

University of Wollongong

## Research Online

---

Faculty of Science, Medicine and Health -  
Papers: part A

Faculty of Science, Medicine and Health

---

1-1-2012

### Combining citizen science, bioclimatic envelope models and observed habitat preferences to determine the distribution of an inconspicuous, recently detected introduced bee (*Halictus smaragdulus* Vachal Hymenoptera: Halictidae) in Australia

Michael B. Ashcroft  
*Australian Museum*, [ashcroft@uow.edu.au](mailto:ashcroft@uow.edu.au)

John R. Gollan  
*Australian Museum*

Michael Batley  
*Australian Museum*

Follow this and additional works at: <https://ro.uow.edu.au/smhpapers>



Part of the [Medicine and Health Sciences Commons](#), and the [Social and Behavioral Sciences Commons](#)

---

#### Recommended Citation

Ashcroft, Michael B.; Gollan, John R.; and Batley, Michael, "Combining citizen science, bioclimatic envelope models and observed habitat preferences to determine the distribution of an inconspicuous, recently detected introduced bee (*Halictus smaragdulus* Vachal Hymenoptera: Halictidae) in Australia" (2012). *Faculty of Science, Medicine and Health - Papers: part A*. 3103.  
<https://ro.uow.edu.au/smhpapers/3103>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: [research-pubs@uow.edu.au](mailto:research-pubs@uow.edu.au)

---

## Combining citizen science, bioclimatic envelope models and observed habitat preferences to determine the distribution of an inconspicuous, recently detected introduced bee (*Halictus smaragdulus* Vachal Hymenoptera: Halictidae) in Australia

### Abstract

Introduced bees may compete with native fauna, spread parasites or pathogens to commercial bee hives, or increase the fecundity of introduced weeds. Therefore, the recent detection of *Halictus smaragdulus*, native to the western Palaearctic, in the Hunter Valley region of New South Wales (NSW, Australia) is cause for concern. However, it is currently difficult to justify control measures, as there is little known on its ecology, impacts and distribution. Determining the current distribution is fundamental to managing introduced species, yet this is difficult with inconspicuous species such as *H. smaragdulus*, especially as recent introductions are often found in low densities. We demonstrated how a combination of approaches could be used to improve the identification of occupied locations in NSW, including bioclimatic envelope models, proximity to known populations, and observed habitat preferences. Members of the public were also trained to collect specimens and improve overall survey efficiency. Bees were collected using pan traps and sweep netting. *H. smaragdulus* was detected at 44 new locations, extending the known distribution from ~1,250 to 46,800 km<sup>2</sup>. While bioclimatic envelope models helped guide survey locations, species detectability was higher when observed habitat preferences and proximity to known populations were also considered. We also demonstrated that with training via the internet and appropriate procedures for returning specimens in the mail, members of the public could successfully collect this small and inconspicuous invertebrate, with potential applications for other similar species.

### Disciplines

Medicine and Health Sciences | Social and Behavioral Sciences

### Publication Details

Ashcroft, M. B., Gollan, J. R. & Batley, M. (2012). Combining citizen science, bioclimatic envelope models and observed habitat preferences to determine the distribution of an inconspicuous, recently detected introduced bee (*Halictus smaragdulus* Vachal Hymenoptera: Halictidae) in Australia. *Biological Invasions*, 14 (3), 515-527.

Article type: Research paper

**Combining citizen science, bioclimatic envelope models and observed habitat preferences to determine the distribution of an inconspicuous, recently detected introduced bee (*Halictus smaragdulus* Vachal Hymenoptera: Halictidae) in Australia**

Running title: Distribution of an introduced bee

Ashcroft, Michael B<sup>\*</sup>

Gollan, John R<sup>1</sup>

Batley, Michael

Australian Museum, 6 College St, Sydney, NSW, 2010, Australia

<sup>\*</sup>Corresponding author: Email: Mick.Ashcroft@austmus.gov.au, Phone: +61-2-93206475, Fax: +61-2-93206021

<sup>1</sup>Current address: Australian Wetlands and Rivers Centre, School of Biological, Earth and Environmental Sciences, University of New South Wales, Kensington, NSW, 2033

Word count: 7385

**Abstract** Introduced bees may compete with native fauna, spread parasites or pathogens to commercial bee hives, or increase the fecundity of introduced weeds. Therefore, the recent detection of *Halictus smaragdulus*, native to the western Palaearctic, in the Hunter Valley region of New South Wales (NSW, Australia) is cause for concern. However, it is currently difficult to justify control measures, as there is little known on its ecology, impacts and distribution. Determining the current distribution is fundamental to managing introduced species, yet this is difficult with inconspicuous species such as *H. smaragdulus*, especially as recent introductions are often found in low densities. We demonstrated how a combination of approaches could be used to improve the identification of occupied locations in NSW, including bioclimatic envelope models, proximity to known populations, and observed habitat preferences. Members of the public were also trained to collect specimens and improve overall survey efficiency. Bees were collected using pan traps and sweep netting. *H. smaragdulus* was detected at 44 new locations, extending the known distribution from ~1,250km<sup>2</sup> to 46,800km<sup>2</sup>. While bioclimatic envelope models helped guide survey locations, species detectability was higher when observed habitat preferences and proximity to known populations were also considered. We also demonstrated that with training via the internet and appropriate procedures for returning specimens in the mail, members of the public could successfully collect this small and inconspicuous invertebrate, with potential applications for other similar species.

**Key words:** ecological niche models; invasive species; participatory science; range extension; species detectability; volunteers.

## Introduction

The impacts of introduced bees potentially include competition with native fauna, transmission of parasites and pathogens to commercial bee hives, disruption of native plant pollination networks, and increased pollination and spread of introduced weeds (Goulson 2003). Hence, extreme care should be taken before any new bee species are introduced outside their native ranges. However, the recent discovery (Gollan et al. 2008) of an introduced bee (*Halictus smaragdulus* Vachal Hymenoptera: Halictidae; native to western Palaearctic) in the Hunter Valley region of New South Wales (NSW), Australia, has received little attention. This is perhaps unsurprising given that controlling an introduced species is expensive and hard to justify unless the species is proven to be harmful and invasive (Mack et al. 2000). The ecology, distribution and impacts of *H. smaragdulus* are poorly known, but should harmful effects be detected in future, any delay in management might mean it is no longer possible to control the species (Sakai et al. 2001).

In this study, we aimed to make an initial assessment of the areas that were potentially affected by *H. smaragdulus* by determining the current distribution of this recently detected introduced species. It was first detected in Australia in riparian areas of the Hunter Valley (NSW) in November 2004 and January 2006 (Gollan et al. 2008). Although it was the second most abundant bee captured in that study, it was assumed to be a recent introduction because it had not been detected in previous surveys. Gollan et al. (2008) also noted that the species had high potential to spread, and indeed it may already have been present at other locations.

Determining the distribution of newly introduced species is problematic, as it is difficult to detect species that are only present in low densities (Harvey et al. 2009). Identifying expanding range boundaries is even more problematic, as populations are often patchy, temporally variable, and at even lower densities near the extremities (Frey 2009; Randin et al. 2006), and may be subject to complex dynamics (Kanda et al. 2009). To ensure early detection of introduced species, and to estimate the invaded range, it is important to develop efficient sampling strategies (Harvey et al. 2009; Hauser and McCarthy 2009).

Climate matching or bioclimatic envelope modelling techniques often give a poor indication of invasion success or extent (Williamson 1999), but due to a lack of ecological knowledge on many introduced species, these are often the only techniques that can be used (Baker et al. 2000). While some argue that climate matching or correlative niche-modelling techniques can reliably predict the geographic region of invasions (e.g. Peterson 2003, 2006; Sutherst 2003), others have suggested that models will not be able to predict the full extent of invasions when species adapt or evolve to cope with new conditions or face a new suite of interacting species (Broennimann et al. 2007; Broennimann and Guisan 2008). Indeed, there is increasing evidence that models cannot always be transferred to new geographic locations (Gray et al. 2009; Randin et al. 2006; Zanini et al. 2009).

There is now increasing recognition that bioclimatic models are more reliable when they are developed using both the native and invaded ranges, but it has been suggested that bioclimatic models developed using only the native range can still identify the initial point of introduction (Broennimann et al. 2007). This is important for the special case of newly detected arrivals, such as *H. smaragdulus*, as there is

insufficient information to include the invaded distribution in models. In this situation it is inevitable that models produced using only the native range will have higher uncertainty. However, when the ultimate objective is to conduct surveys to determine the actual distribution in the invaded range, identifying one true model is unnecessary. It may be preferable to use a range of models to guide surveying effort rather than rely on one model as a definitive prediction.

While coarse-scale bioclimatic models of the potential distribution of introduced species are an important management priority that has received a large amount of recent interest (e.g. Ward 2007), species distributions are affected by other factors at other scales and this affects how well bioclimatic models can be used to detect actual localities where the species is currently present (Loo et al. 2009). For example, in the absence of knowledge on habitat preferences, presence within a coarse grid cell may provide insufficient information to actually locate the species within that cell even if it is present. Determining and incorporating habitat preferences is therefore also an important priority to locate new populations, although there was limited information on habitat preferences for *H. smaragdulus* prior to this study and hence we had to rely on observations made during the course of our study.

Information on the distribution of introduced species can also be obtained by sampling locations in close proximity to known populations (Schmidt et al. 2010). Proximity-based sampling can be particularly effective for detecting populations that have dispersed short distances, but the area to be sampled obviously increases dramatically with dispersal distance. Therefore, it is also difficult and expensive to detect populations where there have been sporadic long-range dispersal events without

targeting environments using knowledge about habitat preferences or climatic suitability.

Members of the general community can also be a valuable resource for supplementing for survey data. Such activities are sometimes referred to as ‘citizen science’ and such participation in environmental and ecological monitoring projects have been described as a global phenomenon (Danielson et al. 2003). Citizen science can be especially informative for ecological niche models where the paucity of relevant scientific data makes it difficult to develop informative predictive models (Kadoya et al. 2009). However, specialised methods need to be developed in cases like the present one, where the species is inconspicuous and requires expertise for accurate identification. Recent technologies such as the internet can deliver training on methodological aspects, social media can support citizens quickly and effectively if questions or difficulties arise, and specimens can be mailed to experts for identification, provided that appropriate procedures are established.

In this study, we demonstrated how a combination of these approaches could be used to determine the distribution of *H. smaragdulus* in Australia. Sampling was undertaken at sites where coarse-grained bioclimatic envelope models predicted that the climate was most suitable, or within ~200km of the locations where the bee had previously been found in Australia. While bioclimatic models may help predict the overall range of a species, they are less suited to predicting the actual localities where the species may be found within the range (Guisan et al. 2006; Williams et al. 2009). To address this issue, we complemented the bioclimatic models by evolving our sampling techniques to target sites with habitats that matched the environments where we observed the bees nesting or foraging. Samples collected by members of the public also



supplemented the survey by providing low cost, albeit less intensive, sampling within at least part of the area of interest.

## Material and methods

### *Bioclimatic envelope models*

Information about the occurrence of *H. smaragdulus* in its native range was obtained from the Royal Belgian Institute of Natural Sciences (1029 records) and the Global Biodiversity Information Facility (GBIF) database ([www.gbif.org](http://www.gbif.org); 46 records). Duplicate locations and records over oceans and seas were discarded, resulting in 688 locations in the native range, geographically separated by up to 7,650 km between Morocco, France, Kazakhstan and Pakistan (Fig. 1). Five forms of the species have been identified based on morphological differences in genitalia (Pauly and Rassel 1982), and only ‘form D’ had been observed in Australia (Batley and Pauly, unpublished data). There were only 19 known locations for form D in the native range, though they still spanned a wide range of environmental conditions from Spain to Iran (~5,350 km).

We used the WorldClim global climate layers for current climate at a 2.5 arc minute resolution (Hijmans et al. 2005) to produce correlative bioclimatic models using Maxent version 3.2.19 with default values for regularisation parameters (Phillips et al. 2006). These regularization parameters are designed to prevent overfitting by producing models with fewer non-zero coefficients, and are suitable for a range of presence-only datasets (Phillips and Dudík 2008). Nevertheless, given the inherent variability of correlative models and the fact that we had insufficient information to include the

invaded range in models, we did not rely on Maxent to select the ‘correct’ predictors and produce one ‘true’ model. Instead, six models were produced using different combinations of predictors and distributional data. Three models were produced using all 688 records for *H. smaragdulus*, and another three with only the 19 form D records. The first model for each used all 19 WorldClim predictors, the second used four commonly used predictors (mean annual temperature, mean annual precipitation, maximum temperature of the warmest month, and minimum temperature of the coldest month), and the third used only mean annual temperature and precipitation. There will obviously be other combinations, statistical methods and environmental factors that will produce different predictions, but these six models provided an indication of how variable the models were, and were sufficient to act as a qualitative guide for the design of our surveys.

There was insufficient information on the distribution in the invaded range (14 locations over a minimum convex polygon area of  $\sim 1,250\text{km}^2$ ) prior to this study to include it in the bioclimatic models. The known locations were spatially autocorrelated relative to the density of records in the native range, spanned a small variation in environmental conditions, and we suspected were an incomplete representation of the invaded range (later justified by results). If we included incomplete knowledge of the Australian range in our models we may have biased predictions towards the known locations and against the unknown locations that were the ones we actually wanted to locate. We therefore elected to base models only on the native range. While this may add to model uncertainty, this uncertainty is inevitable when we do not know the distribution in the invaded range, and cannot assess a priori which model performs best at predicting unknown populations. Given the uncertainties with our bioclimatic

models, the six models we produced were used as a qualitative guide rather than a definitive prediction.

The species records from the native range were randomly divided into ten pools, and ten-fold cross validation was performed using nine pools at a time for training models, and one pool for validation. Model performance was assessed using the average Area Under the receiver operator characteristic Curve (AUC of the ROC) over the ten cross validation models from the native distribution, and the average AUC of these ten models when calculated using Australian distributional data both before and after our surveys. All AUC values were calculated by Maxent by providing a test sample file.

We do not imply that the model with the highest AUC is necessarily the best model, as there is no evidence that the species has reached equilibrium in Australia or that the full distribution is known. Therefore, the model that performs best with the known distribution is not necessarily the best indicator of the actual or future distribution.

### *Surveys of H. smaragdulus*

*H. smaragdulus* was targeted using yellow and white pan traps that were left for approximately 48 hours. Pan traps were used as they are a standardised and passive method of collecting, and were known to capture *H. smaragdulus* in Australia (Gollan et al. 2008). Traps were yellow or white plastic dishes, approximately 15 cm in diameter and 5 cm deep, which were pinned to the ground with wooden skewers and filled with salt water. Detergent was added to reduce surface tension.

We designed three initial surveys to include sites surrounding the locations where *H. smaragdulus* was first detected (Gollan et al. 2008), as well as areas predicted to have high climatic suitability by the six Maxent models. These three surveys did not target any specific habitat, as there was no *a priori* information on the species' preferences. The sites were distributed adjacent to roads (within 50 m) approximately 10 to 15 km apart, and included a variety of soil types and natural and human-dominated land uses, including forests, fields, riparian areas, vineyards, and irrigated areas. We placed two yellow and two white pans at each site. All traps were separated by approximately 5 m.

We conducted our first survey from 14–17 October 2008 at 100 sites in the region between Wollongong, Griffith and Albury (Fig. 2a). This area was targeted as the models consistently predicted it had highest climatic suitability, albeit at some distance from the known populations (Fig. 2). The sites were selected so that they spanned areas that were consistently identified as highly suitable across the six models (Fig. 2), albeit assessed qualitatively and some sites were in marginal habitat according to some of the models.

The second survey was conducted from 3–6 November 2008 at 94 sites in the region between Lithgow, Wellington and Tamworth (Fig. 2b), while the third survey was conducted from 13–16 December 2008 at 97 sites in the coastal region between Sydney and Port Macquarie (Fig. 2c). These two surveys included some sites in the Hunter Valley, but were mainly designed to survey areas in adjacent regions to the known populations. Climatic suitability was not the primary factor, and sites span a range of predicted suitabilities.

These three initial surveys were supplemented by records obtained by members of the public on a voluntary basis. *H. smaragdulus* is a small metallic bee (6 to 8 mm long) that is difficult to identify and distinguish from native species in-situ, making direct observations by untrained members of the public difficult. Therefore, we had to produce protocols and methods, which were not prohibitively expensive, that allowed the collection of samples and their subsequent return to our laboratory for identification.

Volunteers were recruited through groups that were thought to have an interest in this non-native species e.g. The Hunter Valley Amateur Beekeepers Association as well as the Australian Native Bee Research Centre. In the end, a total of 28 members of the public were enlisted all of whom placed four yellow pan traps at a site near their homes for 48 hours on two occasions between November and December 2008, and mailed the collected specimens to the Australian Museum for sorting and identification. To standardise samples as much as possible, a trapping kit was mailed to volunteers. Each kit included: sachets of salt and detergent, yellow dishes (as above), wooden food skewers, specimen jars for return of samples, prepaid envelopes and a set of instructions detailing where and how to deploy the traps. A training video was also produced and made available on the YouTube website (<http://www.youtube.com/watch?v=sDwXAiGyYK8>), and a blogging website was established so that participants could pose questions to us if they had any difficulties or questions. Volunteer samples were predominately from the coastal areas of the Hunter Valley, most likely due to the higher population in this area (Fig. 2d).

*Ad hoc* sweep netting was employed by one of the authors (MB) in late 2008 and early 2009, and targeted introduced flowers (e.g. *Gazania* spp.) in urban areas around Wellington, and between Sydney and Tamworth. Sweep netting is a less standardised

sampling method, but can be employed in one site visit and can rapidly evolve based on perceived habitat preferences. It was employed subjectively to provide information on the distribution of the bee and to provide information on floral resources that the bees were visiting.

Following these initial surveys, our casual observations were that *H. smaragdulus* preferred sandy creeks and rivers with exposed earth (*H. smaragdulus* nests in the ground) and locations where there were introduced weeds (e.g. *Galenia pubescens*) or flowering ground covers (sources of nectar). These apparent preferences were then used as a qualitative guide to target a fourth survey, which was conducted between 3–6 March 2009. This survey included 78 sites near the Goulburn River and the areas around Wellington and Gilgandra (Fig. 2e). These areas were targeted due to the presence of sandy rivers.

A fifth and final survey was conducted in February 2010, targeting 90 sites with suitable habitat near where the species was found during the initial surveys (Fig. 2f). This survey used six yellow pan traps at each site and no white pan traps, as initial results suggested that yellow pans were more effective at capturing the species (Gollan et al. 2011).

Once we had completed our surveys we produced another Maxent model using all records in the native and invaded ranges to estimate the potential future distribution of the species, and to confirm the reliability of the models we produced using only the native range. We limited predictors to the four commonly used predictors as results suggested that the full set of 19 predictors could lead to overfitting (see results).

## Results

The six Maxent models varied substantially in both performance and predicted distribution in Australia, even though they all had high cross-validation performance in the native range ( $AUC > 0.95$ ; Fig. 3). The models based on all 19 predictors had the highest AUC when assessed using cross-validation in the native range, but performed worst when validated using the Australian data. They predicted the Hunter Valley, where the species was known to exist, had low suitability (Maxent logistic output  $< 0.1$ ; Fig. 2). Models with fewer predictors predicted the Australian observations better (Fig. 3), with models based on form D records marginally better than those based on all records despite the much lower sample size (19 records versus 688 records).

All six models suggested that the most climatically suitable habitat was found in the region between Lithgow, Albury and Griffith, yet the first survey was conducted in this area and failed to locate the species. The absence was unlikely due to a lack of activity, as the survey captured 213 bees and 28 species, and *H. smaragdulus* was observed in high abundance in the Hunter Valley just prior to this survey (personal observations by JG and MB).

The second and third surveys captured 310 and 200 bees, consisting of 32 and 26 species respectively. The second survey obtained 13 specimens of *H. smaragdulus* over five new sites in the Hunter Valley, and an additional five specimens at an outlying site near Wellington. This extended the known range west by approximately 170 km. The third survey obtained 18 specimens of *H. smaragdulus* from three new sites in the Hunter Valley, and extended the known range to the east by 10 km and south by 20 km. The volunteers captured 171 bees and 12 species, including one specimen of *H. smaragdulus* that increased the known range to the east a further 15 km.

*Ad hoc* sweep netting located the species at three locations between Tamworth and the Hunter Valley, as well as in Sydney. The only previous specimen from Sydney was opportunistically collected in a dog's water dish, and it was not clear if this was just an isolated individual. Confirmation of more individuals in Sydney extended the known range 155 km further south.

The fourth survey, which targeted sandy creeks and rivers, obtained 539 specimens and 35 species, including 88 specimens of *H. smaragdulus* at 15 new sites. These included numerous locations between the Hunter Valley and the Wellington population along the Goulburn River, and two sites between the Goulburn River and Gilgandra (Fig. 4). Two sites contained dense populations near the eastern distributional limits in the Hunter Valley, but the species was not detected further south or east, despite targeting rivers and creeks in these areas.

The final survey from February 2010 obtained 130 specimens of *H. smaragdulus* over 31 sites. This included 13 new sites around the Hunter Valley and Goulburn River, and 18 sites which had been surveyed previously and which we confirmed the species was present.

Overall, the known range of the species was extended considerably over what was first reported (Gollan et al. 2008). The number of known locations increased from 14 to 58 and the range was extended 100 km north, 155 km south, 30 km east, and 185 km west (Fig. 4). This increased the known area that the species occupies, as determined using minimum convex polygons, from  $\sim 1,250\text{km}^2$  to  $46,800\text{km}^2$ . The final Maxent model we produced using data from both the native and invaded ranges (Fig. 4) was similar to those produced before we had adequate distributional data in the invaded range (Fig. 2). The areas with highest climatic suitability are generally found further



south than the Hunter Valley, so there is great potential for the species to spread further if suitable habitat is available.

## Discussion

### *The distribution of *H. smaragdulus* in Australia*

We have expanded the known range of *H. smaragdulus* in Australia considerably, and it is now known to occupy an area of ~46,800 km<sup>2</sup>. It was found in higher numbers and at a greater proportion of sites in the Hunter Valley than elsewhere (Fig. 4), and therefore it is likely to be having a larger impact in this area (Yokomizo et al. 2009). Potential impacts of introduced bees include increased competition for nest sites and floral resources, the introduction and spread of pathogens and parasites, increased pollination and spread of introduced weeds, and the disruption of pollination of native plants (Barthell et al. 2001; Cook et al. 2007; Dohzono et al. 2008; Goulson 2003; Kenis et al. 2009; Stout et al. 2002; Traveset and Richardson 2006). *H. smaragdulus* could have serious impacts due to its high relative abundance, long seasonal activity, and an apparent preference for introduced plants and declared noxious weeds in New South Wales. These include *Asparagus aphyllus*, *Centaurea* spp., *Convolvulus mauretanicus*, *Conyza canadensis*, *Eryngium campestre*, *Galenia pubescens*, *Gazania* spp., *Portulacca pilosa*, *Thymus capitatus*, *Verbena bonariensis* and other members of the family Umbelliferae (Gollan et al. 2008; Gollan, Batley and Pauly, unpublished data; Herrera 1988; Petanidou and Vokou 1993).

The eastern distributional limits have been well sampled, and it appears that the species does not currently inhabit coastal regions in the Hunter Valley. This is consistent with the bioclimatic envelope models, although the species has been observed in coastal regions near Sydney (Fig. 4). In contrast, the southern limit is currently unreliable, as the survey in the region between Albury, Griffith and Sydney was not targeted towards the specific habitat which we now consider suitable, and may have had low detectability. Re-surveying this area targeting sandy and weedy creeks and rivers is needed to confirm the absence of the species in this area, especially given that this area was consistently identified as the most climatically suitable by the original Maxent models (Fig. 2), and remained the most climatically suitable area in the model produced after our surveys (Fig. 4).

The northern and western limits were inconsistently defined by our models, and were difficult to identify with field sampling due to the sparser and smaller populations near these range limits. The distribution may extend further in these directions. Coarse-grained bioclimatic models may guide future surveying efforts in these areas (e.g. Fig. 4), but efficient sampling also depends on identifying the habitats where the species is most likely to be detected. Expert knowledge of suitable habitat should evolve as new survey data becomes available, and surveying strategies should adapt to these changes. Species distribution models may also prove more useful if fine-scale climate data is used (e.g. Ashcroft et al. 2008) or habitat factors are included (Vanreusel et al. 2007). However, there are still uncertainties regarding the transferability of habitat models to other geographic areas (Bamford et al. 2009; Murray et al. 2009; Vanreusel et al. 2007), and therefore the habitat factors we identified may only be valid in the Hunter Valley. The habitat preferences in the native range are unknown.

There is a strong possibility that *H. smaragdulus* will expand its distribution further. It was one of the most abundant species we captured, occurs over a broad range of climatic conditions in the native range, and is active from at least October to May (no surveys have been conducted yet between June and September). All these attributes suggest it is a generalist species and able to survive a broad range of conditions. In particular, the bioclimatic models suggest *H. smaragdulus* might thrive in Victoria and much of southern Australia (Fig. 4), although this would depend on finding suitable nesting sites and nectar sources.

#### *Performance of bioclimatic models*

Bioclimatic envelope models can make highly variable predictions based on the predictors selected, the climate data used and the statistical method employed. None of the six models we produced were regarded as the only possible prediction, especially as we had insufficient information to include the invaded range in models. The variability amongst the six models provided a good indication of the uncertainty involved in predicting species' distributions, and we suggest that using a range of models as a qualitative guide is more appropriate in the case of newly detected species than trying to select one true model. Even if we used the AUC in the invaded range to determine which model provided the best predictions of the known distribution, this is not necessarily the model that is best at predicting the unknown or future distribution. In any event, the locations where we detected the species with our surveys were the primary output of our study, and the models were simply a tool we used to help achieve that goal. Nevertheless, we also produced a model using both the native and Australian

ranges after our surveys (Fig. 4), and this acts as a guide to potential future distributions.

The models we produced both before (Fig. 2) and after (Fig. 4) our surveys consistently predicted that the most climatically suitable conditions were in the area between Albury and Lithgow, so there is a good chance that the species will eventually colonise this area where there is suitable habitat (e.g. sandy soils and flowering weeds and ground covers). On the other hand, models were inconsistent in their predictions for the Hunter Valley region, where the species was known to be present in high numbers. Models based on all 19 predictors had high AUC when assessed using cross-validation in the native range, but performed poorly when transferred to Australia. Even the models using only the 19 form D locations appeared to overfit to some degree when using 19 predictors (Fig. 3), so the number of predictors appeared to be the primary factor leading to overfitting.

While cross-validation may help detect when a model has been overfitted to the data used to train models, our results suggest that it may provide an overly optimistic view of how transferable the models are. Models that perform well when validated using an ‘independent’ dataset from the same study area (e.g. Elith et al. 2006) may perform poorly when transferred to new study areas (Peterson et al. 2007). Effectively, our results suggest it is possible to overfit to the geographic area where the model is produced, even if the model is not overfitted to the actual data used to produce the model. This highlights the danger of including too many predictors in models (see also Ward 2007) and illustrates the shortcomings of using data from the same study area to perform validation. We suggest it would also be wise to test multiple models for each species before concluding whether or not models are transferable to a new geographic

area, as selecting one model may not be indicative of the transferability of alternative models.

Another notable result from our models was that the models based only on 19 Form D records outperformed those based on all 688 records for the species, however the difference was small when we reduced the number of predictors. This improvement may be due to less spatial autocorrelation when the number of training points is reduced, but it is also possible that it is due to better taxonomic identification. While there is no firm evidence that the different forms of *H. smaragdulus* have different environmental requirements, there is a strong chance of ecotypes given the species' broad environmental range and its widespread distribution in its native range (Dillon Jr. 1984; Hájková et al. 2008; Holman et al. 2003). Models that treat all individuals of a species as genetically and environmentally identical may be suboptimal if there are genotypes or ecotypes with different environmental niches (Boyden et al. 2008; Hampe 2004; Lee 2002; Loehle 1998; Randin et al. 2006; Wright et al. 2006). Given the possibility of ecotypes, and the apparent success of the species in Australia, it would be prudent to prevent further introductions that could diversify the gene pool and broaden the environmental tolerance of the species in Australia.

### *Citizen science*

While the volunteers only captured one specimen of *H. smaragdulus*, it should be noted that many volunteers placed their traps in the coastal area where models predicted the climate was less suitable, and our other surveys did not locate the species in these coastal areas either. In addition, the volunteers conducted their trapping before we had

identified potential habitat preferences, and it would be wise to include this information to better target future surveying efforts.

Involving members of the public in survey work has the potential to accumulate large amounts of data and this is particularly useful when there is a lack of data to construct predictive models for newly detected species. This data may also be prohibitively expensive to collect at large spatial scales using other methods. To further compound this, the spatial distribution of invasive species can be continuously changing (Kadoya et al. 2009), so a high level of temporal replication is required.

While it was not our aim to utilise vast numbers of volunteers spanning the entire study area, our approach demonstrated that it could be achieved, and for the purposes of a dedicated monitoring programme, could be repeated relatively cost-effectively. Data collection by members of the public for research and management purposes is not new, however, it has been largely restricted to large and conspicuous mammals (see examples in Newman et al. 2003) and birds (e.g. McCaffrey 2005). Where data on terrestrial invertebrates have been collected by members of the public, they too have generally been the large and relatively easily recognisable species such as the bumblebee (*Bombus terrestris*) (Kadoya et al. 2009) or charismatic flagship species' such as butterflies (Swengel 1990) and dragonflies (Kadoya and Washitani 2007). Those surveys of terrestrial invertebrates have also relied on methods using field observations that require little, if any, specialist equipment. We have demonstrated that data on even the small and inconspicuous species can be generated by members of the public using standardised collecting techniques that are used by entomologists. Involving the public in invertebrate surveys may also help overcome some of the

‘perception’ problems that invertebrates suffer, especially among people that only recognize ‘the dirty cockroach’ and the ‘nuisance fly’ (Samways 2007).

## **Conclusions**

Our study demonstrated the advantages of using multiple methods to determine the distribution of newly detected introduced species. Proximity based methods are effective because species distributions expand outwards from occupied locations, while bioclimatic envelope models can help target climatically favourable regions at greater distances. However, while both these methods can help identify broad-scale regions of interest, we found they were inefficient to detect actual populations unless we used observed habitat preferences to improve survey efficiency at finer-scales. We also demonstrated that modern technology could be utilised to allow volunteers to collect inconspicuous species, and this also has the potential to improve survey efficiency.

## **Acknowledgements**

This project was supported by a grant from the WV Scott Charitable Trust. Species distribution data in the native range was supplied by Alain Pauly from the Royal Belgian Institute of Natural Sciences. Natalie Sullivan assisted with field work and sorting, and produced the training video for the volunteers with the help of Michael Elliott. Scott Ginn assisted with data basing and incorporating specimens into the Australian Museum’s collection. Thanks to the volunteers who collected samples and

Carsten Dormann and two anonymous referees for providing constructive comments on previous drafts.

## References

- Ashcroft MB, Chisholm LA, French KO (2008) The effect of exposure on landscape scale soil surface temperatures and species distribution models. *Landsc Ecol* 23: 211–225.
- Baker RHA, Sansford CE, Jarvis CH, Cannon RJC, MacLeod A, Walters KFA (2000) The role of climatic mapping in predicting the potential geographic distribution of non-indigenous pests under current and future climates. *Agric Ecosyst Environ* 82: 57–71.
- Bamford AJ, Monadjem A, Anderson MD, Anthony A, Borello WD, Bridgeford M, Bridgeford P, Hancock P, Howells B, Wakelin J, Hardy ICW (2009) Trade-offs between specificity and regional generality in habitat association models: a case study of two species of African vulture. *J Appl Ecol* 46: 852–860.
- Barthell JF, Randall JM, Thorp RW, Wenner AM (2001) Promotion of seed set in yellow star-thistle by honey bees: evidence of an invasive mutualism. *Ecol Appl* 11: 1870–1883.
- Boyden S, Binkley D, Stape JL (2008) Competition among Eucalyptus trees depends on genetic variation and resource supply. *Ecology* 89: 2850–2859.
- Broennimann O, Treier UA, Müller-Schärer H, Thuiller W, Peterson AT, Guisan A (2007) Evidence of climatic niche shift during biological invasion. *Ecol Lett* 10: 701–709.



- Broennimann O, Guisan A (2008) Predicting current and future biological invasions: both native and invaded ranges matter. *Biol Lett* 4: 585–589.
- Cook DC, Thomas MB, Cunningham SA, Anderson DL, De Barro PJ (2007) Predicting the economic impact of an invasive species on an ecosystem service. *Ecol Appl* 17: 1832–1840.
- Danielson F, Mendoza MM, Alviola P, Balete DS, Enghoff M, Poulsen K, Jensen AE (2003) Biodiversity monitoring in developing countries: what are we trying to achieve? *Oryx* 37: 407–409.
- Dillon Jr. RT (1984) Geographic distance, environmental difference, and divergence between isolated populations. *Syst Zool* 33: 69–82.
- Dohzono I, Kunitake YK, Yokoyama J, Goka K (2008) Alien bumble bee affects native plant reproduction through interactions with native bumble bees. *Ecology* 89: 3082–3092.
- Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz G, Nakamura M, Nakazawa Y, Overton JMcC, Peterson AT, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberón J, Williams S, Wisz MS, Zimmermann NE (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129–151.
- Frey JK (2009) Distinguishing range expansions from previously undocumented populations using background data from museum records. *Divers Distrib* 15: 183–187.

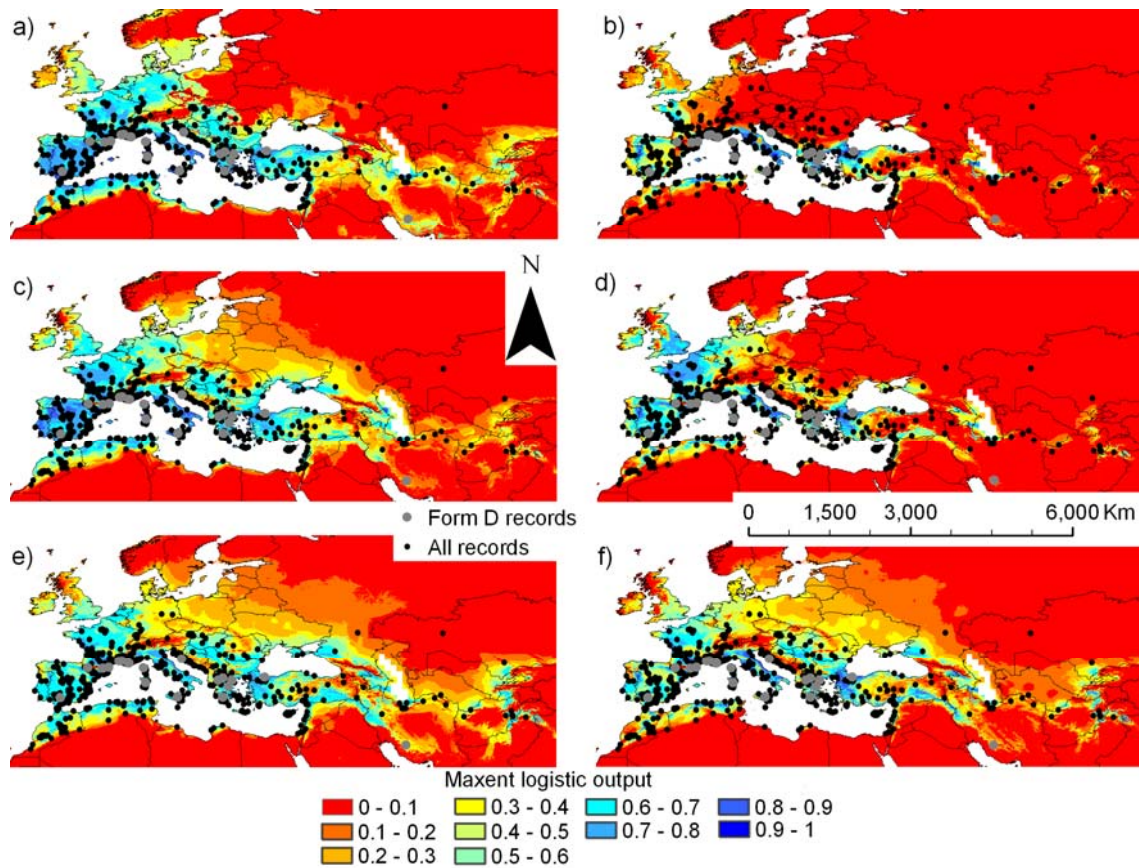
- Gollan JR, Batley M, Reid CAM (2008) The exotic bee *Halictus smaragdulus* Vachal, 1895 (Hymenoptera: Halictidae) in the Hunter Valley, New South Wales: A new genus in Australia. *Aust Entomol* 35: 21–26.
- Gollan JR, Ashcroft MB, Batley M (2011) A comparison of yellow and white pan traps in surveys of bee fauna in New South Wales, Australia (Hymenoptera: Apoidea: Anthophila). *Aust J Entomol* 50: 174–178.
- Goulson D (2003) Effects of introduced bees on native ecosystems. *Annu Rev Ecol Evol Syst* 34: 1–26.
- Gray TNE, Borey R, Hout SK, Chamnan H, Collar NJ, Dolman PM (2009) Generality of models that predict the distribution of species: conservation activity and reduction of model transferability for a threatened bustard. *Conserv Biol* 23: 433–439.
- Guisan A, Broennimann O, Engler R, Vust M, Yoccoz NG, Lehmann A, Zimmermann NE (2006) Using niche-based models to improve the sampling of rare species. *Conserv Biol* 20: 501–511.
- Hájková P, Hájek M, Apostolova I, Zelený D, Dítě D (2008) Shifts in ecological behaviour of plant species between two distant regions: evidence from the base richness gradient in mires. *J Biogeogr* 35: 282–294.
- Hampe A (2004) Bioclimatic envelope models: what they detect and what they hide. *Global Ecol Biogeogr* 13: 469–471.
- Harvey CT, Qureshi SA, MacIsaac HJ (2009) Detection of a colonizing, aquatic, non-indigenous species. *Divers Distrib* 15: 429–437.
- Hauser CE, McCarthy MA (2009) Streamlining ‘search and destroy’: cost-effective surveillance for invasive species management. *Ecol Lett* 12: 683–692.

- Herrera J (1988) Pollination relationships in southern Spanish Mediterranean shrublands. *J Ecol* 76: 274–287.
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* 25: 1965–1978.
- Holman JE, Hughes JM, Fensham RJ (2003) A morphological cline in *Eucalyptus*: a genetic perspective. *Mol Ecol* 12: 3013–3025.
- Kadoya T, Washitani I (2007) An adaptive management scheme for wetland restoration incorporating participatory monitoring into scientific predictions using dragonflies as an indicator taxon. *Global Environ Res* 11: 179–185.
- Kadoya T, Ishii HS, Kikuchi R, Suda S, Washitani I (2009) Using monitoring data gathered by volunteers to predict the potential distribution of the invasive alien bumblebee *Bombus terrestris*. *Biol Conserv* 142: 1011–1017.
- Kanda LL, Fuller TK, Sievert PR, Kellogg RL (2009) Seasonal source-sink dynamics at the edge of a species range. *Ecology* 90: 1574–1585.
- Kenis M, Auger-Rozenberg M-A, Roques A, Timms L, Péré C, Cock MJW, Settele J, Augustin S, Lopez-Vaamonde C (2009) Ecological effects of invasive alien insects. *Biol Invasions* 11: 21–45.
- Lee CE (2002) Evolutionary genetics of invasive species. *Trends Ecol Evol* 17: 386–391.
- Loehle C (1998) Height growth rate tradeoffs determine northern and southern range limits for trees. *J Biogeogr* 25: 735–742.

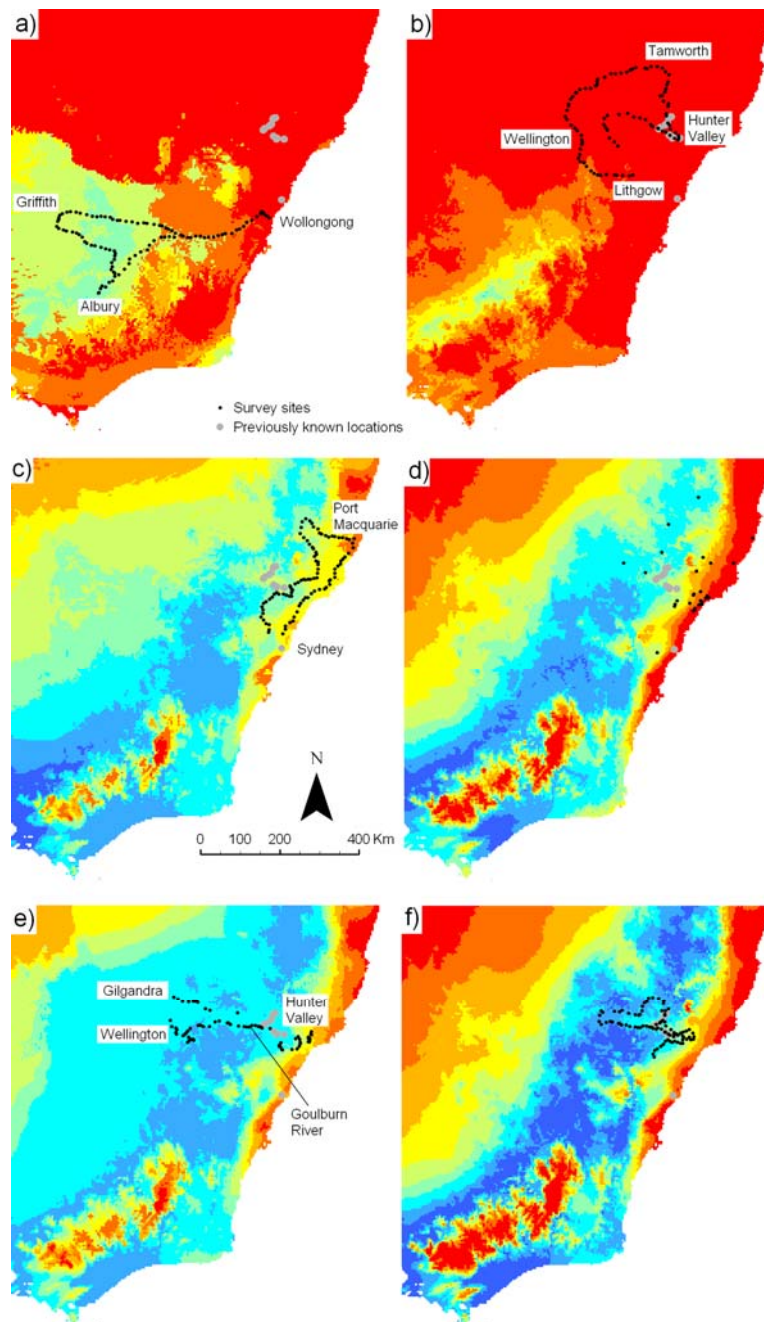
- Loo SE, Mac Nally R, O'Dowd DJ, Thomson JR, Lake PS (2009) Multiple scale analysis of factors influencing the distribution of an invasive aquatic grass. *Biol Invasions* 11: 1903–1912.
- Mack RN, Simberloff D, Lonsdale WM, Evans H, Clout M, Bazzaz FA (2000) Biotic invasions: causes, epidemiology, global consequences, and control. *Ecol Appl* 10: 689–710.
- McCaffrey RE (2005) Using citizen science in urban bird studies. *Urban Habitats* 3: 70–86.
- Murray J, Goldizen AW, O'Leary RA, McAlpine CA, Possingham HP, Choy SL (2009) How useful is expert opinion for predicting the distribution of a species within and beyond the region of expertise? A case study using brush-tailed rock-wallabies *Petrogale penicillata*. *J Appl Ecol* 46: 842–851.
- Newman C, Buesching CD, MacDonald DW (2003) Validating mammal monitoring methods and assessing the performance of volunteers in wildlife conservation—"Sed quis custodiet ipsos custodiet?". *Biol Conserv* 113: 189–197.
- Pauly A, Rassel A (1982) A scanning electron microscopic study of the gonostyli of *Halictus (Seladonia) smaragdulus* Vachal (Hymenoptera, Apoidea Halictidae). *Ann Soc Roy Zool Bel* 112: 137–146.
- Petanidou T, Volou D (1993) Pollination ecology of Labiatae in a phryganic (East Mediterranean) ecosystem. *Am J Bot* 80: 892–899.
- Peterson AT (2003) Predicting the geography of species' invasions via ecological niche modelling. *Q Rev Bot* 78: 419–433.
- Peterson AT (2006) Uses and requirements of ecological niche models and related distributional models. *Biodivers Inform* 3: 59–72.

- Peterson AT, Papeş M, Eaton M (2007) Transferability and model evaluation in ecological niche modelling: a comparison of GARP and Maxent. *Ecography* 30: 550–560.
- Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species distributions. *Ecol Model* 190: 231–259.
- Phillips SJ, Dudík M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31: 161–175.
- Randin CF, Dirnböck T, Dullinger S, Zimmermann NE, Zappa M, Guisan A (2006) Are niche-based species distribution models transferable in space? *J Biogeogr* 33: 1689–1703.
- Sakai AK, Allendorf FW, Holt JS, Lodge DM, Molofsky J, With KA, Baughman S, Cabin RJ, Cohen JE, Ellstrand NC, McCauley DE, O’Neil P, Parker IM, Thompson JN, Weller SG (2001) The population biology of invasive species. *Annu Rev Ecol Syst* 32: 305–332.
- Samways MJ (2007) Insect conservation: a synthetic management approach. *Annu Rev Entomol* 52: 465–487.
- Schmidt D, Spring D, Mac Nally R, Thomson JR, Brook BW, Cacho O, McKenzie M (2010) Finding needles (or ants) in haystacks: predicting locations of invasive organisms to inform eradication and containment. *Ecol Appl* 20: 1217–1227.
- Stout JC, Kells AR, Goulson D (2002) Pollination of the invasive exotic shrub *Lupinus arboreus* (Fabaceae) by introduced bees in Tasmania. *Biol Conserv* 106: 425–434.
- Sutherst B (2003) Prediction of species geographical ranges. *J Biogeogr* 30: 805–816.

- Swengel AB (1990) Monitoring butterfly populations using the fourth of July Butterfly count. *Am Midl Nat* 124: 395–406.
- Traveset A, Richardson DM (2006) Biological invasions as disruptors of plant reproductive mutualisms. *Trends Ecol Evol* 21: 208–216.
- Vanreusel W, Maes D, Van Dyck H (2007) Transferability of species distribution models: a functional habitat approach for two regionally threatened butterflies. *Conserv Biol* 21: 201–212.
- Ward DF (2007) Modelling the potential geographic distribution of invasive ant species in New Zealand. *Biol Invasions* 9: 723–735.
- Williams JN, Seo C, Thorne J, Nelson JK, Erwin S, O'Brien JM, Schwartz MW (2009) Using species distribution models to predict new occurrences for rare plants. *Divers Distrib* 15: 565–576.
- Williamson M (1999) Invasions. *Ecography* 22: 5–12.
- Wright JW, Davies KF, Lau JA, McCall AC, McKay JK (2006) Experimental verification of ecological niche modeling in a heterogeneous environment. *Ecology* 87: 2433–2439.
- Yokomizo H, Possingham HP, Thomas MB, Buckley YM (2009) Managing the impact of invasive species: the value of knowing the density–impact curve. *Ecol Appl* 19: 376–386.
- Zanini F, Pellet J, Schmidt BR (2009) The transferability of distribution models across regions: an amphibian case study. *Divers Distrib* 15: 469–480.

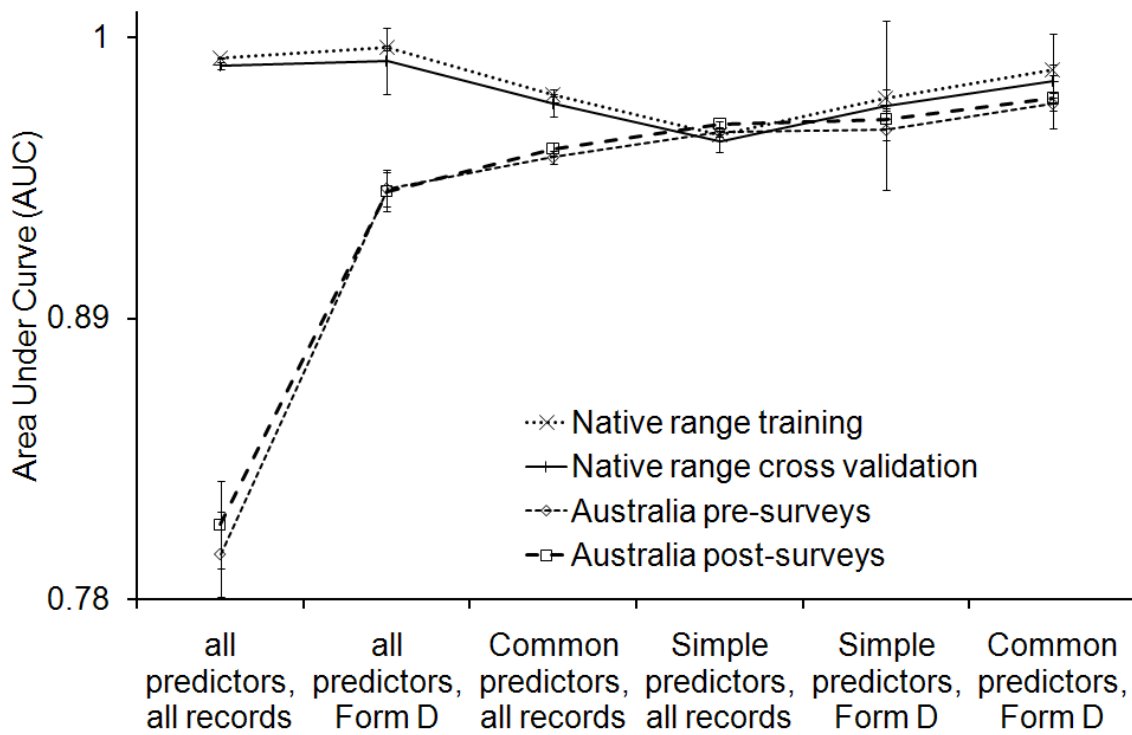


**Fig. 1** The known locations of *Halictus smaragdulus* in its native range. The logistic output from six Maxent models is shown (climatic favourability), with models on the left produced using all 688 records for the species, and the models on the right using the 19 known Form D records. Models from top to bottom were produced using all 19 WorldClim predictors, four commonly used predictors, and a simple model based only on mean annual temperature and precipitation.

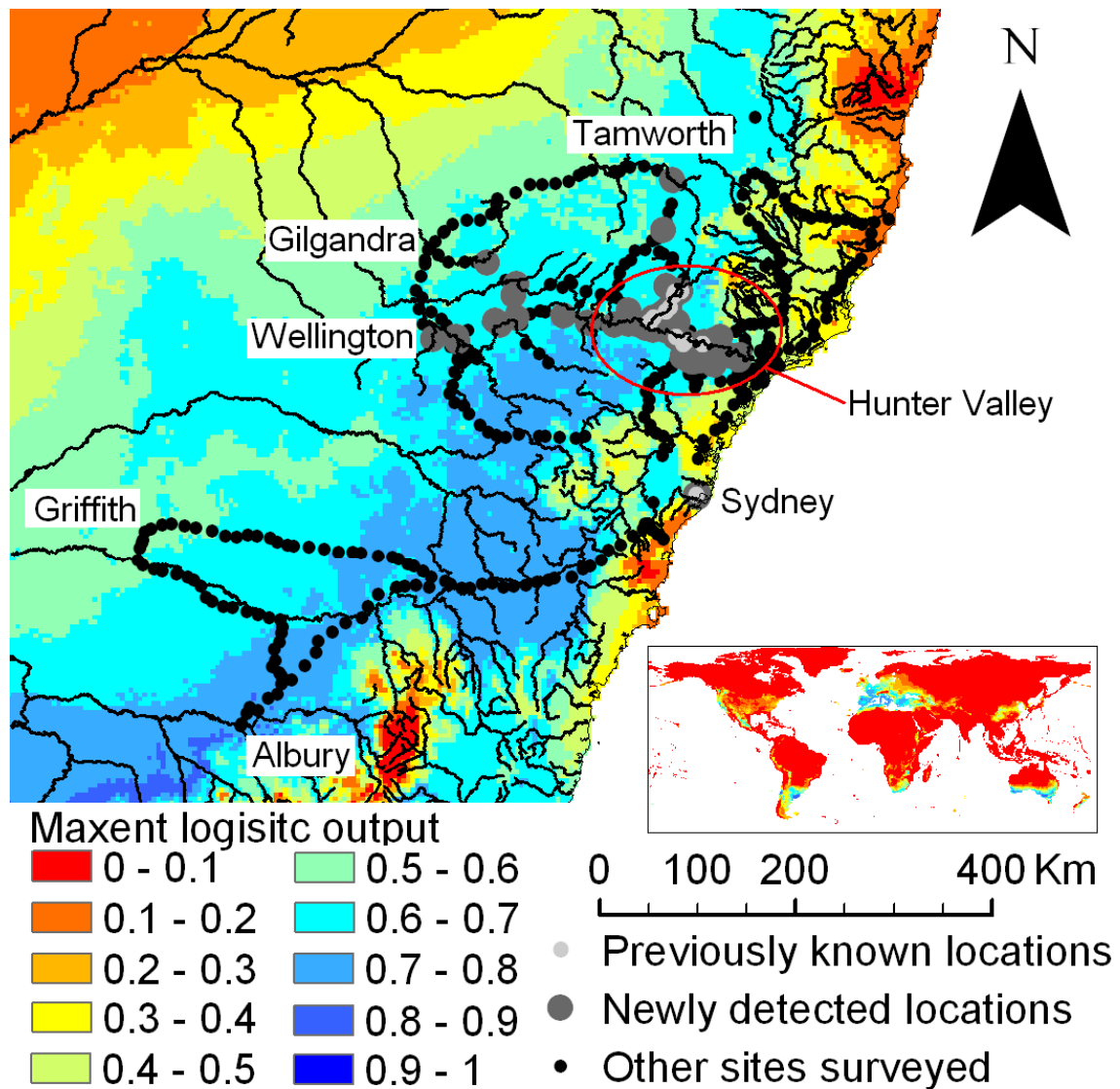


**Fig. 2** The six panels show the Maxent logistic outputs (climatic favourability) for south-east Australia based on the six models from the corresponding panels in Fig. 1. Grey circles illustrate the known locations of *Halictus smaragdulus* prior to our surveys, while the black dots illustrate the locations of the October 2008 survey focused on the areas that had highest climatic suitability in the six models (a), November 2008 survey focused on proximity to the west of known locations (b), December 2008 survey focused on the proximity to the east of the known locations (c), volunteer samples (d), March 2009 survey focused on sandy and weedy creeks and rivers (e) and February 2010 survey targeting areas with suitable habitat near the locations where the species was found in the initial surveys (f).





**Fig. 3** Model performance was assessed using the average AUC of ten cross-validation models. Training AUC was based on the native range data used to produce the models, while the cross validation AUC was calculated using native range data not used to produce models. The AUC for the Australian records was calculated separately for the known locations both before and after our surveys.



**Fig. 4** The dark grey circles illustrate the locations where *Halictus smaragdulus* has been observed in Australia as of July 2010, while the light grey circles represent the known locations prior to our surveys. Black lines indicate rivers. The Maxent logistic output shown was produced using all data in both the native and invaded ranges after our surveys, and was based on four commonly used predictors (see text).