i}\right) \quad \text{(Equation 2)}$$

b_i = number of flower-visiting bee individuals of sampling event i
f_i = number of flowers within the sampling unit of sampling event i

I then used the revisits dataset to estimate the potential number of flower visits conducted by a single bee if it had not been caught (bee_visits). This was done according to Equation 3.

$$bee\_visits = \text{mean}(v_j) \quad \text{(Equation 3)}$$

v_j = number of flower visits for bee j

Then, I estimated the total number of visits per sampling event ($total\_visits$) according to Equation 4.

$$total\_visits = bee\_visitors \times bee\_visits \quad \text{(Equation 4)}$$

Finally, I estimated the flower visitation probability (p) using Equation 5, where '20 min' represent the period of a sampling event. I made the assumptions that 1) flower-visiting bees are active for 12 hours a day (12×60 min), and 2) the flowers are open with receptive stigmas for 2 days.

$$p = total\_visits \times \frac{12{\times}60 \text{ min}}{20 \text{ min}} \times 2 \text{ days} \qquad \text{(Equation 5)}$$

In addition, I calculated 95% confidence intervals for both probability estimates using bootstrap (Efron & Tibshirani, 1986). Detailed calculations made to generate the probabilities and confidence limits can be found in Appendix E.

3 RESULTS

The main fieldwork resulted in 168 sampling events with records of flower visitors to soybean in central Buenos Aires during 2016. Only 68 bee individuals (89.7% honeybees) and 6 hoverfly individuals were observed during 38 of the sampling events (i.e. in 130 of the sampling events no flower visitors were observed). An overview of the taxonomical diversity of the flower visitors is listed in Table 3.1.

Table 3.1: Taxonomical diversity of the observed flower visitors on soybean during 168 sampling events of 20 min in central Buenos Aires, Argentina, in 2016.

Lowest taxonomical rank (species, genus or family)		Total number
Apoidea	*Apis mellifera*	61
	Apidae (not *Apis*)	3
	Megachile sp.	3
	Halictidae	1
Syrphidae	*Palpada* sp.	4
	Toxomerus sp.	2

3.1 Factors explaining flower visitor frequency

The best model of the observed flower visitor frequency IE included the fixed effects 'humidity', 'temperature' and 'sampling method' (Table 3.2). A quadratic term of the humidity variable was also included (Table 3.2), suggesting an optimum in flower visitor frequency when the relative humidity is 57.29% (Figure 3.1). The observed frequency increased with temperature (Figure 3.2), and decreased when using transect walks (sampling unit 100×2) compared to plot samplings (sampling units 4×4, 6×6 and 20×4) (Figure 3.1 & 3.2). The best model included the binary variable 'sampling methods' rather than the 'unit area' (numeric, ΔBIC = 1.2) or 'sampling unit' (4 levels, ΔBIC = 9.8) variables. The humidity measured by the handheld weather recorder explained the data better than the humidity measured by the

weather loggers (ΔBIC = 10.8). This was reversed for temperature, i.e. the logger-temperatures explained the data better than those measured by the handheld recorder (ΔBIC = 3.8). Latitude was not included in the best model, and it did not represent any real climate gradient (Appendix F).

Among the environmental variables, humidity explained most of the variation (Table 3.3). The random effects 'field section' (spatial variation on within site scale) and 'observer' were also included in the best model, comprising about 35% of the variation in flower visitor frequency (Table 3.3).

Table 3.2: The GLMM that best explained the flower visitor frequency (bees) on soybean along a latitudinal gradient (35°04' S - 36°44' S) in Buenos Aires in 2016, based on 168 sampling events, and according to the BIC value. **Humidity =** the relative humidity (%) obtained from the handheld weather recorder, measured prior to the sampling event. **Temperature** = the temperature (°C) obtained from the nearest weather logger, measured within half an hour from the start of the sampling event. **Sampling method** = binary variable distinguishing between 'transect' (sampling unit 100×2) and 'plot' (sampling unit 4×4, 6×6 and 20×4), where 'plot' is the reference level. **SE** = standard error. The **95% confidence limits** were calculated as $estimate \pm 1.96 \times SE$. Random effects are 'field section' (n = 26) and 'observer' (n = 3).

			95% confidence limits	
Fixed effect	Estimate	SE	Lower	Upper
Intercept	−48.4	13.1	−74.1	−22.7
Humidity	1.21	0.446	0.331	2.08
Humidity^2	−0.0105	0.00380	−0.0180	−0.00307
Sampling method (transect)	−0.824	0.294	−1.40	−0.248
Temperature	0.109	0.0430	0.0251	0.194

Table 3.3: The relative contribution of explained variation in flower visitor frequency (assumes that all covariates are independent). **Humidity =** the relative humidity (%) obtained from the handheld weather recorder, measured prior to the sampling event. **Temperature** = the temperature (°C) obtained from the nearest weather logger, measured within half an hour from the start of the sampling event. **Sampling method** = binary variable distinguishing between 'transect' (sampling unit 100×2) and 'plot' (sampling unit 4×4, 6×6 and 20×4). **Field section** = factor variable describing the spatial variation on the 'within site' level. **Observer** = factor variable distinguishing between the field workers. The fixed effect variances are that of the model estimate multiplied by the raw data of the variable (var($estimate \times \boldsymbol{variable}$)). The random effect variances are model estimates.

	Covariate	Variance contribution (%)
Fixed effects	Humidity	55.84
	Temperature	7.16
	Sampling method	4.07
Random effects	Field section	23.36
	Observer	9.57

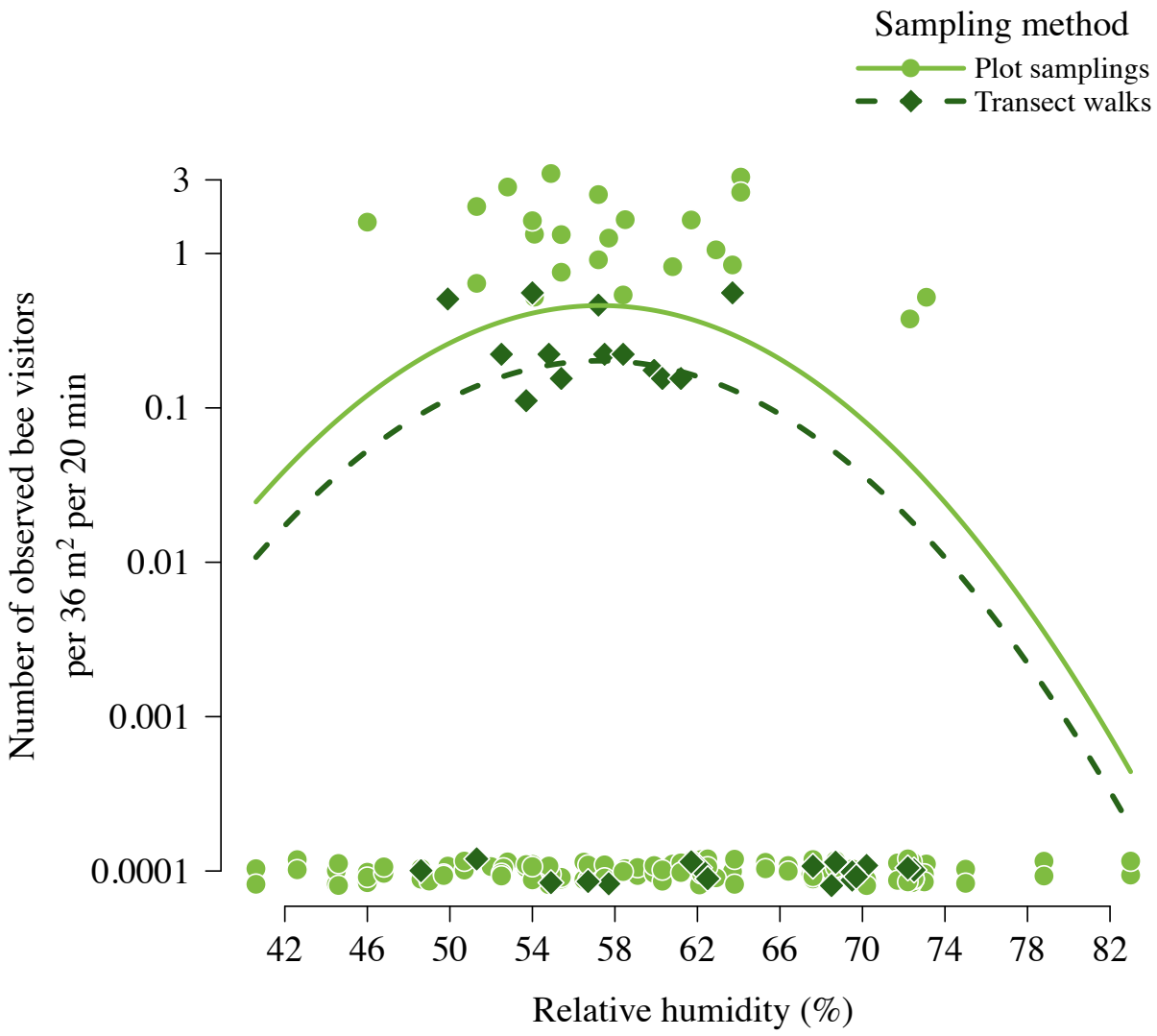


Figure 3.1: The lines describe the expected effect of humidity on the observed flower visitor frequency for two sampling methods (plots and transect) when the temperature is 30.28°C (mean value from dataset). The points are raw data estimates ($\frac{\#bees}{\#flowers}$). Both the model and raw data estimates are multiplied with the mean number of flowers in the 6×6 plots (12,053.39) to get more comprehensible and relevant values on the y axis. The y axis is log scaled, and 0.0001 was added to all the raw data values so that the sampling events with zero observations could be included in the plot. The 'jitter' function in R was also applied to these values (= 0.0001) to show overlapping data points.

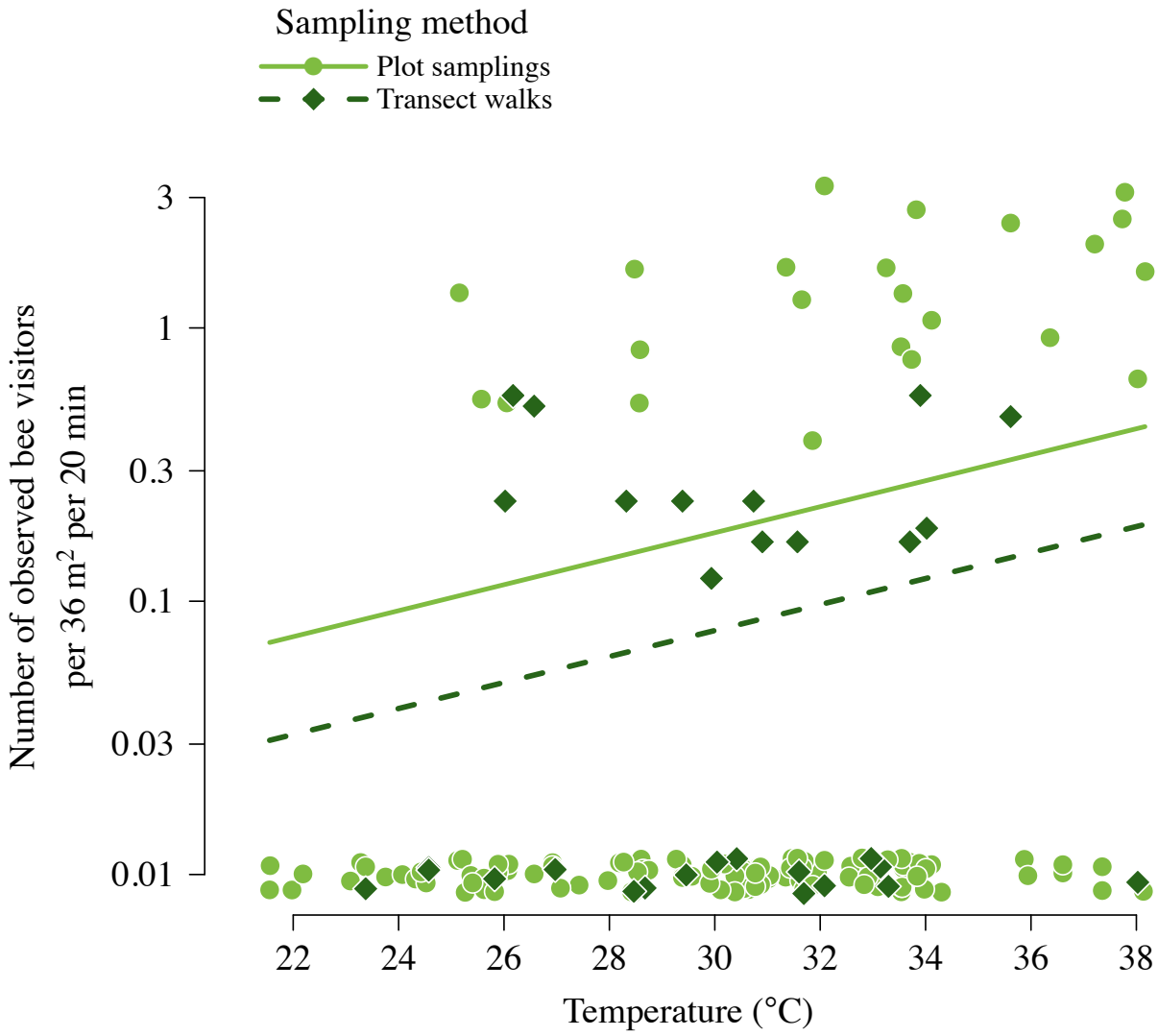


Figure 3.2: The lines describe the expected effect of temperature on the observed flower visitor frequency for two sampling methods (plots and transect) when the relative humidity is 60.18% (mean value from dataset). The points are raw data estimates ($\frac{\#bees}{\#flowers}$). Both the model and raw data estimates are multiplied with the mean number of flowers in the 6×6 plots (12,053.39) to get more comprehensible and relevant values on the y axis. The y axis is log scaled, and 0.01 was added to all the raw data values so that the sampling events with zero observations could be included in the plot. The 'jitter' function in R was also applied to these values (= 0.01) to show overlapping data points.

3.2 Comparison of sampling methods

3.2.1 Flower visitor frequency

The best model of perceived flower visitor frequency EE included the fixed effect 'sampling method' (Table 3.4). The negative estimate suggests that the number of observed flower visitors per flower (per sampling event of 20 min) was significantly lower when using transect walks (Table 3.4; Figure 3.3). The random effects 'unexplained variation' and 'field section' were also included in the model, and accounted for about 85% of the variation in observed flower visitors (Table 3.5).

Table 3.4: The GLMM that best explained the perceived flower visitor frequency (bees and hoverflies) on soybean in central Buenos Aires in 2016, based on 324 sampling events, and according to the BIC value. **Sampling method** = binary variable distinguishing between 'transect' (sampling unit 100×2) and 'plot' (sampling unit 4×4, 6×6 and 20×4), where 'plot' is the reference level. **SE** = standard error. The **95% confidence limits** were calculated as *estimate* ± 1.96×SE. Random effects are 'unexplained variation' (n = 324) and 'field section' (n = 87).

			95% confidence limits	
Fixed effect	Estimate	SE	Lower	Upper
Intercept	−11.6	0.364	−12.3	−10.9
Sampling method (transect)	−1.21	0.326	−1.85	−0.570

Table 3.5: The relative contribution of variation in perceived flower visitor frequency (assumes that all contributors are independent). **Sampling method** = binary variable distinguishing between 'transect' (sampling unit 100×2) and 'plot' (sampling unit 4×4, 6×6 and 20×4). **Unexplained variation** = unique ID for each sampling event. **Field section** = factor variable describing the spatial variation on the 'within site' scale. The fixed effect variance is that of the model estimate multiplied by the raw data of the variable (var(*estimate*×***variable***)). The random effect variances are model estimates.

	Covariate	Variance contribution (%)
Fixed effects	Sampling method	14.54
Random effects	Unexplained variation	51.46
	Field section	34.00

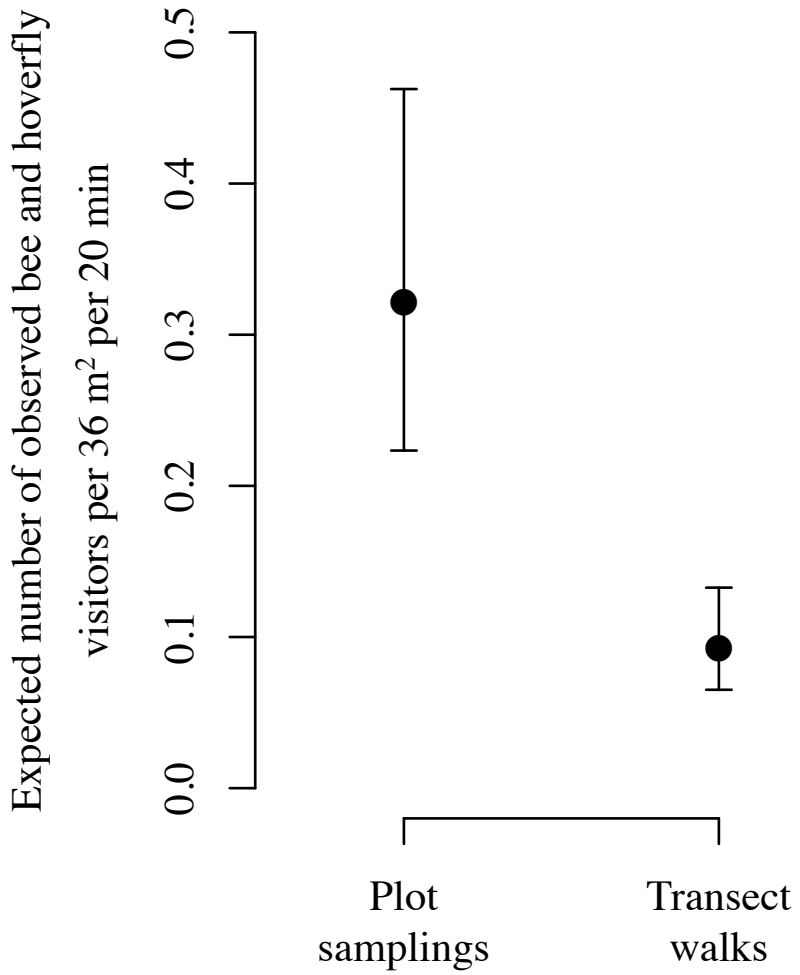


Figure 3.3: The expected effect of the sampling method on the perceived flower visitors per flower per unit time (frequency). **Plot samplings** = Sampling unit is 4×4, 6×6 or 20×4. **Transect walks** = Sampling unit is 100×2. The error bars demonstrate the standard error of the estimates. The estimates and standard errors are multiplied with the mean number of flowers in the 6×6 plots (12,053.39) to get more comprehensible and relevant values on the y axis.

3.2.2 Flower visitor count

I based the flower visitor count model on the frequency model IE. Contrary to what was the case in the frequency models, the estimate of the 'sampling method' variable was significantly positive (Table 3.6). This means that the number of observed bees per unit time (20 min) was significantly higher when using transect walks compared to plot samplings (Figure 3.4).

Table 3.6: The flower visitor frequency model IE without offset variable. **Humidity =** the relative humidity (%) obtained from the handheld weather recorder, measured prior to the sampling event. **Temperature** = the temperature (°C) obtained from the nearest weather logger, measured within half an hour from the start of the sampling event. **Sampling method** = binary variable that distinguishes between 'transect' (sampling unit 100×2) and 'plot' (sampling unit 4×4, 6×6 and 20×4), where 'plot' is the reference level. **SE** = standard error. The **95% confidence limits** were calculated as *estimate* ± 1.96×SE. Random effects are 'field section' (n = 26) and 'observer' (n = 3).

			95% confidence limits	
Fixed effect	Estimate	SE	Lower	Upper
Intercept	−41.5	29.3	−98.9	15.9
Humidity	1.32	1.02	−0.667	3.32
Humidity^2	−0.0117	0.00874	−0.0287	0.00536
Sampling method (transect)	**1.06**	**0.298**	**0.476**	**1.646**
Temperature	0.0935	0.0452	0.00492	0.182

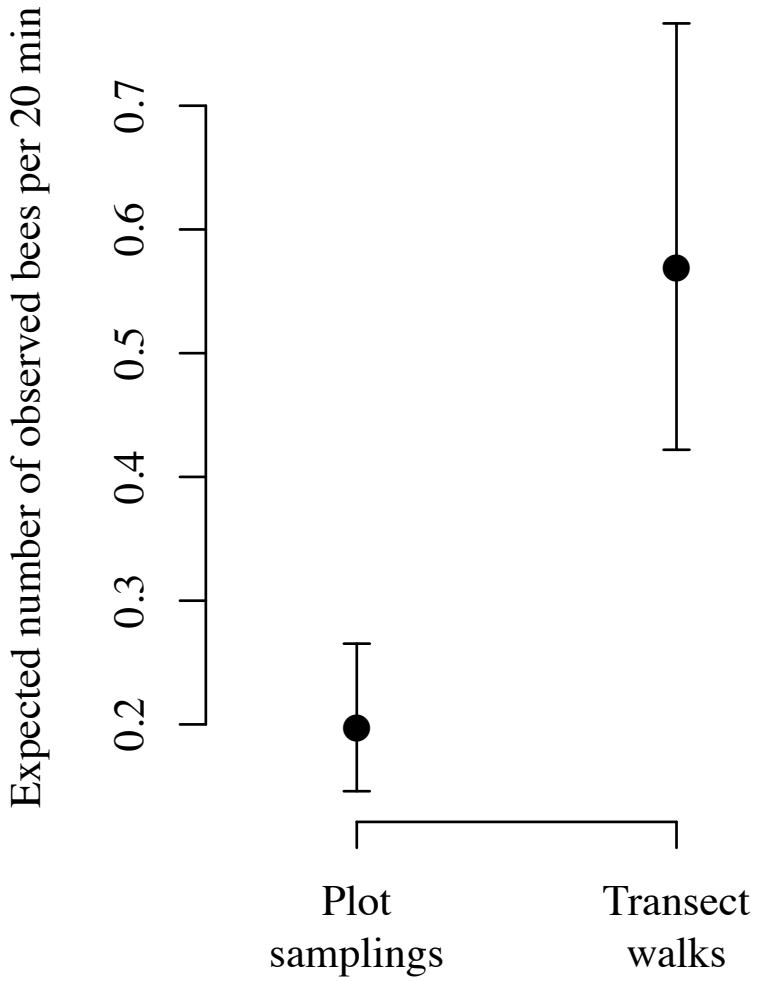


Figure 3.4: The expected effect of the sampling method on the perceived number of flower visitors per unit time (count). **Plot samplings** = Sampling unit is 4×4, 6×6 or 20×4. **Transect walks** = Sampling unit is 100×2. The error bars demonstrate the standard error of the estimates.

3.3 Flower visitation probability

I made two estimates of the flower visitation probability (i.e. the probability for a soybean flower to be visited at least once during the time it is open). The first was calculated using the frequency model IE, and the second was calculated from the raw data. The mean values of temperature (30.28°C) and humidity (60.18%) were used to calculate the model estimate, and the sampling method was set to be 'plot'. When calculating the raw data estimate, I included only the sampling events that had been conducted using plots.

The revisits dataset contained the number of visits from 30 individual bees. The mean number of flower visits the bees had made before flying away was 29.13 ± 4.54 SE. The flower visitation probability was estimated to be lower than 6% (Table 3.7).

Table 3.7: Estimates of the probability for a soybean flower in central Buenos Aires in 2016 to be visited at least once during the time it is open and the stigma is receptive to pollen.

		Bootstrap 95% confidence levels	
	Probability estimate (%)	Lower	Upper
Model estimate	5.23	2.04	8.57
Raw data estimate	4.22	1.61	6.21

4 DISCUSSION

Overall, extremely few insects were observed to visit soybean flowers during the fieldwork; on average, we observed only ~0.4 individual flower visitors per sampling event of 20 min. This was not unexpected, as the study area is under intensive agricultural management and mostly wind-pollinated crops surrounded the focal soybean fields. The most abundant flower visitor was the honeybee (~82% of the observed flower-visiting insects, Table 3.1), consistent with previous studies in the region (~90% in Milfont et al., 2013; ~55% in Monasterolo et al., 2015; 70.5% in Santos et al., 2013).

4.1 Environmental effects

To assess how the pollinator frequency may be affected by environmental conditions (Objective I), I measured the ambient temperature, relative humidity and wind speed associated with each sampling event. By comparing the hourly measured environmental variables of the three study locations forming the latitudinal gradient (Figure 2.3), I found that they did not represent a significant climate gradient (Appendix F). Nevertheless, the environmental conditions varied substantially among sampling events, and explained ~63% of the variation in flower visitor frequency (Table 3.3). Separately, temperature and humidity explained ~7 and ~56%, respectively, of the variation. However, environmental variables are often correlated, and this was also the case in my study (Pearson's r between 'temperature' and 'humidity' = –0.36, $p < 0.001$ using 'cor.test'). This makes the distribution of the variance contributions among the environmental variables relatively uncertain, and it is impossible to conclude on the relative importance of the humidity and temperature effects. Their combined variance contribution is more reliable, however, as there was no correlation between 'sampling method' and either 'humidity' or 'temperature' ($p > 0.05$ using 'cor.test'). The best frequency model IE therefore indicates that the flower visitor frequency in the area is clearly affected by changes in environmental conditions (humidity and temperature), giving support to hypothesis H1.

Humidity effects

The best frequency model IE suggests an optimum in flower visitor frequency when the relative humidity is 57.29%. For this humidity value, the expected visitor frequency is 0.46 bees per 6×6 plot sampling of 20 min (at 30.28°C, Figure 3.1). When decreasing or increasing the relative humidity by 10% (i.e. to 47.29 or 67.29%), the expected visitor frequency drops to 0.16, i.e. by ~65% (Figure 3.1). This is quite a dramatic change, especially when the relative humidity during sampling (handheld weather recorder) ranged from ~41 to ~83% (weather logger measurements ranged from ~32 to ~90%). Some studies have investigated the influence of relative humidity levels within honeybee hives (e.g. Human et al., 2006; Lindauer, 1955), but I could find none that have reported effects of the relative humidity on flower visitor abundance or activities in the field. In their Fifth Assessment Report (AR5), the IPCC (2014) predicted that regional precipitation patterns worldwide will change in contrasting ways during the 21st century. In particular, the already high inter-annual variability in precipitation caused by ENSO is predicted to intensify (IPCC, 2014). In the Argentinian Pampas, ENSO phases have already been shown to affect yield anomalies in soybean and other crops (Podestá et al., 1999). My results indicate that pollinator frequencies may also be affected by this phenomenon, as ambient humidity levels are highly correlated with precipitation frequency and intensity. Unfortunately, I could not consider inter-annual variability in my analyses, as my data were collected during one season only. It should also be noted that the particular season of data collection coincided with the strong 2015-16 El Niño event (National Weather Service), and that Argentina had an overall positive anomaly in summer precipitation of ~25% (Servicio Meteorológico Nacional).

Temperature effects

The IPCC (2014) predicts that the global temperatures will increase with 2 (low emission scenario) to 5.7°C (high emission scenario) within 2100. The most rapid increases are predicted to occur in arctic areas, while the temperature change in the Argentinian Pampas is predicted to be just below the global mean increase (IPCC, 2014). In my study, the flower visitor frequency increased exponentially within the temperature range observed during field samplings (~22-38°C, Figure 3.2). Even though samplings were carried out at temperatures as high as 38°C, and the average

daily maximum temperature was ~34°C during the fieldwork period[3], no optimal value for flower visitor frequency was found. This was unexpected, in particular because the study area's average daily maximum temperature is only 30°C in January (Servicio Meteorológico Nacional). The best frequency model IE thus suggests that the bees in the study area may not necessarily be affected negatively by elevating temperatures. However, I have only assessed the climatic effects on the frequency of flower-visiting bee *individuals*. The behaviour (e.g. visitation frequency) of the bees might show different responses to temperature (and humidity), affecting pollination in subtle ways. Optimal temperature values for bee visitation frequencies (i.e. number of *visits* per flower) have been observed in other parts of the world; Nielsen et al. (2017) found an optimum in honeybee visitation frequency to domestic raspberry (*Rubus idaeus* L.) at 24.1°C in Southeast Norway. Similarly, Rader et al. (2013) found honeybee visits to watermelon (*Citrullus lanatus* (Thunb.) Matsum. & Nakai) to be most frequent at ~27°C in the eastern part of the US. Furthermore, during a 5-days period of unusually warm weather in Lithuania, Blažytė-Čereškienė et al. (2010) found a negative correlation between temperature (27-45°C) and honeybee visitation to oilseed rape (*Brassica napus* L.). These findings strongly imply that the expected elevating temperatures of the future will negatively affect some bee populations and the ecosystem service they provide. However, as pointed out by Goulson et al. (2015), and briefly discussed here in the previous paragraph, climate change is not solely associated with warming. Extreme weather events including storms and floods may be more threatening to bees in some regions (Goulson et al., 2015). When it comes to the bees in the focal area of this study, increased changes in precipitation patterns may seem to represent a more severe threat than temperature elevations.

Recording of environmental variables

I used weather loggers and a handheld weather recorder to measure the environmental variables. The best model included logger-temperature and recorder-humidity. When I exchanged the logger-temperature variable with the recorder-temperature in the model, and did the opposite exchange with the humidity variable,

[3] Calculated within each study site with the values measured every hour by the weather loggers, and then averaged over the nine study sites.

the BIC value increased (i.e. resulting in poorer models). As the loggers were installed 25 cm above ground, they were measuring the conditions of the microenvironment within the crop canopy. Conversely, the values obtained from the handheld recorder were measured at chest height (~120 cm above ground), i.e. above the canopy. A possible explanation for the best humidity variable is therefore that the loggers were placed too close to the ground to reflect the bees' flight environment accurately. Ideally, the loggers should therefore have been placed higher above the ground. The most likely reason for the preferred temperature variable is that the recorder is more susceptible to sun exposure, possibly resulting in somewhat erroneous temperature measures.

4.2 Effects of different sampling methods

To assess how different sampling methods may influence the estimation of flower visitor frequencies and counts (Objective II), the data collection involved both transect walks and plot samplings with three different plot sizes.

Flower visitor frequency

The perceived flower visitor frequencies were significantly higher when using plot samplings compared to transect walks (Figures 3.1-3), supporting hypothesis H2-a. According to the best frequency model IE, the use of transect walks leads to a ~56% reduction in the perceived frequency ($e^{-0.824} \approx 0.44$, Table 3.2). The best frequency model EE implies an even stronger effect, with a ~70% reduction ($e^{-1.208} \approx 0.30$, Table 3.4). However, the frequency model EE was unable to explain more than 50% of the variation in the perceived flower visitor frequency (Table 3.5), most likely because it was based on a dataset without recordings of the environmental conditions. The strong effect of sampling method in the frequency model EE may imply that most of the additional sampling events added to the primary dataset (all transect walks) were conducted during relatively poor climatic conditions. Because of the strong effect that environmental conditions evidently have on the flower visitor frequency, the model IE is superior to the model EE, and it is reasonable to assume that the real reduction in the perceived frequency is closer to 56%. This is still a significant effect. The main reason for the reduction is that the flower exposure used

as an offset variable was highly overestimated for the transect walks. It is therefore reasonable to assume that plot samplings yield more correct estimates. Interestingly, the variable distinguishing between transect walks and plot samplings (2 levels) explained the variation in perceived flower visitor frequency better than 'unit area' (numeric variable describing the size of the sampling unit). This may imply that few of the flower visitors entering the sampling unit during the 20 min were missed within the three plot types covering different areas. It is, however, doubtful that this is true also for the 20×4 sampling unit. The lower sampling effort for this particular sampling unit (Table 2.1) may explain this outcome.

Flower visitor count

The model of the flower visitor count per unit time revealed that the transect walks performed better than the plot samplings when the goal is to observe as many flower visitors as possible (Figure 3.4). The model indicates that transect walks may yield $e^{1.06} \approx 2.9$ times as many observations per unit time as plot samplings (Table 3.6), giving support to hypothesis H2-b. Similarly, Gibson et al. (2011) concluded that using transect walks is a quicker way of collecting large quantities of data when assessing entire plant-pollinator communities. From their results, I calculated that the transect walks yielded ~4.4 ± 1.0 SE times as many observations per unit time as the plot samplings (see Appendix G.1 for calculations). In contrast to my study, Gibson et al. (2011) carried out the fieldwork in a meadow containing several plant species. They claim that, in their system, plot samplings are more likely to yield fewer observations per unit time "… because of the increased probability of encountering abundant and/or highly attractive plant species in transects and the reduced time spent at plant species with no visitors" (p. 830).

Nielsen et al. (2011) also conducted both plot samplings and transect walks in an agricultural system (olive groves), assessing the efficiency of different sampling methods when measuring bee species richness. They caught ~21 times as many individuals per unit time using transect walks as with plot samplings (see Appendix G.2 for calculations). This large increase might be due to the huge difference in area coverage, as the plots were only 1 m × 2 m (six plots per site, so 12 m^2 in total) and the transects were 250 m long and 4 m wide (1000 m^2). When including the area

covered in the calculation, plot samplings yielded ~2.2 times as many individuals per m^2 per unit time (Appendix G.2), which is also in line with the present study's results (Table 3.2). It should be noted that olive groves are, in contrast to soybean fields, holding a diverse plant community displaying a high diversity of flowers. The explanation by Gibson et al. (2011) as to why transect walks yield more observations than plot samplings thus may be valid for Nielsen et al. (2011) as well. Here, I have showed that transect walks yield more observations per unit time also in a more homogeneous system consisting of only one plant species. The reason for this is most likely linked to the behaviour of the flower visitors, as bees tend to stay in the same area for a while, visiting neighbouring flowers (M. Devoto, *pers. comm.*). Covering a larger area during each sampling event will therefore increase the probability of detecting flower visitors.

Even if my study system consisted of one plant species only, both *best models* included the random term 'field section' (spatial variation within the fields), suggesting that the system is not as homogeneous as one might think at first sight. This hidden heterogeneity is difficult to explain, but it may have to do with the landscape context, i.e. nearby hives or natural nesting areas. Variation on the landscape level, however, was not considered in my study, and could be an interesting variable to include in future studies.

Which is the best sampling method?

As noted in the introduction, and now demonstrated by the thesis' results, different methods for sampling pollination data have both benefits and drawbacks. Which is the best sampling method, depends on the particular objectives to be addressed. When it comes to the estimation of visitor or visitation frequencies, there is clearly a trade-off between obtaining accuracy and reducing sampling effort. Systems with low visitation frequencies have the additional difficulty that a longer observation period per sampling event is needed in order to obtain accurate frequency estimates (Fijen & Kleijn, 2017). In such cases, a possibility might be to define plots that cover a slightly or moderately larger area than what the observer is able to view simultaneously, if the system allows for movement within the sampling unit. When the visitor frequency is low, it should be easier to detect all the flower visitors

entering the plot area, as compared to when the frequency is high. This may be particularly true if the main flower visitors are bees; one of the experiences I gained during the fieldwork was that it was often easier to detect flower-visiting bees by hearing rather than by sight. This perception was probably related to the dense formation of the crop foliage, which made it difficult to view all the small soybean flowers within the crop canopy.

If obtaining a large quantity of observations is more important than to get accurate frequency estimates, my results imply that transect walks is a better approach than plot samplings. Obtaining large sample sizes at the cost of accuracy may be useful when assessing pollinator species richness or, like in the present study, when assessing how different factors, such as climatic conditions, are affecting flower visitor populations. Obtaining many observations may also be advantageous when constructing plant-pollinator networks, although the network structure itself may be influenced by the sampling method used; Gibson et al. (2011) found that networks constructed using plot samplings had a significantly higher number of unique interactions than those constructed using transect walks.

4.3 Insect pollination in a uniform landscape

My impression in the field was that the number of insects visiting soybean flowers in the study area was extremely low. In order to make a quantification of the flower visitation frequency (Objective III), I calculated the probability that a soybean flower would be visited by a bee at least once while being open. Both the model and raw data estimates suggested that this probability was less than 6%, and that its upper 95% bootstrap confidence limit was less than 9%. For comparison, even if to a very different system, Nielsen et al. (2017) estimated that each raspberry flower surveyed in their study received, on average, ~60 visits per day. Assuming that a flower's life span is 2.5 days (as done in Sáez et al., 2014), it received ~150 visits during the time it stayed open. This means that a raspberry flower in Nielsen et al.'s study system on average received ~2,500 times more visits than a soybean flower in my study system ($\frac{150}{0.06}$ = 2,500). This comparison gives support to my third hypothesis (H3). However, when comparing my findings to previously observed visitation frequencies to soybean (Monasterolo et al., 2015), the outcome is different. Using Monasterolo et al.'s results

to calculate a corresponding flower visitation probability (Appendix G.3), I found that there was only a ~1.58% chance for a flower in their system to receive a visit while being open.

Monasterolo et al. (2015) conducted their study in the Chaco Serrano district, part of the dry forest biome, further north in Argentina. The natural vegetation of Chaco Serrano is fragmented within intensively managed agricultural areas. Still, patches of native vegetation remain in a higher degree than in the humid Pampas; the nine study sites of Monasterolo et al. (2015) were comprised of ~33% natural vegetation cover, while the semi-natural vegetation near my study sites was restricted to linear patches between the crop fields and the road. Furthermore, when the soybean cultivar in my study sites was sprayed with glyphosate, the non-cultivated plants closest to the field margins were killed, leaving behind very little alternative food sources for the flower-visiting insects in the area. Because of the difference in the amount of natural vegetation surrounding the soybean fields in the two studies, I expected the flower visitation probability to be lower in my study as compared to in Monasterolo et al.'s study, but this was not the case. There are, however, several possible explanations to the difference observed between the two study systems. Firstly, both studies were conducted during one season only, and in different years. The different results may thus be related to inter-annual climatic variation. Secondly, the climate is considerably dryer in the Chaco Serrano than in the humid Pampas. Therefore, the soybean varieties are most likely quite different, and their attractiveness to insects may differ as well. A third possibility is that the differences in natural vegetation cover between the two areas had the opposite effect of what I expected; López-Carretero et al. (2017) assessed plant-pollinator networks near San Claudio (my mid location) during the summers of 2015-16. They found that the pollinator species in the area preferred to forage on the non-cultivated plants in the field margins, and that the soybean flowers seemed to work only as an alternative resource. Their result may suggest that soybean flowers (at least those in their study area) are not particularly attractive to pollinators, and that the flower visitation probability was higher in my study because the bees had no other forage alternatives.

Regardless of the divergence between the flower visitation probability observed in my study compared to that of Monasterolo et al. (2015), the overall low estimates from both studies is quite remarkable and still in support of hypothesis H3.

Potential for increasing soybean yields by improving pollination?

One of the motivations for this thesis was the potential for augmenting soybean yields by facilitating for increased insect pollination. The degree to which soybean production will benefit from this, however, is still uncertain (Gazzoni, 2016). Some authors claim to have shown increases in soybean yields ($\frac{kg}{ha}$) corresponding to 18.9 (Milfont et al., 2013) or 57.73% (Chiari et al., 2005b) when the flowers were available to pollinators as compared to when they were not during enclosure experiments. Both of these studies' experimental design involved introducing honeybee hives to the system, but neither assessed the visitor or visitation frequency. It is thus not evident that the yield increases were accounted for by flower visitors. Instead, methodological issues, such as shading by the enclosure structures, could just as well be the reason for the observed differences in yield. To my knowledge, the only researchers that have carried out enclosure experiments on soybean, while simultaneously recording the visitation frequency, are Monasterolo et al. (2015). Their experiment involved marking a number of floral buds and enclosing half of them in voile bags to exclude flower visitors. The bags stayed on for a month during flowering. The open treatment yielded significant increases in fruit weight, seed weight and reproductive success[4], and a reduction of ~20% in aborted flowers. It is, however, doubtful that flower-visiting bees were the factor accounting for the different outcomes of the treatments, as the flower visitation probability was less than 2%, i.e. each flower received, on average, 0.02 visits during the time they were open. It may be more plausible that the enclosure structures they used (i.e. the voile bags) had a negative effect on the reproductive output; they may have caused damage to the buds or flowers, or affected the microenvironment within the bags (e.g. increased shading and altered humidity levels).

[4] Estimated as $\frac{\#fruits}{\#flowers} \times \frac{\text{mean}(\#seeds\ per\ fruit)}{\text{mean}(\#ovules\ per\ flower)}$

4.4 Summary

The overall aim of this thesis was to assess the flower visitor and visitation frequencies to GM soybean in an intensively managed agricultural system, and to identify the feasibility of different methods when sampling flower visitor data. I found that environmental conditions had a significant impact on the frequency of soybean flower visitors in the area. My findings imply that the predicted changes in precipitation patterns may represent a severe threat to the pollinator fauna in the Argentinian Pampas. Furthermore, I found that plot samplings, rather than transect walks, should be used when the aim is to obtain accurate visitor or visitation frequency estimates. On the other hand, if the aim is to detect as many flower visitors or visits as possible, transect walks is a better approach, also in more homogeneous systems consisting of one plant species only. Finally, I found that the flower visitation frequency to soybean in the humid Pampas was less than 0.06 visits per flower during its life span, which is extremely low compared to other systems. This observation may be explained by the combination that the surroundings represent a poor insect habitat (and thus inhabit few pollinators), and that the soybean flowers are not particularly attractive to bees.

4.5 Future steps

My results indicate that the bees in the humid Pampas are affected by changing climatic conditions within a season, but I was not able to assess the effects of climatic variation among years. As the inter-annual variability of precipitation in the Argentinian Pampas is high, and is predicted to increase even further (IPCC, 2014), this could be a relevant objective for future research. In such studies, it will be beneficial to obtain as many observations as possible, as this will likely facilitate the detection of contrasts in temperature or humidity responses. Therefore, according to my findings, transect walks should be used.

My results, and the estimate obtained from Monasterolo et al. (2015), also imply that under natural conditions, soybean flowers in modern agricultural systems are receiving extremely few visits. The studies that have observed yield increases when introducing managed honeybees to their focal soybean fields have not assessed the

visitation frequencies (Chiari et al., 2005b; Milfont et al., 2013). More research is therefore needed to evaluate to what degree promoting insect pollination will improve soybean production, and how this in turn may facilitate the conservation of natural habitats threatened by field expansions. Such future studies should control for methodological effects (e.g. shading from enclosure structures) and assess visitation frequencies to verify that the 'open treatment' flowers are visited sufficiently for the bees to be the factor accounting for possible yield increases. According to my findings, plot samplings should be used to obtain accurate visitation frequency estimates. Furthermore, the hidden heterogeneity within soybean fields should be accounted for; a possibility is to reduce the number of repeated sampling events per 'field section' while increasing the number of sections distributed within each study field. Recordings of the temperature and relative humidity should also be included in the analyses to strengthen the reliability of the visitation frequency estimate.

REFERENCES

Ausden, M. & Drake, M. (2006). Invertebrates. In W. J. Sutherland (Ed.), *Ecological census techniques: a handbook* (2 ed., pp. 214-249): Cambridge University Press.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Biesmeijer, J. C., Roberts, S. P., Reemer, M., Ohlemüller, R., Edwards, M., Peeters, T., Schaffers, A., Potts, S. G., Kleukers, R., Thomas, C., Settele, J. & Kunin, W. E. (2006). Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. *Science, 313*(5785), 351-354.

Blažytė-Čereškienė, L., Vaitkevičienė, G., Venskutonytė, S. & Būda, V. (2010). Honey bee foraging in spring oilseed rape crops under high ambient temperature conditions. *Zemdirbyste-Agriculture, 97*(1), 61-70.

Blettler, D. C., Fagúndes, G. A. & Caviglia, O. P. (2017). Contribution of honeybees to soybean yield. *Apidologie*, 1-11.

Bohart, G. E. (1960). Insect pollination of forage legumes. *Bee World, 41*(4), 85-97.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H. & White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution, 24*(3), 127-135.

Breeze, T. D., Bailey, A. P., Balcombe, K. G. & Potts, S. G. (2011). Pollination services in the UK: How important are honeybees? *Agriculture, Ecosystems and Environment, 142*(3), 137-143.

Brown, M. J. & Paxton, R. J. (2009). The conservation of bees: a global perspective. *Apidologie, 40*(3), 410-416.

Burkart, R., Bárbaro, N. O., Sánchez, R. O. & Gómez, D. A. (1999). *Eco-regiones de la Argentina*: Presidencia de la Nación-Secretaría de Recursos Naturales y Desarrollo Sustentable.

Burton, J. W. (1997). Soyabean (*Glycine max* (L.) Merr.). *Field Crops Research, 53*(1), 171-186.

Cane, J. H., Minckley, R. L. & Kervin, L. J. (2000). Sampling bees (Hymenoptera: Apiformes) for pollinator community studies: pitfalls of pan-trapping. *Journal of the Kansas Entomological Society*, 225-231.

Chiari, W. C., Toledo, V., Ruvolo-Takasusuki, M. C. C., Attencia, V. M., Costa, F. M., Kotaka, C. S., Sakaguti, E. S. & Magalhães, H. R. (2005a). Floral biology and behavior of Africanized honeybees *Apis mellifera* in soybean (*Glycine max* L. Merril). *Brazilian Archives of Biology and Technology, 48*(3), 367-378.

Chiari, W. C., Toledo, V. d. A. A. d., Ruvolo-Takasusuki, M. C. C., Oliveira, A. J. B. d., Sakaguti, E. S., Attencia, V. M., Costa, F. M. & Mitsui, M. H. (2005b). Pollination of soybean (*Glycine max* L. Merril) by honeybees (*Apis mellifera* L.). *Brazilian Archives of Biology and Technology, 48*(1), 31-36.

Clark, J. D., Beyene, Y., WoldeGabriel, G., Hart, W. K., Renne, P. R., Gilbert, H., Defleur, A., Suwa, G., Katoh, S., Ludwig, K. R., Bolsserle, J.-R., Asfaw, B. & White, T. D. (2003). Stratigraphic, chronological and behavioural contexts of Pleistocene *Homo sapiens* from Middle Awash, Ethiopia. *Nature, 423*(6941), 747-752.

de Jong, P. & Heller, G. Z. (2008). *Generalized linear models for insurance data* (Vol. 10): Cambridge University Press.

Dellafiore, C. (a). Southern South America: Southeastern Argentina. Retrieved from https://www.worldwildlife.org/ecoregions/nt0806

Dellafiore, C. (b). Southern South America: Eastern Argentina. Retrieved from https://www.worldwildlife.org/ecoregions/nt0803

Dormann, C. F., Schweiger, O., Arens, P., Augenstein, I., Aviron, S., Bailey, D., Baudry, J., Billeter, R., Bugter, R., Bukácek, R., Burel, F., Cerny, M., de Cock, R., de Blust, G., deFilippi, R., Diekötter, T., Dirksen, J., Durka, W., Edwards, P. J., Frenzel, M., Hamersky, R., Hendrickx, F., Herzog, F., Klotz, S., Koolstra, B., Lausch, A., le Coeur, D., Liira, J., Maelfait, J. P., Opdam, P., Roubalova, M., Schermann-Legionnet, A., Schermann, N., Scmidt, T., Smulders, M. J. M., Speelmans, M., Simova, P., Verboom, J., van Wingerden, W. & Zobel, M. (2008). Prediction uncertainty of environmental change effects on temperate European biodiversity. *Ecology Letters, 11*(3), 235-244.

Duke, S. O. & Powles, S. B. (2008). Glyphosate: a once-in-a-century herbicide. *Pest Management Science, 64*(4), 319-325. doi:10.1002/ps.1518

Efron, B. & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 54-75.

Erickson, E. & Garment, M. (1979). Soya-bean flowers: nectary ultrastructure, nectar guides, and orientation on the flower by foraging honeybees. *Journal of Apicultural Research, 18*(1), 3-11.

Fearnside, P. M. (2001). Soybean cultivation as a threat to the environment in Brazil. *Environmental Conservation, 28*(1), 23-38.

Fijen, T. P. & Kleijn, D. (2017). How to efficiently obtain accurate estimates of flower visitation rates by pollinators. *Basic and Applied Ecology, 19*, 11-18.

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., Mueller, N. D., O'Connell, C., Ray, D. K., West, P. C., Balzer, C., Bennett, E. M., Carpenter, S. R., Hill, J., Monfreda, C., Polasky, S., Rockström, J., Sheehan, J., Siebert, S., Tilman, D. & Zaks, D. P. M. (2011). Solutions for a cultivated planet. *Nature, 478*(7369), 337-342.

Food and Agriculture Organization of the United Nations (FAO). (2016). FAOSTAT Database. Retrieved 15 October 2017 http://www.fao.org/faostat/en/ - data/QC

Freitas, B. M., Imperatriz-Fonseca, V. L., Medina, L. M., Kleinert, A. d. M. P., Galetto, L., Nates-Parra, G. & Quezada-Euán, J. J. G. (2009). Diversity, threats and conservation of native bees in the Neotropics. *Apidologie, 40*(3), 332-346.

Fry, J., Barnason, A. & Horsch, R. B. (1987). Transformation of *Brassica napus* with *Agrobacterium tumefaciens* based vectors. *Plant Cell Reports, 6*(5), 321-325.

Garibaldi, L. A., Steffan-Dewenter, I., Winfree, R., Aizen, M. A., Bommarco, R., Cunningham, S. A., Kremen, C., Carvalheiro, L. G., Harder, L. D., Afik, O., Bartomeus, I., Benjamin, F., Boreux, V., Cariveau, D., Chacoff, N., Dudenhöffer, J., Freitas, B. M., Ghazoul, J., Greenleaf, S., Hipólito, J., Holzschuh, A., Howlett, B., Isaacs, R., Javorek, S., Kennedy, C., Krewenka, K., Krishnan, S., Mandelik, Y., Mayfield, M., Motzke, I., Munyuli, T., Nault, B., Otieno, M., Petersen, J., Pisanty, G., Potts, S. G., Rader, R., Ricketts, T., Rundlöf, M., Seymour, C., Schüepp, C., Szentgyörgyi, H., Taki, H., Tscharntke, T., Vergara, C., Viana, B., Wanger, T., Westphal, C., Williams, N. & Klein, A. M. (2013). Wild pollinators enhance fruit set of crops regardless of honey bee abundance. *Science, 339*(6127), 1608-1611.

Gazzoni, D. L. (2016). *Soybean and bees*. Brazil: Embrapa Soja.

Ge, L., Yu, J., Wang, H., Luth, D., Bai, G., Wang, K. & Chen, R. (2016). Increasing seed size and quality by manipulating BIG SEEDS1 in legume species. *Proceedings of the National Academy of Sciences*, 201611763.

Gibson, R. H., Knott, B., Eberlein, T. & Memmott, J. (2011). Sampling method influences the structure of plant–pollinator networks. *Oikos, 120*(6), 822-831.

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M. & Toulmin, C. (2010). Food security: the challenge of feeding 9 billion people. *Science, 327*(5967), 812-818.

Goulson, D., Lye, G. C. & Darvill, B. (2008). Decline and conservation of bumble bees. *Annual Review of Entomology, 53*, 191-208.

Goulson, D., Nicholls, E., Botías, C. & Rotheray, E. L. (2015). Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. *Science, 347*(6229), 1255957.

Grau, H. R., Aide, T. M. & Gasparri, N. I. (2005). Globalization and soybean expansion into semiarid ecosystems of Argentina. *Ambio, 34*(3), 265.

Greenwood, J. J. D. & Robinson, R. A. (2006). General census methods. In W. J. Sutherland (Ed.), *Ecological census techniques: a handbook* (2 ed., pp. 87-185): Cambridge University Press.

Hegland, S. J., Nielsen, A., Lázaro, A., Bjerknes, A. L. & Totland, Ø. (2009). How does climate warming affect plant-pollinator interactions? *Ecology Letters, 12*(2), 184-195.

Hickling, R., Roy, D. B., Hill, J. K., Fox, R. & Thomas, C. D. (2006). The distributions of a wide range of taxonomic groups are expanding polewards. *Global Change Biology, 12*(3), 450-455.

HighQuest Partners & Soyatech. (2008). *How the global oilseed and grain trade works*. Retrieved from the U.S. Soybean Export Council: https://ussec.org/resources/how-the-global-oilseed-and-grain-trade-works/

Hinchee, M. A., Connor-Ward, D. V., Newell, C. A., McDonnell, R. E., Sato, S. J., Gasser, C. S., Fischhoff, D. A., Re, D. B., Fraley, R. T. & Horsch, R. B. (1988). Production of transgenic soybean plants using *Agrobacterium*-mediated DNA transfer. *Nature Biotechnology, 6*(8), 915-922.

Human, H., Nicolson, S. W. & Dietemann, V. (2006). Do honeybees, *Apis mellifera scutellata*, regulate humidity in their nest? *Naturwissenschaften, 93*(8), 397-401.

Hymowitz, T. (1970). On the domestication of the soybean. *Economic Botany, 24*(4), 408-421.

Intergovernmental Panel on Climate Change (IPCC). (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (C. W. Team, R. Pachauri, & L. Meyer Eds.). IPCC, Geneva, Switzerland.

International Service for the Acquisition of Agri-biotech Applications (ISAAA). (2016). Global status of commercialized biotech/GM crops: 2016 *ISAAA Brief* (Vol. 52). Ithaca, New York: ISAAA.

Jaycox, E. (1970). Ecological relationships between honey bees and soybeans. II. The plant factors. *American Bee Journal*.

Kaimowitz, D. & Smith, J. (2001). Soybean technology and the loss of natural vegetation in Brazil and Bolivia. In A. Angelsen & D. Kaimowitz (Eds.), *Agricultural technologies and tropical deforestation* (pp. 195-211). Wallingford, UK: CAB International.

Kearns, C. A. & Inouye, D. W. (1993). *Techniques for pollination biologists*: University Press of Colorado.

Keil, P., Biesmeijer, J. C., Barendregt, A., Reemer, M. & Kunin, W. E. (2011). Biodiversity change is scale-dependent: an example from Dutch and UK hoverflies (Diptera, Syrphidae). *Ecography, 34*(3), 392-401.

Kerr, J. T., Pindar, A., Galpern, P., Packer, L., Potts, S. G., Roberts, S. M., Rasmont, P., Schweiger, O., Colla, S. R., Richardson, L. L., Wagner, D., Gall, L., Sikes, D. & Pantoja, A. (2015). Climate change impacts on bumblebees converge across continents. *Science, 349*(6244), 177-180.

Kjøhl, M., Nielsen, A. & Stenseth, N. C. (2011). *Potential effects of climate change on crop pollination*. Rome: Food and Agriculture Organization of the United Nations (FAO).

Klein, A. M., Vaissiere, B. E., Cane, J. H., Steffan-Dewenter, I., Cunningham, S. A., Kremen, C. & Tscharntke, T. (2007). Importance of pollinators in changing landscapes for world crops. *Proceedings of the Royal Society B: Biological Sciences, 274*(1608), 303-313.

Komatsu, S., Tougou, M. & Nanjo, Y. (2015). Proteomic techniques and management of flooding tolerance in soybean. *Journal of Proteome Research, 14*(9), 3768-3778.

Lindauer, M. (1955). The water economy and temperature regulation of the honeybee colony. *Bee World, 36*(5), 81-92.

López-Carretero, A., Monasterolo, M., Chamer, M., Medan, D. & Devoto, M. (2017). *Mass flowering of soybean has little effect on plant-pollinator networks in field margins*. Manuscript in preparation.

Masuda, T. & Goldsmith, P. D. (2009). World soybean production: area harvested, yield, and long-term projections. *International Food and Agribusiness Management Review, 12*(4), 143-162.

McCormick, S., Niedermeyer, J., Fry, J., Barnason, A., Horsch, R. & Fraley, R. (1986). Leaf disc transformation of cultivated tomato (*L. esculentum*) using *Agrobacterium tumefaciens*. *Plant Cell Reports, 5*(2), 81-84.

McDougall, I., Brown, F. H. & Fleagle, J. G. (2005). Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature, 433*(7027), 733-736.

McGregor, S. E. (1976). *Insect pollination of cultivated crop plants*. Washington D.C.: Agricultural Research Service, US Department of Agriculture.

Medan, D., Torretta, J. P., Hodara, K., Elba, B. & Montaldo, N. H. (2011). Effects of agriculture expansion and intensification on the vertebrate and invertebrate diversity in the Pampas of Argentina. *Biodiversity and Conservation, 20*(13), 3077-3100.

Messina, C., Hansen, J. & Hall, A. (1999). Land allocation conditioned on El Niño-Southern Oscillation phases in the Pampas of Argentina. *Agricultural Systems, 60*(3), 197-212.

Milfont, M. d. O., Rocha, E. E. M., Lima, A. O. N. & Freitas, B. M. (2013). Higher soybean production using honeybee and wild pollinators, a sustainable alternative to pesticides and autopollination. *Environmental chemistry letters, 11*(4), 335-341.

Monasterolo, M., Musicante, M. L., Valladares, G. R. & Salvo, A. (2015). Soybean crops may benefit from forest pollinators. *Agriculture, Ecosystems and Environment, 202*, 217-222.

Musicante, M. (2013). *Efectos de la fragmentación del hábitat sobre himenópteros antófilos (Insecta) en el Bosque Chaqueño Serrano*. (Thesis), Universidad Nacional de Córdoba.

Nabhan, G. P. & Buchmann, S. L. (1997). Services provided by pollinators. In G. Daily (Ed.), *Nature's Services: societal dependence on natural ecosystems* (pp. 133-150): Island Press.

National Research Council. (2007). Status of pollinators. In K. Kelly & P. Whitacre (Eds.), *Status of pollinators in North America* (pp. 34-74): National Academies Press.

National Weather Service. Cold & warm episodes by season. Retrieved from http://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php

Nielsen, A., Reitan, T., Rinvoll, A. W. & Brysting, A. K. (2017). Effects of competition and climate on a crop pollinator community. *Agriculture, Ecosystems & Environment, 246*, 253-260.

Nielsen, A., Steffan-Dewenter, I., Westphal, C., Messinger, O., Potts, S. G., Roberts, S. P., Settele, J., Szentgyörgyi, H., Vaissière, B. E., Vaitis, M., Woyciechowski, M., Bazos, I., Biesmeijer, J. C., Bommarco, R., Kunin, W. E., Tscheulin, T., Lamborn, E. & Petanidou, T. (2011). Assessing bee species richness in two Mediterranean communities: importance of habitat type and sampling techniques. *Ecological Research, 26*(5), 969-983.

Observatory of Economic Complexity. (2016). Argentina. Retrieved from https://atlas.media.mit.edu/en/profile/country/arg/

Ollerton, J., Winfree, R. & Tarrant, S. (2011). How many flowering plants are pollinated by animals? *Oikos, 120*(3), 321-326.

Omacini, M., Chaneton, E. J., León, R. J. C. & Batista, W. B. (1995). Old-field successional dynamics on the Inland Pampa, Argentina. *Journal of Vegetation Science, 6*(3), 309-316.

Palmer, M., Bernhardt, E., Chornesky, E., Collins, S., Dobson, A., Duke, C., Gold, B., Jacobson, R., Kingsland, S., Kranz, R., Mappin, M., Martinez, M. L., Micheli, F., Morse, J., Pace, M., Pascual, M., Palumbi, S., Reichman, O. J., Simons, A., Townsend, A. & Turner, M. (2004). Ecology for a crowded planet. *Science, 304*(5675), 1251-1252.

Palmer, R. G., Perez, P., Ortiz-Perez, E., Maalouf, F. & Suso, M. J. (2009). The role of crop-pollinator relationships in breeding for pollinator-friendly legumes: from a breeding perspective. *Euphytica, 170*(1), 35-52.

Parmesan, C., Ryrholm, N., Stefanescu, C., Hill, J. K., Thomas, C. D., Descimon, H., Huntley, B., Kaila, L., Kullberg, J., Tammaru, T., Tennent, W. J., Thomas, J. A. & Warren, M. (1999). Poleward shifts in geographical ranges of butterfly species associated with regional warming. *Nature, 399*(6736), 579-583.

Phipps, R. H. & Park, J. (2002). Environmental benefits of genetically modified crops: global and European perspectives on their ability to reduce pesticide use. *Journal of Animal and Feed Sciences, 11*(1), 1-18.

Podestá, G. P., Messina, C. D., Grondona, M. O. & Magrin, G. O. (1999). Associations between grain crop yields in central-eastern Argentina and El Niño–Southern Oscillation. *Journal of Applied Meteorology, 38*(10), 1488-1498.

Popic, T. J., Davila, Y. C. & Wardle, G. M. (2013). Evaluation of common methods for sampling invertebrate pollinator assemblages: net sampling out-perform pan traps. *PloS one, 8*(6), e66665.

Potts, S. G., Biesmeijer, J. C., Kremen, C., Neumann, P., Schweiger, O. & Kunin, W. E. (2010a). Global pollinator declines: trends, impacts and drivers. *Trends in Ecology & Evolution, 25*(6), 345-353.

Potts, S. G., Roberts, S. P., Dean, R., Marris, G., Brown, M. A., Jones, R., Neumann, P. & Settele, J. (2010b). Declines of managed honey bees and beekeepers in Europe. *Journal of Apicultural Research, 49*(1), 15-22.

Probst, A. & Judd, R. (1973). Origin, U.S. history and development, and world distribution. In B. E. Caldwell, R. W. Howell, & H. W. Johnson (Eds.), *Soybeans: Improvement, Production, and Uses* (pp. 1-15): American Society of Agronomy, Inc.

Qaim, M. (2009). The economics of genetically modified crops. *Annual Review of Resource Economics, 1*(1), 665-694.

Qaim, M. & Zilberman, D. (2003). Yield effects of genetically modified crops in developing countries. *Science, 299*(5608), 900-902.

R Core Team. (2016). R: A language and environment for statistical computing (Version 3.3.2): R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.r-project.org/

Rader, R., Reilly, J., Bartomeus, I. & Winfree, R. (2013). Native bees buffer the negative impact of climate warming on honey bee pollination of watermelon crops. *Global Change Biology, 19*(10), 3103-3110. doi:10.1111/gcb.12264

Reitan, T. & Nielsen, A. (2016). Do not divide count data with count data; a story from pollination ecology with implications beyond. *PloS one, 11*(2), e0149129.

Rindes y Cultivos DAS. (2012). Soja DM 3070. Retrieved from http://www.rycdas.com.ar/es/productos/305/sojadm3070.html

RStudio Inc. (2016). R Markdown. Retrieved from http://rmarkdown.rstudio.com/

Rust, R. W., Mason, C. E. & Erickson, E. H. (1980). Wild bees on soybeans, *Glycine max*. *Environmental Entomology, 9*(2), 230-232.

Sáez, A., Morales, C. L., Ramos, L. Y. & Aizen, M. A. (2014). Extremely frequent bee visits increase pollen deposition but reduce drupelet set in raspberry. *Journal of Applied Ecology, 51*(6), 1603-1612.

Santos, E., Mendoza, Y., Vera, M., Carrasco-Letelier, L., Díaz, S. & Invernizzi, C. (2013). Aumento en la producción de semillas de soja (*Glycine max*) empleando abejas melíferas (*Apis mellifera*). *Agrociencia Uruguay, 17*(1), 81-90.

Schweiger, O., Settele, J., Kudrna, O., Klotz, S. & Kühn, I. (2008). Climate change can cause spatial mismatch of trophically interacting species. *Ecology, 89*(12), 3472-3479.

Servicio Meteorológico Nacional. Clima en Argentina. Retrieved from https://www.smn.gob.ar/caracterización-estad%C3%ADsticas-de-largo-plazo

Settele, J., Kudrna, O., Harpke, A., Kühn, I., Van Swaay, C., Verovnik, R., Warren, M. S., Wiemers, M., Hanspach, J., Hickler, T., Kühn, E., van Halder, I., Veling, K., Vliegenthart, A., Wynhoff, I. & Schweiger, O. (2008). *Climatic risk atlas of European butterflies*: Pensoft

Severson, D. & Erickson Jr., E. (1984). Quantitative and qualitative variation in floral nectar of soybean cultivars in southeastern Missouri. *Environmental Entomology, 13*(4), 1091-1096.

Sinclair, T. R., Purcell, L. C., King, C. A., Sneller, C. H., Chen, P. & Vadez, V. (2007). Drought tolerance and yield increase of soybean resulting from improved symbiotic N_2 fixation. *Field Crops Research, 101*(1), 68-71.

Sinclair, T. R., Purcell, L. C. & Sneller, C. H. (2004). Crop transformation and the challenge to increase yield potential. *Trends in Plant Science, 9*(2), 70-75.

Southwood, T. R. E. & Henderson, P. A. (2000). Absolute population estimates by sampling a unit of habitat: air, plants, plant products, and vertebrate hosts *Ecological methods* (3 ed.): Blackwell Science Ltd.

Tilman, D. (1999). Global environmental impacts of agricultural expansion: The need for sustainable and efficient practices. *Proceedings of the National Academy of Sciences, 96*(11), 5995-6000.

Umbeck, P., Johnson, G., Barton, K. & Swain, W. (1987). Genetically transformed cotton (*Gossypium hirsutum* L.) plants. *Nature Biotechnology, 5*(3), 263-266.

United Nations (UN), Department of Economic and Social Affairs, Population Division. (2017). *World population prospects: The 2017 revision, key findings and advance tables* (Working Paper No. ESA/P/WP/248). Retrieved from https://esa.un.org/unpd/wpp/Publications/Files/WPP2017_KeyFindings.pdf

University of Oslo (UiO). (2012, 13 March 2017). About Abel. Retrieved from http://www.uio.no/english/services/it/research/hpc/abel/more/

Valliyodan, B., Ye, H., Song, L., Murphy, M., Shannon, J. G. & Nguyen, H. T. (2016). Genetic diversity and genomic strategies for improving drought and waterlogging tolerance in soybeans. *Journal of Experimental Botany, 68*(8), 1835-1849.

vanEngelsdorp, D., Hayes Jr, J., Underwood, R. M. & Pettis, J. (2008). A survey of honey bee colony losses in the U.S., fall 2007 to spring 2008. *PloS one, 3*(12), e4071.

Westphal, C., Bommarco, R., Carré, G., Lamborn, E., Morison, N., Petanidou, T., Potts, S. G., Roberts, S. P., Szentgyörgyi, H., Tscheulin, T., Vaissière, B., Woyciechowski, M., Biesmeijer, J. C., Kunin, W. E., Settele, J. & Steffan-Dewenter, I. (2008). Measuring bee diversity in different European habitats and biogeographical regions. *Ecological Monographs, 78*(4), 653-671.

Williams, L. F. (1950). Structure and genetic characteristics of the soybean. *Soybeans and Soybean Products*, 111-134.

Willmer, P. (2011). *Pollination and floral ecology*: Princeton University Press.

Wilson, J. S., Griswold, T. & Messinger, O. J. (2008). Sampling bee communities (Hymenoptera: Apiformes) in a desert landscape: are pan traps sufficient? *Journal of the Kansas Entomological Society, 81*(3), 288-300.

Zuur, A. F., Ieno, E. N., Walker, N., Saveliev, A. A. & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R*: Springer-Verlag New York.

APPENDICES

A Crop morphology

This appendix shows the morphological difference between the soybean crop in the different study locations (Figure A.1). The plants in the south location differed significantly in morphology from those in the other two locations, and were probably of a different variety. As the crop in the S1 and S2 sites did not flower until about halfway into the fieldwork period, a possibility is that the plants in south were of the second soybean cohort.

Figure A.1: Representative pictures of the crop morphology in the three study locations. The crop formed a dense scenery and grew quite tall in **a)** north and **b)** mid, while in **c)** south, the crop grew much shorter and was therefore also less dense.

B Datasets

This appendix provides an overview of the three datasets used in this thesis.

The primary dataset

The primary dataset included 168 sampling events and 31 variables (Table B.1).

Table B.1 (continues to the next page): Overview of the variables included in the primary dataset.

Variable	Summary			
Cos1: Numeric variable in relation to daily rhythm variation	Min: -1.00 Median: -0.57		Mean: 0.50 Median: 0.38	
Cos2: Numeric variable in relation to daily rhythm variation	Min: -1.00 Median: -0.35		Mean: -0.21 Max: 1.00	
Day: Numeric variable describing the number of days after fieldwork initiation	Min: 0.00 Median: 13.50		Mean: 11.34 Max: 18.00	
Day factor: Day as factor variable, 18 levels	D6: 9 D7: 9 D8: 4 D10: 6 D11: 4	D12: 8 D13: 6 D14: 8 D15: 4 D16: 6	D17: 6 D18: 10 D19: 4 D20: 6	D21: 18 D22: 36 D23: 6 D24: 18
Distance from field border: Numeric variable (m^2)	Min: 30.00 Median: 50.00		Mean: 63.27 Max: 120.00	
Distance from field border factor: Factor variable, 5 levels	30: 6 50: 115 80: 6		100: 35 120: 6	
Field section: Factor variable, 26 levels	M1-S1: 3 M1-S2: 5 M1-S3: 3 M1-S4: 5 M1-S5: 3 M3-S1: 3 M3-S2: 4	M3-S3: 3 M3-S4: 4 M3-S5: 3 M4-S1: 24 M4-S2: 27 M4-S3: 24 M4-S4: 3	N1-S2: 5 N1-S4: 5 N2-S2: 5 N2-S4: 5 N3-S2: 9 N3-S4: 1	S1-S2: 3 S1-S4: 3 S2-S2:3 S2-S4: 3 S3-S2: 6 S3-S4: 6
Flower density: Numerical variable, the flower density of the sampling unit (m^{-1})	Min: 13.2 Median: 267.6		Mean: 356.2 Max: 918.4	
Total number of flowers: The number of flowers within the sampling unit	Min: 475.2 Median: 13,800.0		Mean: 24,900.0 Max: 108,000.0	
Focus: Factor variable distinguishing between sampling of environmental variation and different sampling methods, two levels	Environmental variation: 78 Different sampling methods: 90			

Table B.1 (continued): Overview of the variables included in the primary dataset.

Humidity (loggers): The relative humidity (%) measured by the loggers	Min: 40.60 Median: 60.30		Mean: 60.18 Max: 83.00
Humidity (recorder): The relative humidity (%) measured by the recorder	Min: 31.97 Median: 61.11		Mean: 61.71 Max: 90.17
Latitude: The latitude where the sampling event took place in decimals	Min: -36.94 Median: -36.04		Mean: -36.02 Max: -35.26
Observed bees: The number of observed bees	Min: 0.00 Median: 0.00		Mean: 0.40 Max: 5.00
Observed hoverflies: The number of observed hoverflies	Min: 0.00 Median: 0.00		Mean: 0.04 Max: 3.00
Observer: ID of the person carrying out the sampling event	O1: 55 O2: 57		O3: 56
Sampling method: Factor variable, two levels	Plots: 138 Transects: 30		
Sampling unit: Factor variable, four levels.	100×2: 30 4×4: 24		6×6: 98 20×4: 16
Sin1: Numeric variable in relation to daily rhythm variation	Min: -1.00 Median: -0.82		Mean: -0.41 Max: 0.79
Sin2: Numeric variable in relation to daily rhythm variation	Min: -1.00 Median: 0.05		Mean: 0.03 Max: 1.00
Study location: Factor variable, three levels	North: 30 Mid: 114		South: 24
Study site: Factor variable, nine levels	N1: 10 N2: 10 N3: 10	M1: 19 M2: 17 M3: 78	S1: 6 S2: 6 S3: 12
Temperature (loggers): The temperature (°C) measured by the loggers	Min: 21.55 Median: 30.66		Mean: 30.28 Max: 38.17
Temperature (recorder): The temperature (°C) measured by the recorder	Min: 20.90 Median: 29.40		Mean: 29.02 Max: 35.30
Time factor: Factor variable describing the time of the sampling event	143 levels		
Time of day: Numeric variable describing the time of day when the sampling event took place	Min: 0.36 Median: 0.65		Mean: 0.60 Max: 0.81
Unexplained variation: Factor variable, unique ID for each sampling event	168 levels		
Unit area: Numeric variable of the area (m^2) of the sampling unit	Min: 16.00 Median: 36.00		Mean: 66.62 Max: 200.00
Weather: Factor variable describing the weather when the sampling event took place, three levels	Cloudy: 16 Partly cloudy: 45		Sunny: 107
Wind: Wind speed ($m\ s^{-1}$) measured by the recorder	Min: 0.00 Median: 1.45		Mean: 1.65 Max: 5.50

The extended dataset

The extended dataset was created by adding 156 additional sampling events to the primary dataset. The extended dataset then included 324 sampling events and 21 variables (no environmental variables were included, Table B.2).

Table B.2 (continues to the next page): Overview of the variables included in the extended dataset.

Variable	Summary			
Day: Numeric variable describing the number of days after fieldwork initiation	Min: 0.00 Median: 14.00		Mean: 16.77 Max: 53.00	
Day factor: Day as factor variable, 25 levels	D0: 11 D1: 9 D2: 4 D3: 2 D4: 9 D5: 6 D6: 18	D7: 13 D8: 18 D9: 6 D10: 14 D11: 16 D12: 20	D13: 4 D14: 16 D15: 26 D16: 46 D17: 13 D18: 28	D48: 9 D49: 9 D50: 6 D51: 8 D52: 7 D53: 6
Distance from field border: Numeric variable (m^2)	Min: 30.00 Median: 50.00		Mean: 70.59 Max: 120.00	
Distance from field border factor: Factor variable, 5 levels	30: 6 50: 193 80: 6		100: 113 120: 6	
Field section: Factor variable	87 levels			
Flower density: Numerical variable, the flower density of the sampling unit (m^{-1})	Min: 1.0 Median: 144.5		Mean: 136.3 Max: 251.0	
Total number of flowers: The number of flowers within the sampling unit	Min: 475.2 Median: 19,860.0		Mean: 29,380.0 Max: 128,000.0	
Focus: Factor variable distinguishing between the focus of the sampling events, three levels	Environmental variation: 78 Different sampling methods: 90 Networks: 156			
Latitude: The latitude where the sampling event took place in decimals	Min: -36.94 Median: -36.00		Mean: -35.99 Max: -35.26	
Month: Numeric variable describing the month (January=1, February=2)	Min: 1.00 Median: 1.00		Mean: 1.14 Max: 2.00	
Month factor: Month as factor, two levels	January: 279 February: 45			
Observed bees: The number of observed bees	Min: 0.00 Median: 0.00		Mean: 0.30 Max: 6.00	
Observed hoverflies: The number of observed hoverflies	Min: 0.00 Median: 0.00		Mean: 0.03 Max: 3.00	

Table B.2 (continued): Overview of the variables included in the extended dataset.			
Observer: ID of the person carrying out the sampling event	O1: 14 O2: 53 O3: 37	O4:52 O5: 55	O6: 57 O7: 56
Sampling method: Factor variable, two levels	Plots: 138 Transects: 186		
Sampling unit: Factor variable, four levels.	100×2: 186 4×4: 24		6×6: 98 20×4: 16
Study location: Factor variable, three levels	North: 30 Mid: 270		South: 24
Study site: Factor variable	39 levels		
Team: Factor variable distinguishing between the two observational teams conducting the sampling, 2 levels	Team1: 168 Team2: 156		
Unexplained variation: Factor variable, unique ID for each sampling event	324 levels		
Unit area: Numeric variable of the area (m^2) of the sampling unit	Min: 16.0 Median: 200.0		Mean: 130.8 Max: 200.0

The revisits dataset

The revisits dataset included 30 observations and two variables (Table B.3).

Table B.3: Overview of the variables included in the revisits dataset.

Variable	Summary	
Bee: ID of the observed bee	30 levels	
Flowers visited: Numeric variable, the number of flower visits the bee made before flying away	Min: 2.00 Median: 26.50	Mean: 29.13 Max: 108.00

C Model selection function

The function I used in the model selection procedure ('regress.ic.search.2step') was written by Trond Reitan and can be found at the URL: http://folk.uio.no/trondr/R/regress_search.R. A simplified version, but functional for my use, is shown at the end of this appendix*.

When calling the function, a data file, the response variable, a vector containing possible covariates, the family and an offset variable were specified:

```
bestFreqModIE <-
       regress.ic.search.2step(data=primary_data, response='Bees', covs=covs,
                               family='poisson', offset='offset(log(Flowers))')

bestFreqModEE <-
       regress.ic.search.2step(data=extended_data, response='Bees + Hoverflies',
                               covs=covs, family='poisson',
                               offset='offset(log(Flowers))')
```

The model selection function starts by making a null model, and then goes into a while-loop. For each iteration, a multitude of new models are made, based on either the null model (the first iteration only) or the outcome of the last iteration. The models are generated by, using the elements in the list of potential covariates, attempting to 1) add one covariate, 2) add two covariates, 3) remove one covariate, 4) remove two covariates, 5) replace one covariate, 6) replace one while removing another covariate, 7) replace one while adding another covariate and 8) replace two covariates. The combination of covariates that yields the model with the lowest BIC value is returned ('status.new'), and a new iteration begins with this same combination ('status'). The while-loop goes on until the returned combination ('status.new') is the same as it was at the beginning of the iteration. Then the generated model with the lowest BIC value (the *best model*) is returned from the model selection function.

*Below is a simplified version (some if-commands, superfluous for my use, were removed) of the function:

```
regress.ic.search.2step=function(data,response,covs,family="normal",
 offset="",strata="",use.glmmADMB=F, IC="BIC",threshold="flexible",
 check.est=F,check.se=F, do.return.status=F, start.status=NA) {
  numcov=length(covs)

  # Assume random factors start with "(":
```

```
israndf=substr(covs,1,1)=="("

# Create regression string:
getmodel=function(status) {
 if(sum(status)==0) {
   ret=sprintf("%s~1",response) }
 else {
   ret=sprintf("%s~",response)
   index=which(status==1)
   ret=sprintf("%s%s",ret,covs[index[1]])
   if(length(index)>1)
    for(i in index[2:length(index)])
     ret=sprintf("%s+%s", ret, covs[i]) }
 if(offset!="")
  ret=sprintf("%s+%s",ret,offset)

 if(strata!="")
  ret=sprintf("%s+strata(%s)",ret,strata)

 return(ret) }

# Make zero model:
if(family!="normal" & family!="clogit" & family!="ordinal") {
  if(offset!="")
    res0=glm(as.formula(sprintf("%s~1+%s",response,offset)),data=data,family=family)

# Run regression:
getres=function(status) {
  str=getmodel(status)

  if(sum(israndf[status!=0])>0) {
    #print.srcref(sprintf("glmer(as.formula(%s), data=data, family=%s)",str,family))
    return(withRestarts(tryCatch(glmer(as.formula(str), data=data,family=family) ),
           abort=function() {return(res0)})) }

  return(withRestarts(tryCatch(glm(as.formula(str), data=data,family=family) ),
         abort=function() {return(res0)})) }

# IC instead of stepwise hypothesis testing
# allows for going up and down and sideways in the model search

# Run one step up/down/sideways (stop when the IC doesn't improve):
step.ic=function(status) {
res0=getres(status)

ic0=best.ic=BIC(res0)
print.srcref(sprintf("Best model: %s",getmodel(status)))
print.srcref(sprintf("IC=%f",ic0))
best.status=status

# Add one:
#show("Add one")
for(i in which(status==0)) {
 status1=status
 status1[i]=1
 res1=getres(status1)
 ic1=BIC(res1)
 if(!is.na(ic1)) {
  if(is.na(ic0)) {
   best.ic=ic1
   best.status=status1 }
  if(!is.na(ic0))
   if(ic1<best.ic) {
   best.ic=ic1
   best.status=status1 }
```

```
 }
}

# Add two:
#show("Add two")
for(i in which(status==0))
 for(j in which(status==0))
  if(i<j) {
 status1=status
 status1[i]=1
 status1[j]=1
 res1=getres(status1)
 ic1=BIC(res1)
 if(!is.na(ic1)) {
  if(is.na(ic0)) {
   best.ic=ic1
   best.status=status1 }
  if(!is.na(ic0))
   if(ic1<best.ic) {
   best.ic=ic1
   best.status=status1 }
 }
}

#Remove one:
#show("Remove one")
for(i in which(status==1)) {
 status1=status
 status1[i]=0
 res1=getres(status1)
 ic1=BIC(res1)
 #show(c(covs[i],i,ic1))
 if(!is.na(ic1)) {
  if(is.na(ic0)) {
   best.ic=ic1
   best.status=status1 }
  if(!is.na(ic0))
   if(ic1<best.ic | (ic1==best.ic & sum(status1)<sum(status))) {
   best.ic=ic1
   best.status=status1 }
 }
}

#Remove two:
#show("Remove")
for(i in which(status==1))
 for(j in which(status==1))
  if(i<j) {
 status1=status
 status1[i]=0
 status1[j]=0
 res1=getres(status1)
 ic1=BIC(res1)
 if(!is.na(ic1)) {
  if(is.na(ic0)) {
   best.ic=ic1
   best.status=status1 }
  if(!is.na(ic0))
   if(ic1<best.ic | (ic1==best.ic & sum(status1)<sum(status))) {
   best.ic=ic1
   best.status=status1 }
 }
}

#Replace one:
```

```
#show("Replace")
for(i in which(status==1))
 for(j in which(status==0)) {
 status1=status
 status1[i]=0
 status1[j]=1
 res1=getres(status1)
 ic1=BIC(res1)
 if(!is.na(ic1)) {
  if(is.na(ic0)) {
   best.ic=ic1
   best.status=status1 }
  if(!is.na(ic0))
   if(ic1<best.ic | (ic1==best.ic & sum(status1)<sum(status))) {
   best.ic=ic1
   best.status=status1 }
 }
}

#Replace one, remove one:
#show("Replace one, remove one")
for(i in which(status==1))
 for(j in which(status==0))
  for(k in which(status==1))
   if(i>k) {
 status1=status
 status1[i]=0
 status1[j]=1
 status1[k]=0
 res1=getres(status1)
 ic1=BIC(res1)
 if(!is.na(ic1)) {
  if(is.na(ic0)) {
   best.ic=ic1
   best.status=status1 }
  if(!is.na(ic0))
   if(ic1<best.ic | (ic1==best.ic & sum(status1)<sum(status))) {
   best.ic=ic1
   best.status=status1 }
 }
}

#Replace one, add one:
#show("Replace one, add one")
for(i in which(status==1))
 for(j in which(status==0))
  for(k in which(status==0))
   if(j>k) {
 status1=status
 status1[i]=0
 status1[j]=1
 status1[k]=1
 res1=getres(status1)
 ic1=BIC(res1)
 if(!is.na(ic1)) {
  if(is.na(ic0)) {
   best.ic=ic1
   best.status=status1 }
  if(!is.na(ic0))
   if(ic1<best.ic | (ic1==best.ic & sum(status1)<sum(status))) {
   best.ic=ic1
   best.status=status1 }
 }
}
```

```
#Replace two:
#show("Replace two")
for(i in which(status==1))
 for(j in which(status==0))
  for(k in which(status==1))
   for(l in which(status==0))
     if(i>k & j>l) {
 status1=status
 status1[i]=0
 status1[j]=1
 status1[k]=0
 status1[l]=1
 res1=getres(status1)
 ic1=BIC(res1)
 if(!is.na(ic1)) {
  if(is.na(ic0)) {
   best.ic=ic1
   best.status=status1 }
  if(!is.na(ic0))
   if(ic1<best.ic | (ic1==best.ic & sum(status1)<sum(status))) {
   best.ic=ic1
   best.status=status1 }
 }
}

return(best.status) }

status=rep(0,numcov)
if(sum(is.na(start.status))==0 & length(start.status)==numcov)
  status=start.status
status.new=step.ic(status)
while(sum(status.new!=status)>0) {
  status=status.new
  status.new=step.ic(status) }
status=status.new

if(!do.return.status)
  return(getres(status))
return(list(status=status, res=getres(status))) }
```

D Potential covariates

This appendix provides an overview of the covariates that were included in the model selection procedure for the frequency models based on the primary (Table D.1) and the extended (Table D.2) datasets.

D.1 From the primary dataset

Table D.1 (continues to the next page): List of covariates that were included in the model selection procedure to find the best model of flower visitor frequency, including environmental variation, on soybean in Buenos Aires in 2016, based on the primary dataset.

	Covariate	Description
	Fixed effects	
Weather variables	Temperature (°C) Relative humidity (%)	Temperature and relative humidity variables obtained from two sources were included: 1) Handheld weather recorder, measured prior to each sampling event 2) Weather logger placed within the site, measured within half an hour from the sampling event
	Wind speed (m s^{-1})	Measured from handheld weather recorder
	Weather	Factor variable defining the weather during the observation (sunny, partly cloudy, overcast)
Spatial variables	Study location	Factor variable (north, mid, south)
	Latitude (°S)	The latitude of the study site
	Sampling method	Factor variable (plot, transect)
	Sampling unit	Factor variable (100×2, 20×4, 6×6, 4×4)
	Unit area (m^2)	The area of the sampling unit
	Distance from field border (m)	As numerical and factor variable (30, 50, 80, 100, 120)
Temporal variables	Day	Number of days after fieldwork initiation
	Time of day	Three variables were included to test daily rhythms: 1) Time of day from 0 to 1 (t) 2) $\cos(2\pi t) + \sin(2\pi t)$ 3) $\cos(2\pi t) + \sin(2\pi t) + \cos(4\pi t) + \sin(4\pi t)$
Other variables	Focus	Factor variable (environmental variation, different sampling methods)
	Flower density (m^{-1})	The flower density of the sampling unit
Transformations		
Quadratic terms		Quadratic terms were included to test for non-linear effects. They were included for both of the temperature and humidity variables, all the numerical spatial variables, and for wind speed, day and flower density

Table D.1 (continued): List of covariates that were included in the model selection procedure to find the best model of flower visitor frequency, including environmental variation, on soybean in Buenos Aires in 2016, based on the primary dataset.

Logarithmic terms		Logarithmic terms were included in order to test for power law relationships. Such terms were included for flower density, unit area and the numerical variable of distance from field border The total number of flowers were also included as a logarithmic term in case the number of flower visitors do not scale proportional to the number of flowers.
Statistical interactions		
Temperature and relative humidity in interactions with other factors		Both of the temperature and humidity variables were tested for interactions with latitude, area and time of day (as continuous). This was to see if temperature and/or humidity affected how latitude, area and/or time of day affected the pollinator frequency (or vice versa)
Weather variable interactions		Both of the temperature and humidity variables were tested for interactions with each other and with wind
Random effects		
Spatial variables	Study site	Factor variable (N1, N2, N3, M1, M2, M3, S1, S2, S3)
	Field section	Factor variable testing for spatial variation on within field scale (26 levels from 9 fields (sites)
Temporal variables	Day factor	Day as factor (18 levels)
	Time factor	Date and hour of the day as factor (143 levels on 168 observations)
Other variables	Observer	ID of the person carrying out the sampling event (3 levels)
	Unexplained variation	Factor variable (168 levels). Unique ID of each sampling event to account for possible overdispersion
Random slopes		
Observer – sampling method interactions		The variables sampling method, unit area, ln(unit area), and sampling unit were included
Field section – sampling method interactions		The variables sampling method, unit area, ln(unit area), and sampling unit were included
Day factor – temperature		Temperature from nearest weather logger
Day factor – time of day		Time of day from 0 to 1 (t)
Study site – day		Day as continuous variable

D.2 From the extended dataset

Table D.2 (continues to the next page): List of covariates that were included in the model selection procedure to find the best explanation of flower visitor frequency, excluding environmental variation, on soybean in Buenos Aires in 2016, based on the extended dataset.

Covariate		Description
Fixed effects		
Spatial variables	Study location	Factor variable (north, mid, south)
	Latitude (°S)	The latitude of the study site
	Sampling method	Factor variable (plot, transect)
	Sampling unit	Factor variable (100×2, 20×4, 6×6, 4×4)
	Unit area (m^2)	The area of the sampling unit
	Distance from field border (m)	As numerical and factor variable (30, 50, 80, 100, 120)
Temporal variables	Day	Number of days after fieldwork initiation
	Month	Numerical variable where January=1 and February=2
Other variables	Focus	Factor variable (environmental variation, different sampling methods)
	Flower density (m^{-1})	The flower density of the sampling unit
Transformations		
Quadratic terms		Several quadratic terms were included in order to test for non-linear effects. Such terms were included for all of the numerical spatial variables, as well as for the variables day and flower density
Logarithmic terms		Logarithmic terms were included in order to test for power law relationships. Such terms were included for flower density, unit area and the numerical variable of distance from field border The total number of flowers were also included as a logarithmic term in case the number of flower visitors do not scale proportional to the number of flowers.
Random effects		
Spatial variables	Study site	Factor variable (39 levels)
	Field section	Factor variable testing for spatial variation on within field scale (87 levels from 39 fields (sites)
Temporal variables	Day factor	Day as factor (25 levels)
	Month factor	Month as factor (January, February)
Other variables	Observer	ID of the person carrying out the sampling event (7 levels)
	Team	Factor variable distinguishing between the two observational teams conducting the sampling (2 levels)
	Unexplained variation	Factor variable (324 levels). Unique ID of each sampling event to account for possible overdispersion

Table D.2 (continued): List of covariates that were included in the model selection procedure to find the best explanation of flower visitor frequency, excluding environmental variation, on soybean in Buenos Aires in 2016, based on the extended dataset.

Random slopes	
Observer – sampling method interactions	The variables sampling method, unit area, ln(unit area), and sampling unit were included
Team – sampling method interactions	The variables sampling method, unit area, ln(unit area), and sampling unit were included
Field section – sampling method interactions	The variables sampling method, unit area, ln(unit area), and sampling unit were included
Study site – day	Day as continuous variable

E Flower visitation probability

This appendix shows the detailed calculations of the flower visitation probability estimates.

```
library(lme4)  # Required package

# Use the best frequency model including
# environmental variation: bestFreqModIE

# Get coefficients
coefs <- fixef(bestFreqModIE)

# Get variances for random factors
vars <- as.data.frame(VarCorr(bestFreqModIE))$vcov
```

Using Equation 1 for $bee\_visitors_{mod}$, and Equation 2 for $bee\_visitors_{raw}$:

```
# Finding estimate of number of expected bee visitors
# per flower per sampling event from model (bee_visitors.mod)
# Set sampling method to be 'plot', i.e. 0
(bee_visitors.mod <- mean(exp(coefs[1]
                            + coefs[2] * primary_data$Humidity
                            + coefs[3] * 0
                            + coefs[4] * primary_data$logger.temp
                            + coefs[5] * primary_data$Humidity2
                            + vars[1]/2 + vars[2]/2)))
## [1] 2.495691e-05
# Finding estimate of number of bee visitors
# per flower per sampling event from raw data (bee_visitors.raw)
# Use only sampling events where plots were used
(bee_visitors.raw <-
  mean(primary_data$Bees[primary_data$Transect==F]/
         primary_data$Flowers[primary_data$Transect==F]))
## [1] 2.013384e-05
```

Using Equation 3 to find *bee_visits*:

```
# Estimate mean number of visits per bee individual (bee_visits)
(bee_visits <- mean(revisits_data$flowers_visited))
## [1] 29.13333
```

Using Equation 4 to find *total_visits* for both of the estimates:

```
# Estimate the total number of visits
# per sampling event (total_visits)
(total_visits.mod <- bee_visitors.mod * bee_visits)
## [1] 0.0007270781
(total_visits.raw <- bee_visitors.raw * bee_visits)
## [1] 0.000586566
```

Using Equation 5 to find the flower visitation probabilities p_{mod} and p_{raw}:

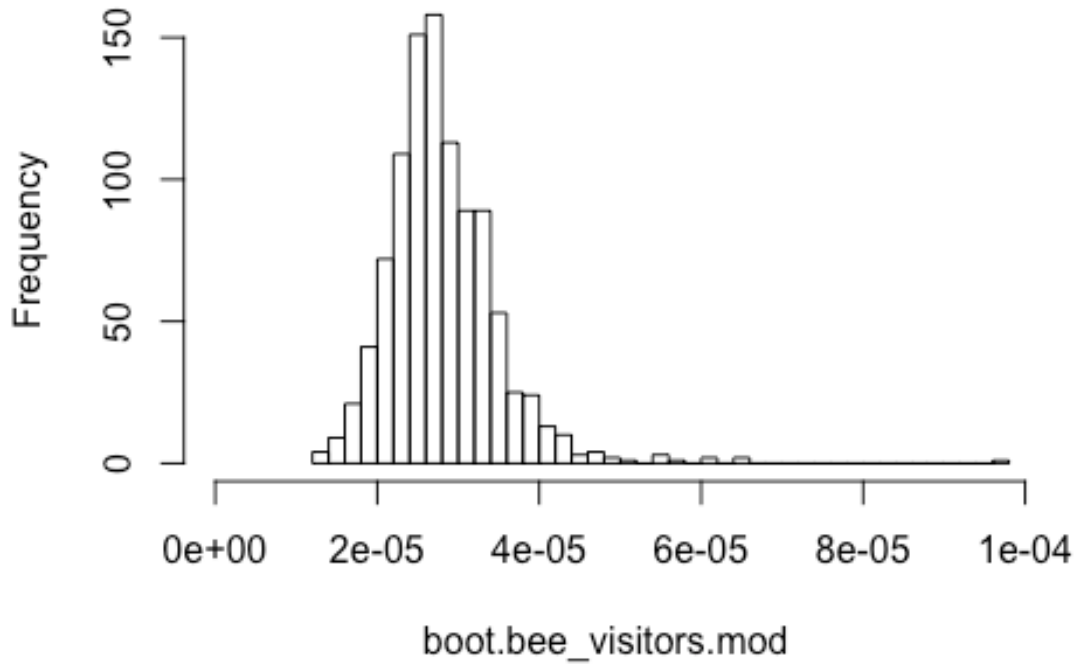
```
# Estimate of total flower visits per 2 days,
# given that pollinators are active 12 hours a day (p)
(p.mod <- total_visits.mod * ((12*60)/20)*2)
## [1] 0.05234962
(p.raw <- total_visits.raw * ((12*60)/20)*2)
## [1] 0.04223275
```

I then used bootstrap to make 95% confidence intervals for the probability estimates. This was done by generating 1000 new datasets, based on the primary dataset and with replacements. For each new dataset, I calculated $bee\_visitors_{mod}$ and $bee\_visitors_{raw}$, and saved them in a list (boot.bee_visitors.X):

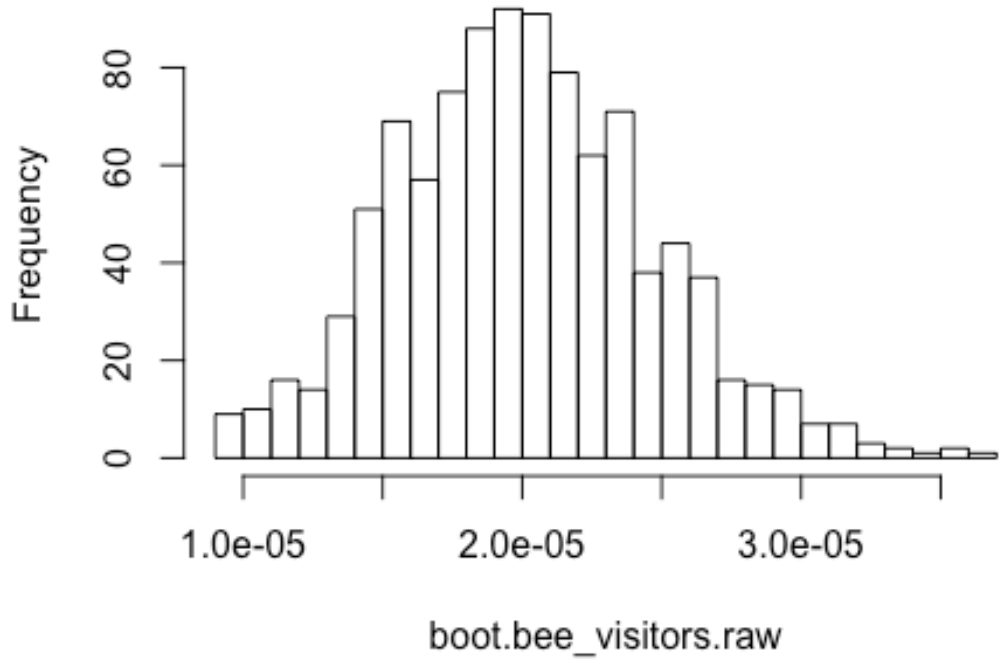
```
boot.bee_visitors.mod <- vector(mode="numeric", length=1000)
boot.bee_visitors.raw <- vector(mode="numeric", length=1000)

# Making a dataset including only plot samplings to use
# in raw data estimates
primary_data.plot <- primary_data[primary_data$Transect==0,]

for (i in 1:1000) {
  # Sample a new dataset with replacement
  boot.df.primary <-
    primary_data[sample(nrow(primary_data), size=nrow(primary_data), replace=T), ]

  # Remove possible levels that have not been included
  boot.df.primary <- droplevels(boot.df.primary)

  # MODEL
  # Make the model with the new dataset
  boot.bestFreqModIE <- glmer(Bees ~ Humidity + Transect
                                  + logger.temp + Humidity2
                                  + offset(log(Flowers))
                                  + (1 | Observer) + (1 | Section),
                              data=boot.df.primary, family="poisson")

  # Get coefficients
  boot.coefs <- fixef(boot.bestFreqModIE)

  # Get variance for random effects
  boot.vars <- as.data.frame(VarCorr(boot.bestFreqModIE))$vcov

  # Calculate estimate of number of expected visitors
  # per flower per sampling event from model
  boot.bee_visitors.mod[i] <- mean(exp(boot.coefs[1]
                                  + boot.coefs[2]
                                  * boot.df.primary$Humidity
                                  + boot.coefs[3] * 0
                                  + boot.coefs[4]
                                  * boot.df.primary$logger.temp
                                  + boot.coefs[5]
                                  * boot.df.primary$Humidity2
                                  + boot.vars[1]/2 + boot.vars[2]/2))

  boot.df.primary.plot <-
    primary_data.plot[sample(nrow(primary_data.plot),
                             size=nrow(primary_data.plot), replace=T), ]
  # RAW DATA
  boot.bee_visitors.raw[i] <-
```

```
    mean(boot.df.primary.plot$Bees/boot.df.primary.plot$Flowers)
}
```

```
hist(boot.bee_visitors.mod, xlim=c(0, max(boot.bee_visitors.mod)), breaks=50, main="")
```




```
hist(boot.bee_visitors.raw, breaks=30, main="")
```



I also generated 1000 new revisits datasets, and calculated *bee_visits* for each time, and saved the values in a list (boot.bee_visits):

```
# Mean number of visits per bee individual
bee_visits
## [1] 29.13333
boot.bee_visits <- vector(mode="numeric", length=1000)

for (i in 1:1000) {
  boot.bee_visits[i] <- mean(sample(revisits_data$flowers_visited,
                                    size = nrow(revisits_data),
                                    replace = T))
}
```

Finally, I found the 95% confidence limits as follows:

```
# Mean number of expected flower visits
boot.total_visits.mod <- boot.bee_visitors.mod * boot.bee_visits
boot.total_visits.raw <- boot.bee_visitors.raw * boot.bee_visits

# Total flower visits per 2 days,
# given that bees are active 12 hourse a day
boot.p.mod <- boot.total_visits.mod * ((12*60)/20)*2
boot.p.raw <- boot.total_visits.raw * ((12*60)/20)*2

# 95% confidence intervals model
lower.mod <- quantile(boot.p.mod, 0.025)
upper.mod <- quantile(boot.p.mod, 0.975)
mean.mod <- mean(boot.p.mod)
(CIlower.mod <- 2*mean.mod-upper.mod)
##      97.5%
## 0.02042829
(CIupper.mod <- 2*mean.mod-lower.mod)
##       2.5%
## 0.08573787
# 95% confidence intervals raw data
lower.raw <- quantile(boot.p.raw, 0.025)
upper.raw <- quantile(boot.p.raw, 0.975)
mean.raw <- mean(boot.p.raw)
(CIlower.raw <- 2*mean.raw-upper.raw)
##      97.5%
## 0.01607209
(CIupper.raw <- 2*mean.raw-lower.raw)
##       2.5%
## 0.06208869
```

F Environmental variation of the latitudinal gradient

There was no difference in temperature between the three study locations (Figure F.1, top). There was a slightly clearer difference in humidity (Figure F.1, bottom); however, as the loggers were placed only 25 cm above ground, this could be due to the morphological differences of the cultivar (Appendix A).

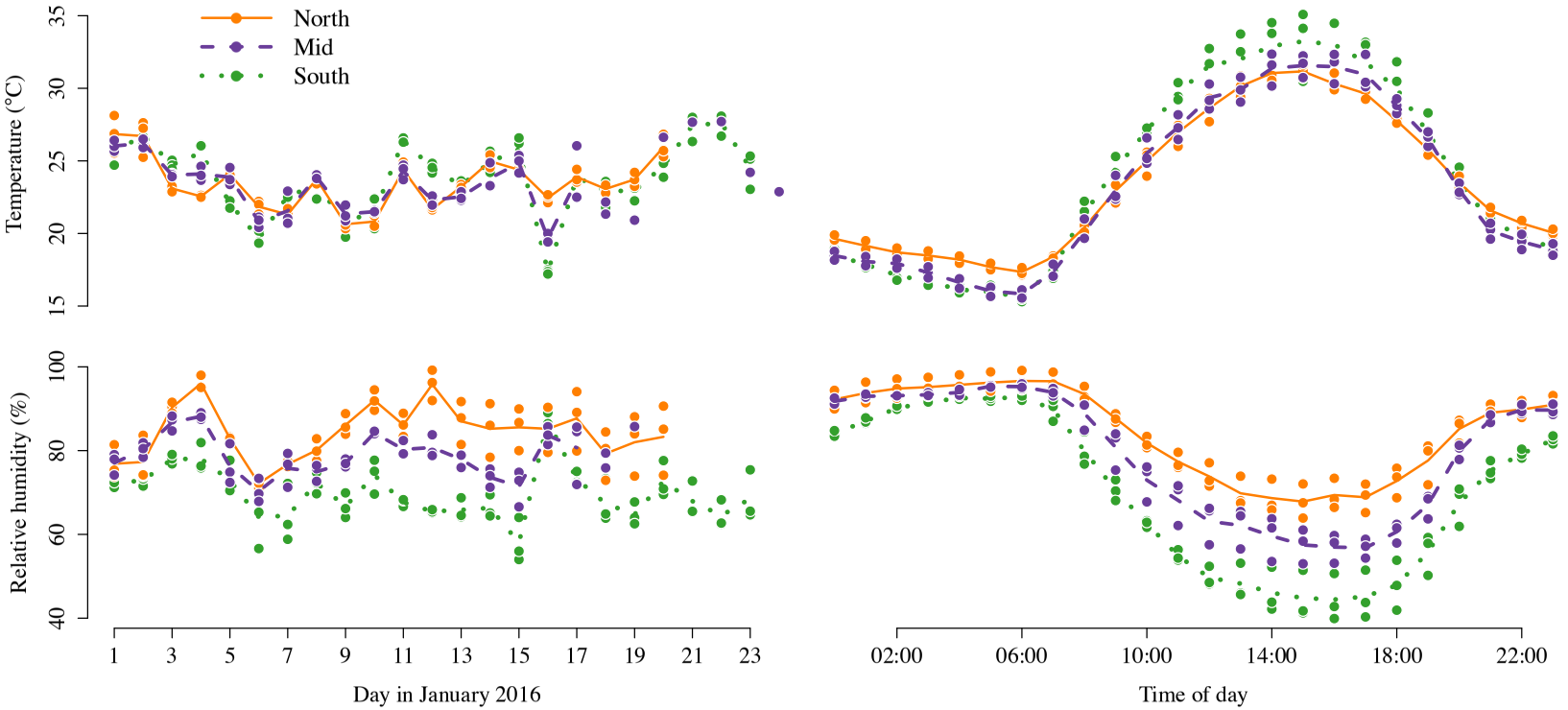


Figure F.1: The temporal variation of temperature and humidity between 1 and 23 January 2016 measured by the weather loggers. **Left:** The points are the daily averages of each study site, and the lines are the averages of the three sites within each study location. **Right:** The points are the averages for each hour of the day during the whole period (1-23 January) of each study site, and the lines are the averages of the three sites within each study location.

G Calculations for discussion

G.1 Gibson et al. (2011)

In Table 2, Gibson et al. (2011) presented the number of sampled insects per min, using transect walks and plot samplings (referred to as 'transect sampling' and 'timed observations', respectively, in their study) for three different years (Table G.1).

Figure G.1: A simplification of Table 2 in Gibson et al. (2011).

Year	Sampling method	Sampled insects (min^{-1})
2003	Plot samplings	0.094
	Transect walks	0.469
2004	Plot samplings	0.148
	Transect walks	0.352
2008	Plot samplings	0.137
	Transect walks	0.789

I first made a data frame with the values from Figure G.1:

```
(gibson <- data.frame(year=c("2003", "2004", "2008"),
                      plots=c(0.094, 0.148, 0.137),
                      transects=c(0.469, 0.352, 0.789)))
##   year plots transects
## 1 2003 0.094     0.469
## 2 2004 0.148     0.352
## 3 2008 0.137     0.789
```

I then found the relation between how many insects were sampled with transect walks compared to plot samplings for each year:

```
(gibson$frac <- gibson$transects/gibson$plots)
## [1] 4.989362 2.378378 5.759124
```

Finally, I found the mean between the years, and found the SE for the mean:

```
(mean(gibson$frac))
## [1] 4.375621
(sd(gibson$frac)/sqrt(3))
## [1] 1.023046
```

G.2 Nielsen et al. (2011)

In Figure 2a, Nielsen et al. (2011) presented the average number of individuals caught per site in olive groves with transect walks and plot samplings (referred to as 'standardized transect walks' and 'observation plots', respectively, in their study). I also looked at Table 3 to get an idea of what the exact values were, and concluded that they caught ~230 individuals per site using transect walks, and ~13 individuals per site using plot samplings. They explained that "each study site was sampled ten times using each method" (p. 972).

Transect walks

They spent a total of 50 min throughout the entire standardized transect in each round of observations (p. 972). Thus, the total time spent per site was $10 \times 50 = 500$ min. This means that on average, they caught $\frac{\sim 230}{500} \approx 0.46$ individuals per min using transect walks. The transects were 250 m long and 4 m wide (1000 m^2). The number of individuals caught per min per m^2 were then $\frac{0.46}{1000\ m^2} = 0.00046$.

Plot samplings

In each study site, they had 10 plots and each plot sampling lasted 6 min (p. 972). Thus, the total time spent per site was $10 \times 10 \times 6 = 600$ min. This means that on average, they caught $\frac{\sim 13}{600} \approx 0.022$ individuals per min using plot samplings. The plots were 1 m × 2 m (2 m^2). As there were 10 plots per site, the number of individuals caught per min per m^2 using plot samplings were then $\frac{0.022}{10 \times 2\ m^2} = 0.0011$.

Summary

As illustrated in Figure G.2, transect walks yielded ~21 times more observations per unit time, but plot samplings yielded a ~2.2 times higher visitor frequency (observations min^{-1} m^{-2}).

Figure G.2: Summary of the calculations made from Nielsen et al. (2011).

	Transect walks	Plot samplings	Fraction
Observations per min	0.46	0.022	$\frac{0.46}{0.022} \approx 21$
Observations per min per m^2	0.00046	0.001	$\frac{0.001}{0.00046} \approx 2.2$

G.3 Monasterolo et al. (2015)

In Table 1, Monasterolo et al. (2015) presented the observed visitation frequency (visits per flower per hour multiplied by 1000) of five different bee species to soybean. The values were 0.4, 0.03, 0.03, 0.03 and 0.17, equalling a total of $\frac{0.66}{1000} = 0.00066$ visits per flower per hour. This corresponds to $\frac{0.00066}{3} = 0.00022$ visits per flower per 20 min. I then put this value into Equation 5:

$$p_{Monasterolo} = 0.00022 \times \frac{12 \times 60 \text{ min}}{20 \text{ min}} \times 2 \text{ days} = 0.01584 \approx 1.58\%$$