Article:

Hassall, C (2012) Predicting the distributions of under-recorded Odonata using species distribution models. Insect Conservation and Diversity, 5 (3). 192 - 201 . ISSN 1752-458X

https://doi.org/10.1111/j.1752-4598.2011.00150.x

Predicting the distributions of under-recorded Odonata using species distribution models, Hassall (2012) - SELF-ARCHIVED COPY

http://onlinelibrary.wiley.com/doi/10.1111/j.1752-4598.2011.00150.x

HASSALL, C. 2012. Predicting the distributions of under-recorded Odonata using species distribution models. *Insect Conservation and Diversity*, 5**,** 192-201.

Predicting the distributions of under-recorded Odonata using species distribution models

Running title: Modelling distributions of Odonata

Author: Christopher Hassall

Affiliation: Carleton University

Address: Biology Department, Carleton University, Ottawa, ON, K1S 5B6, Canada

Telephone: +1 613 520 2600 (ext. 3866)

Fax: +1 613 520 3539

Email: chassall@connect.carleton.ca

ABSTRACT

1. Absences in distributional data may result either from the true absence of a species or from a false absence due to lack of recording effort. I use species distribution models both to elucidate this problem in North American Odonata and present a potential solution.
2. I use AIC model selection methods to evaluate the ability of water energy variables and measures of recorder effort to explain patterns of odonate diversity in the USA and Canada.
3. I model the ranges of 176 species that are found solely in the USA and Canada using model averaging of eight different methods. These give predictions of hypothetical "true" distributions of each of the 176 species, which I compare with observed distributions to identify areas where potential under-recording may be occurring.
4. Water-energy variables explain a large proportion of the variance in odonate diversity, but the addition of recorder effort significantly improves the models.
5. Under-recording appears to be highest in northern Canada, Alaska and Quebec and recording is higher in coastal areas of the USA than the interior. Levels of recording are inversely related to population density. Maps for individual species have been made freely available online via Odonata Central (http://www.odonatacentral.org/) to facilitate recording in the future.
6. This analysis has both illustrated a problem with current odonate recording in the form of unbalanced recorder effort. However, the SDM approach also provides the solution, targeting recorder effort in such a way as to maximise returns from limited resources.

Keywords: biological recording, damselfly, distributions, dragonfly, North America, Odonata, recorder effort, species distribution models.

INTRODUCTION

Datasets of distributional data have been used extensively for a number of purposes including global change biology (Hickling et al., 2006), conservation (Polasky et al., 2001), pest management (Worner & Gevrey, 2006) and fisheries (Perry et al., 2005). However, animals vary in their detectability and so in many cases the datasets that are available are for the more charismatic animals, particularly birds (e.g. Thomas & Lennon, 1999) and mammals (e.g. Andrews & O'Brien, 2000). Insect conservation ecology has been hampered by the relative under-reporting of insect sightings (Dunn, 2005), making distributional data patchy.

Attempts at collecting distributional data generally take one of two forms. The first form of recording is the standardised census. Examples of extensive recording of this type are rare, although the Butterfly Monitoring Scheme and a variety of aphid suction trap networks have accumulated a vast amount of such data. Standardised datasets have been used extensively in macroecological research (Harrington et al., 2007; Hill et al., 2002). With standardised methods the data from such surveys are relatively straightforward to analyse and trends can be obtained with a degree of certainty. The second, and historically more common, form of recording is the collection of reports of sightings. Due to the stochastic nature of record submission and the heavy reliance on the enthusiasm of the public both to observe and report specific taxa, these datasets tend to be extremely biased in time and space. Special analytical approaches are required to account for such biases and these can reduce the certainty associated with results (Hassall & Thompson, in press). In particular, for data that does not follow a standardised methodology such as the Pollard walk (Pollard & Yates, 1993), there is a great concern that variation in sampling effort will produce artifacts in the data. There is, therefore, an interest both in quantifying variation in recorder effort and in directing future recording effort in such a way as to maximise returns from limited resources (Murdoch et al., 2007).

Early work on the causes of patterns of diversity in such data proposed a large number of hypotheses, many of which were shown to be tautologies or unsupported by data (Rohde, 1992). Numerous factors have been proposed as influencing richness, including spatial heterogeneity (Roth, 1976; Tews et al., 2004), climatic variability (Knapp et al., 2002), environmental age (Montoya et al., 2007) and energy availability (Currie, 1991). Of all of these hypotheses, the relationship between the water-energy balance and diversity appears the best substantiated (Currie et al., 2004; Hawkins et al., 2003). These complementary factors are of particular significance, as water (precipitation) and energy (related to environmental temperature) are predicted to vary considerably under projections of global climate change. A link between water-energy balances and patterns of diversity has been demonstrated in a range of different vertebrate, invertebrate and plant taxa (for a review see Hawkins et al., 2003).

Recorder effort is rarely included in investigations of patterns of diversity at larger geographical scales. Previous work on odonate species richness in Europe (Keil et al., 2008) examined only biological and meteorological predictors of diversity, omitting any consideration of the human aspect of recording. However, the requirement of taking into account temporal variation in recorder effort which has become commonplace in global change analyses certainly extends to spatial variation. Furthermore, analyses of diversity in Europe pose a number of problems as a result of the region's history and topography (Whittaker et al., 2007). While conclusions drawn from Keil et al.'s work are convincing, the consistency of those conclusions remains to be tested.

One way in which spatial variation in recorder effort can be detected is by analysing the input to datasets. Recorder effort may be influencing species recording if the number of records being submitted from a given region is related to the number of species even when other covariates (such as water and energy) are accounted for. The alternative method for identifying spatial variation in recording is to look at the output of the dataset. In this case we can predict hypothetical requirements for species based on where they have been observed and predict geographical ranges where they may be present without having been observed.

Species distribution models (SDMs) use recorded species presence and the conditions associated with those records to infer unrecorded presence where conditions are similar (Elith & Leathwick, 2009). Aquatic insects, and the Odonata (dragonflies and damselflies) in particular, should be strongly influenced by water-energy variables. Odonata exhibit relationships between various aspects of their biology and temperature (Hassall & Thompson, 2008) and these suggest a tropical evolutionary history (Pritchard & Leggott, 1987), with temperate species possessing adaptations to cooler climates which permit the colonisation of habitat that is available during interglacial periods. Odonates rely strongly and nearly universally on persistent water bodies for breeding and larval habitats. Two previous studies have investigated the use of species distribution models for odonate distributions. The first investigated the distributions of 160 species of South African Odonata using the BIOCLIM method of distribution modelling due to its ease of implementation within GIS (Finch et al., 2006). However, this yielded over-predictions of generalists and species with distributional outliers. Finch *et al.* recommend the use of probabilistic modelling methods instead of BIOLCLIM for more accurate modelling. The second study used only two species, *Schistolobos boliviensis* (Daigle, 2007) and *Tuberculobasis inversa* (Selys, 1876) and was restricted to the tropics (De Almeida et al., in press). No studies have yet been conducted on North American Odonata, although this approach has been advocated previously as it may provide assistance in searching for previously unrecorded species in new areas (Samways in Bried & Mazzacano, 2010).

The Odonata have been highlighted as being particularly susceptible to both environmental warming (Hassall & Thompson, 2008) and water quality (Menetrey et al., 2005) as well as indicating general species richness (Sahlén & Ekestubbe, 2001). As such they could be potentially be considered "barometers" of environmental change. However, in order to be used as such, accurate data concerning their distributions is required. I demonstrate that geographical variation in recorder effort affects the distributional recording of North American Odonata in addition to water-energy variables. By comparing SDMs of 176 North American Odonata with their observed distributions, I highlight regions that may be more species poor than those models suggest. I further provide maps for individual species such that recorders may search for particular species in a more focused way.

METHOD

Data sources

Odonata distributions

The distributions of North American Odonata species were determined using the Odonata Central database (Abbott, 2007). All records in the database are verified by an expert before being entered and the dabatase is actively managed to ensure accuracy. The database was cleaned to remove all sightings of species outside Canada and the USA and any records lacking geographical data, leaving 109,674 records of 417 species. Temporal information was only available for 10% of records, so controlling for variation in time was not possible. However, of the dates available the majority (77%) were from between 1990 and 2009. Distributional data were in the format of presence/absence at a county-level in the USA and on a 0.5 degree latitude-longitude grid in Canada. A two-degree latitude-longitude grid was superimposed on the geographical data and presence/absence calculated for each cell. To account for the difference in grid cell size between latitudes, the area of each cell was entered into all models described below. Presences for all species were summed for each cell to give a total species richness. With a dataset that has not been collected systematically, as is the case here, the assumption has to be made that recorder effort between cells is even. While this is highly unlikely given the variations in both accessibility of some grid cells and the variation in human population densities to provide sightings, it is likely that all grid cells have been visited. For example, there exist atlases of Odonata from Ontario, Alaska, the Northwest Territories and the Yukon. While the territory of Nunavut has received relatively little attention, only the southern part of this region is included in this analysis.

Water energy

Global Vegetation Index (GVI) data were obtained from The United Nation Environment Program (UNEP, http://www.grid.unep.ch/, data code GNV179). The data are 10 categories ranging from 1 (highest GVI class – most green) to 10 (lowest GVI class – least green) and represent the maximum GVI values averaged over the years 1983-1990. GVI data originally comprised a raster dataset with a resolution of 0.5 degrees and were averaged for each cell in the two-degree grid. Potential evapotranspiration (PET) and actual transpiration (AET) were also obtained from the UNEP (dataset GNV183). These datasets both have a resolution of 0.5 degrees and values were averaged for each two-degree cell. Climate data (mean annual precipitation, mean annual temperature and mean summer (June, July, August) temperature) were taken from the Climate Research Unit (CRU) TS2.10 dataset averaged between 1991 and 2000. This dataset is available at a 0.5 degree resolution and was averaged for each two-degree grid cell. The area of freshwater in each cell was calculated from the GlobCover map version 2.2 (http://ionia1.esrin.esa.int/), which has a grid cell size of 300m but is ground-truthed against the 30m resolution Shuttle Radar Topography Mission Water Body Dataset (SWDB). The proportion of 300m cells containing water was calculated for each two-degree grid cell. Altitude and altitudinal range were extracted from the WORLDCLIM dataset (http://www.worldclim.org/). WORLDCLIM is available at a resolution of 30 arc-second and was averaged for each two-degree cell. Twenty-one cells lacked GVI data and only those cells entirely on land were included in the analysis to reduce errors associated with maritime climates, leaving 517 cells altogether (Fig. 1).

Recorder effort and population density

Recorder effort was determined by the number of records submitted by individuals (i.e. not part of the Dot Map Project which forms the bulk of the dataset (Donnelly, 2004a, b, c) or records from museum specimens). North American population density data gridded at 2.5' were obtained for 2000 from the Socioeconomic Data and Applications Centre (SEDAC, http://sedac.ciesin.columbia.edu/gpw/). Values were averaged for each of the grid cells.

Describing diversity

Multicollinearity exists between environmental predictors, making stepwise model selection potentially unreliable. To circumvent this problem, an exhaustive model selection procedure was carried out using the regsubsets function in the leaps package (Lumley, 2009) in R (R Development Core Team, 2009). This function calculates the residual sum of squares for each possible model and yields the best model of each size from 1 to 10 variables, thus there is no penalty for model size at this point. The best model can be found by comparing the Schwartz criterion for each of the 10 resulting models. Species richness in each cell was the response variable and the following predictor variables (and their quadratic terms) were entered into the selection process: mean annual temperature, mean summer temperature, mean annual precipitation, altitudinal range, mean altitude, PET, AET, GVI and area of freshwater. The area of the cell was included in all models. Due to the potentially confounding effects of spatial autocorrelation, a simultaneous autoregression (SAR) model was used to analyse the selected model in SAM v3.0 (Rangel et al., 2006).

In addition to this model, four general linear models were constructed: (i) a null model, (ii) water-energy model (WE) as described above, (iii) water-energy + population density(WE+PD), (iv) water-energy + recorder effort(WE+RE), (v) water energy + population density + recorder effort (WE+PD+RE). Due to the variation in size between grid cells which are equal latitude-longitude but not equal area, an area term (in km^2) was included in every model. Models were compared using ΔAIC and Akaike weights. A ΔAICc value between 0 and 2 corresponds to substantial empirical support, a value between 4 and 7 corresponds to considerably less support and when ΔAICc > 10 there is essentially no support for that model (Burnham & Anderson, 2002).

Species distribution models

Only species whose ranges fall entirely within the USA and Canada (range locations were confirmed using Needham et al. (2000) and Westfall and May (1996)) were selected for predictive modelling. Species were further excluded if they were present in fewer than 20 grid cells to ensure accurate model fitting leaving a total of 176 species for the analysis. The BIOMOD package (Thuiller et al., 2009) was used in R to carry out the modelling. This package uses an ensemble approach to quantify variation inherent within the different predictive models and produces a probability of occurrence based on the consensus of eight modelling methods (classification tree analysis (CTA), general additive models (GAM), generalised boosting regression (GBM), generalised linear models (GLM), multiple adaptive regression splines (MARS), mixture discriminant analysis (MDA), random forests (RF) and surface range envelope (SRE)). Artificial neural networks are also available in the program, but this modelling approach performs poorly where datasets have few preferences as was the case with some species in this analysis.

Two sets of pseudo-absences were generated for each species for each model. All grid cells that were outside of a radius of 330km from each presence were selected as absences. Models were calibrated using 70% of the data selected at random for each species and evaluated using the remaining 30% (Guisan & Thuiller, 2005). This process was repeated three times and then the final model was calibrated on 100% of the data (see Thuiller et al., 2009 for details). Each of the eight models produced a probability of occurrence and this probability was averaged across models. Methods to determine objective thresholds for presence from probability / suitability data have recently been reviewed (Liu et al., 2005). One of the methods recommended by this review uses the mean predicted probability / suitability from the output as the threshold (Cramer, 2003) with any probabilities above this mean being regarded as "present" and this is the method applied here. Model prediction accuracy was assessed using the average area under curve (AUC), Kappa and true skill statistic (TSS) values for the eight modelling methods. A "good" or "high/excellent" predictive ability is indicated when AUC >0.8 or Kappa or TSS >0.6 (Thuiller et al., 2009).

The total number of species for which the climate space was predicted to be suitable was calculated for each cell. The number of species actually observed in those cells was also calculated from the Odonata Central dataset. The proportion of species that are predicted but not observed is a potential indicator of the potential need to invest greater recorder effort in an area. Maps of both the regions where there may be under-recording of Odonata in general and maps for individual species were produced.

RESULTS

Present diversity

Species richness in North American Odonata does not conform to the usual latitudinal gradient that is seen as a general global pattern. Instead the majority of odonate species are found in the south-east of the continent (Fig. 1), with richness decreasing both northward as temperature declines and westward as water availability declines. This is the same pattern that has been established for trees (Currie & Paquin, 1987) and amphibians (Currie, 1991). As has been found for European and North African Odonata, as well as a range of other taxa, water-energy balances contributed the most to determining patterns of species richness in North American Odonata. Exhaustive variable selection resulted a model containing the terms PET, GVI and mean annual precipitation as well as the area which was included by force (area, $t=0.098$, $p=0.922$, PET, $t=9.887$, $p<0.001$, GVI, $t=-15.159$, $p<0.001$, precipitation, $t=4.543$, $p<0.001$, $R^2=0.770$).

The high explanatory power of the model excluding spatial autocorrelation corresponds well with that found in Europe ($R^2=0.79-0.81$, Keil et al., 2008). Partial correlations show that higher PET, lower GVI (i.e. more greenery) and higher precipitation are related to greater species richness in this taxon in this region (see Fig. 2 for partial plots). Incorporating spatial autocorrelation using SAR did not alter the results of the model and all variables were still significant, but explanatory power was reduced (area, $t=0.123$, $p=0.902$, PET, $t=8.015$, $p<0.001$, GVI, $t=-14.351$, $p<0.001$, precipitation, $t=2.563$, $p=0.011$, $R^2=0.555$).

The addition of population density (PD) and recorder effort (RE) resulted in models with substantially greater support than the water-energy (WE) model ($\Delta AICc \approx 4$, Table). However, the models WE+PD+RE and WE+RE were statistically indistinguishable with Akaike weights of 0.46 and 0.41, respectively. The model WE+PD yielded a model with considerably less support, suggesting that it is recorder effort which is affecting the model more so than population density.

Projected climate space

Model performance statistics showed that almost all models performed well in describing current species distributions (for an example see Fig. 3). Out of 176 species that were modelled, models for 163 (Kappa), 167 (TSS) and 172 (AUC) species were classified as having a "good" or "excellent" fit, using the criteria outlined above (Fig. 4). Thus I can be confident that the projections will not be hindered by poor predictive ability of the models.

Despite the very good performance of the models, there were a number of cases in which presence was predicted but not observed in the data. 168 out of 176 species were predicted to have larger ranges than were recorded in the dataset. Assuming that this is a result of under-recording, there are a number of factors that could contribute. Taxonomic confusion and small body size may affect the level of recording of individual species. The number of species and the mean difference between recorded and predicted range size (in grid cells) was calculated was calculated for each genus and a Pearson correlation used to test for an association. Body size was calculated as the first principal component (explaining 94.8% of the variance) from maximum and minimum values for abdomen length, total body length and hind wing length in males (data from Needham et al., 2000; Westfall & May, 1996) with Pearson correlation was used to test for an association. Neither congener number (r=-0.173, p=0.268, n=43) nor body size (r=0.052, p=0.500, n=176) showed any association with the difference between predicted and observed range size.

The patterns of under-recording appear to be focused in Canada, particularly in the boreal regions (Fig. 5). The Yukon is the exception, with a large proportion of the predicted species having been recorded. The Northwest Territories stands out as being of greatest interest in terms of recording as it has a number of species predicted to occur but very few actually recorded. A large area of northern Quebec and Alaska are similarly poorly recorded. Aside from these boreal regions, there are other areas of interest. While much of the eastern half of the USA appears well-recorded, Missouri, Nebraska and South Dakota are less well-recorded. Both boreal regions and the Great Plains are known to suffer from a lack of recording (T.W. Donnelly, *pers. comm.*). On the west coast, there are a number of species for which potentially viable climate space occurs but these regions are geographically isolated from their range on the east coast. As might be expected, there is a significant relationship between the number of species predicted but not observed in a grid cell and (log transformed) human population density (r=-0.417, p<0.001, data summarised in Fig. 6).

DISCUSSION

Predicting distributions

Distributional datasets comprising anecdotal reports of sightings frequently have uneven distributions of records in space and time. Here I demonstrate two pieces of evidence for this for North American Odonata recording and the single largest dataset dedicated to that endeavour: (i) odonate diversity is related not only to water energy variables as seen in previous analyses but also to the number of records submitted from each area, and (ii) some areas are more species-poor than SDMs suggest based on local climate. It is my intention through this work to provide a useful resource for Odonata recording that may help to investigate and remedy this problem. I provide both a guide to regions which may be under-recorded as well as maps of individual species distributions which will be made available online.

Variations in the degree of recording of taxa in general may be related to any number of factors that are difficult to quantify. The accessibility of a site, the number of visitor attractions in a region or, as tested in this study, the population density of a given area. There is even a suggestion that the recorded species richness of 19th Century England was related to the density of Anglican clergymen.

Such a diverse potential range of factors affecting recording requires the inclusion of at least population density if not a measure of recorder effort. For individual species, there have been issues involving taxonomic confusion (Cotton, 1981) and taxonomic "charisma" (Dunn, 2005). A consideration of potential confusion and body size did not reveal any patterns, but there may still be species traits that affect "recordability".

Using the maps produced in this study, it is possible to predict which species would be found in those regions which have been surveyed least. While atlas data do exist for Alaska and the Northwest Territories, a systematic survey of habitats in these regions has not yet been undertaken. Similarly, Quebec has neither atlas nor survey work, particularly in the north, and lacks even a field guide to the species found there. These boreal surveys would not only test the predictions set out in the models but also assist in cataloguing a fauna occupying the region which will experience twice the global mean warming over the next 50 years (Intergovernmental Panel on Climate Change (IPCC), 2001).

In addition to surveying more remote areas, there is a need to fill gaps in recorder effort in more accessible regions. The west coast of North America and further inland from the east coast are relatively well recorded, but there are regions of Nebraska, South Dakota and Idaho as well as southern Manitoba and Saskatchewan which are highlighted in this analysis as being under-recorded. As well as surveying, there are no doubt collections in museums and universities which have not been digitised and submitted. This kind of work is essential for investigating the present distributions but also for investigating past changes where specimens are above a certain age.

Patterns of diversity

Water-energy patterns have been analysed in a wide range of taxa using data that arise from a variety of sources . The North American data on odonate distributions, like the European data on odonate distributions which as previously been studied (Keil et al., 2008), is the result of the merging of museum and survey data with records submitted by the public. While the overall patterns of water-energy and richness in North America concur with the findings of Keil et al. (with a similar level of explanatory power) from Europe the addition of recorder effort still improves the model considerably, suggesting that this is having an effect on the distributional data. It is advisable to quantify recorder effort for future analyses.

Factors affecting SDM accuracy

The high explanatory power of the SDMs is not surprising given aforementioned relationships between temperature and water and odonate biology. The strength of the fit suggests that such models could be applied in other geographical locations. However, SDMs predicted much larger ranges for many species than have been observed. This may be due to recorder effort, but there are additional factors which must be borne in mind. The climate space that was predicted may occupy two discrete geographical locations. In particular, the species which are found on the eastern part of North America were occasionally predicted to have suitable climate space on the west coast where patterns of precipitation and temperature were similar. However, these two areas are separated by the Great Plains region which bisects the continent and prevents colonisation. Increasing precipitation is predicted for the northern Great Plains region, while decreasing precipitation is predicted for the south (Karl et al., 2009). Increasing precipitation may facilitate movements across the continent by providing additional habitat.

While biological predictors are almost always omitted from SDMs due to their potential colinearity with climate variables (Elith & Leathwick, 2009), although exceptions occur when host species are required (Araújo & Luoto, 2007). Odonata have no such strong biological interactions, being both generalist predators and preyed on by a range of species in return. Within taxonomic groups, a variety of species traits have previously been shown to influence the accuracy of SDMs, and this will be explored in a later publication.

Conclusion

Species distribution models have a variety of applications including predicting responses to climate change (Lawler et al., 2006), evolutionary biology (Graham et al., 2004) and conservation biology (Guisan & Thuiller, 2005). Here I apply them to supplementing existing distributional data by predicting regions and particular species that may be under-recorded. As well as highlighting the potential problem, these models also indicate how best to solve it along with providing testable predictions for species presence in rarely surveyed regions.

ACKNOWLEDGEMENTS

I am extremely grateful to John Abbott for permitting me to use the Odonata Central dataset and for advising on the nature of the data. Thanks also to those diligent recorders who have made the dataset such a valuable resource. Conversations with Nick Donnelly substantially affected the direction of the study. Jeremy Kerr and Jay Fitzsimmons provided helpful advice on an earlier version of the manuscript. Funding was provided by a Government of Canada Postdoctoral Fellowship.

REFERENCES

Abbott, J.C. (2007) OdonataCentral: An online resource for the distribution and identification of Odonata. Texas Natural Science Center, The University of Texas at Austin, available at http://www.odonatacentral.org. (Accessed: April 01, 2010).

Andrews, P. & O'Brien, E. (2000) Climate, vegetation, and predictable gradients in mammal species richness in southern Africa. *Journal of Zoology*, **251**, 205-231.

Araújo, M.B. & Luoto, M. (2007) The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography*, **16**, 743-753.

Bried, J.T. & Mazzacano, C.A. (2010) National review of state wildlife action plans for Odonata species of greatest conservation need. *Insect Conservation and Diversity*, **3**, 61-71.

Burnham, K.P. & Anderson, D.R. (2002) Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd edn. Springer-Verlag, New York.

Cotton, D.C.F. (1981) Some new records and an appraisal of the published records for *Ischnura pumilio* (Charpentier) (Odonata: Coenagrionidae) in Ireland. *Entomologist's Gazette*, **32**, 59-64.

Cramer, J.S. (2003) Logit models: from economics and other fields Cambridge University Press, Cambridge.

Currie, D.J. (1991) Energy and large-scale patterns of animal and plant species richness. *American Naturalist*, **137**, 27-49.

Currie, D.J., Mittelbach, G.G., Cornell, H.V., Kaufman, D.M., Kerr, J.T., Oberdorff, T., & Guegan, J.-F. (2004) A critical review of species-energy theory. *Ecology Letters*, **7**, 1121-1134.

Currie, D.J. & Paquin, V. (1987) Large-scale biogeographical patterns of species richness of trees. *Nature*, **329**, 326-327.

De Almeida, M.C., Côrtes, L.G., & De Marco Júnior, P. (in press) New records and a niche model for the distribution of two Neotropical damselflies: *Schistolobos boliviensis* and *Tuberculobasis inversa* (Odonata: Coenagrionidae). *Insect Conservation and Diversity*.

Donnelly, T.W. (2004a) Distribution of North American Odonata. Part I: Aeshnidae, Petaluridae, Gomphidae, Cordulegastridae. *Bulletin of American Odonatology*, **7**, 61-90.

Donnelly, T.W. (2004b) Distribution of North American Odonata. Part II: Macromiidae, Corduliidae and Libellulidae. *Bulletin of American Odonatology*, **8**, 1-32.

Donnelly, T.W. (2004c) Distribution of North American Odonata. Part III: Calopterygidae, Lestidae, Coenagrionidae, Protoneuridae, Platystictidae, with data sources and bibliography, parts 1-III. *Bulletin of American Odonatology*, **8**, 33-99.

Dunn, R.R. (2005) Insect extinctions, the neglected majority. *Conservation Biology*, **19**, 1030-1036.

Elith, J. & Leathwick, J.R. (2009) Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 677-697.

Finch, J.M., Samways, M.J., Hill, T.R., Piper, S.E., & Taylor, S. (2006) Application of predictive distributions modeling to invertebrates: Odonata in South Africa. *Biodiversity and Conservation*, **15**, 4239-4251.

Graham, C.H., Ron, S.R., Santos, J.C., Schneider, C.J., & Moritz, C. (2004) Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution*, **58**, 1781-1793.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, **8**, 993-1009.

Harrington, R., Clark, S.J., Welham, S.J., Verrier, P.J., Denholm, C.H., Hullé, M., Maurice, D., Rounsevell, M.D., & Cocu, N. (2007) Environmental change and the phenology of European aphids. *Global Change Biology*, **13**, 1550-1564.

Hassall, C. & Thompson, D.J. (2008) The impacts of environmental warming on Odonata: a review. *International Journal of Odonatology*, **11**, 131-153.

Hassall, C. & Thompson, D.J. (in press) Accounting for recorder effort in the detection of range shifts from historical data. *Methods in Ecology and Evolution*.

Hawkins, B.A., Field, R., Cornell, H.V., Currie, D.J., Guegan, J.-F., Kaufman, D.M., Kerr, J.T., Mittelbach, G.G., Oberdorff, T., O'Brien, E.M., Porter, E.E., & Turner, J.R.G. (2003) Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, **84**, 3105-3117.

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**, 1-6.

Hill, J.K., Thomas, C.D., Fox, R., Telfer, M.G., Willis, S.G., Asher, J., & Huntley, B. (2002) Responses of butterflies to twentieth century climate warming: Implications for future ranges. *Proceedings of the Royal Society - Biological Sciences (Series B)*, **269**, 2163-2171.

Intergovernmental Panel on Climate Change (IPCC) (2001) Climate Change 2000—Third Assessment Report Cambridge University Press, Cambridge.

Karl, T.R., Melillo, J.M., & Peterson, T.C., eds. (2009) Global Climate Change Impacts in the United States. Cambridge University Press, Cambridge.

Keil, P., Simova, I., & Hawkins, B.A. (2008) Water-energy and the geographical species richness pattern of European and North African dragonflies (Odonata). *Insect Conservation and Diversity*, **1**, 142-150.

Knapp, A.K., Fay, P.A., Blair, J.M., Collins, S.L., Smith, M.D., Carlisle, J.D., Harper, C.W., Danner, B.T., Lett, M.S., & McCarron, J.K. (2002) Rainfall Variability, Carbon Cycling, and Plant Species Diversity in a Mesic Grassland. *Science*, **298**, 2202-2205.

Lawler, J.J., White, D., Neilson, R.P., & Blaustein, A.R. (2006) Predicting climate-induced range shifts: model differences and model reliability. *Global Change Biology*, **12**, 1568-1584.

Liu, C., Berry, P.M., Dawson, T.P., & Pearson, R.G. (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, **28**, 385-393.

Lumley, T. (2009) leaps: regression subset selection (using Fortran code by Alan Miller), http://CRAN.R-project.org/package=leaps.

Menetrey, N., Sager, L., Lachavanne, J.B., & Oertli, B. (2005) Looking for metrics to assess the trophic state of ponds. Macroinvertebrates and amphibians. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **15**, 653-664.

Montoya, D., Rodriguez, M.A., Zavala, M.A., & Hawkins, B.A. (2007) Contemporary richness of holarctic trees and the historical pattern of glacial retreat. *Ecography*, **30**, 173-182.

Murdoch, W.W., Polasky, S., Wilson, K.A., Possingham, H.P., Kareiva, P., & Shaw, R. (2007) Maximising return on investment in conservation. *Biological Conservation*, **139**, 375-388.

Needham, J.G., Westfall, M.J., & May, M.L. (2000) Dragonflies of North America, revised edition Scientific Publishers, Gainesville, FL.

Perry, A.L., Low, P.J., Ellis, J.R., & Reynolds, J.D. (2005) Climate change distribution shifts in marine fishes. *Science*, **308**, 1912-1915.

Polasky, S., Csuti, B., Vossler, C.A., & Meyers, S.M. (2001) A comparison of taxonomic distinctness versus richness as criteria for setting conservation priorities for North American birds. *Biological Conservation*, **97**, 99-105.

Pollard, E. & Yates, T.J. (1993) Monitoring butterflies for ecology and conservation Chapman and Hall, London.

Pritchard, G. & Leggott, M. (1987) Temperature, incubation rates and the origins of dragonflies. *Advances in Odonatology*, **3**, 121-126.

R Development Core Team (2009) R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria.

Rangel, T.F.L.V.B., Diniz-Filho, J.A.F., & Bini, L.M. (2006) Towards an intergrated computational tool for spatial analysis in macroecology and biogeography. *Global Ecology and Biogeography*, **15**, 321-327.

Rohde, K. (1992) Latitudinal gradients in species diversity: the search for the primary cause. *Oikos*, **65**, 514-527.

Roth, R., R. (1976) Spatial heterogeneity and bird species diversity. *Ecology*, **57**, 773-782.

Sahlén, G. & Ekestubbe, K. (2001) Identification of dragonflies (Odonata) as indicators of general species richness in boreal forest lakes. *Biodiversity and Conservation*, **10**, 673-690.

Tews, J., Brose, U., Grimm, V., Tielbörger, K., Wichmann, M.C., Schwager, M., & Jeltsch, F. (2004) Animal species diversity driven by habitat heterogeneity/diversity: the importance of keystone structures. *Journal of Biogeography*, **31**, 79-92.

Thomas, C.D. & Lennon, J.J. (1999) Birds extend their ranges northwards. *Nature*, **399**, 213.

Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M.B. (2009) BIOMOD - a platform for ensemble forecasting of species distributions. *Ecography*, **32**, 369-373.

Westfall, M.J. & May, M.L. (1996) Damselflies of North America Scientific Publishers, Gainesville, FL.

Whittaker, R.J., Nogués-Bravo, D., & Araújo, M.B. (2007) Geographical gradients of species richness: a test of the water-energy conjecture of Hawkins et al. (2003) using European data for five taxa. *Global Ecology and Biogeography*, **16**, 76-89.

Worner, S.P. & Gevrey, M. (2006) Modelling global insect pest species assemblages to determine risk invasion. *Journal of Applied Ecology*, **43**, 858-867.

Tables

Table 1 – Comparison of models explaining geographical variation in odonate diversity. Area=area of the grid cells, WE=water energy variables, RE=recorder effort, PD=population density.

Model	k	AICc	ΔAIC	Akaike weight
Area+WE+PD+RE	8	4619.30	0.00	0.46
Area+WE+RE	7	4619.55	0.25	0.41
Area+WE	6	1623.23	3.93	0.06
Area+WE+PD	7	4623.30	4.00	0.06
Area	3	5027.73	408.44	0.00

Figure legends

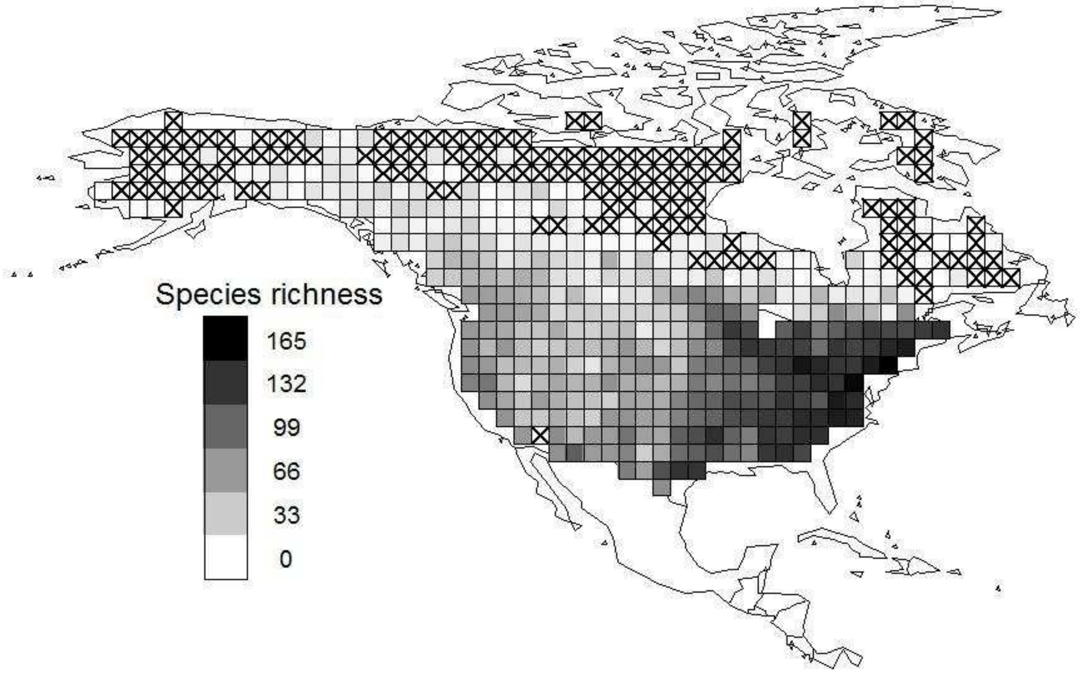

Fig 1 – Recorded diversity of North American Odonata (457 species in total). "X" indicates no species recorded in the Odonata Central dataset.

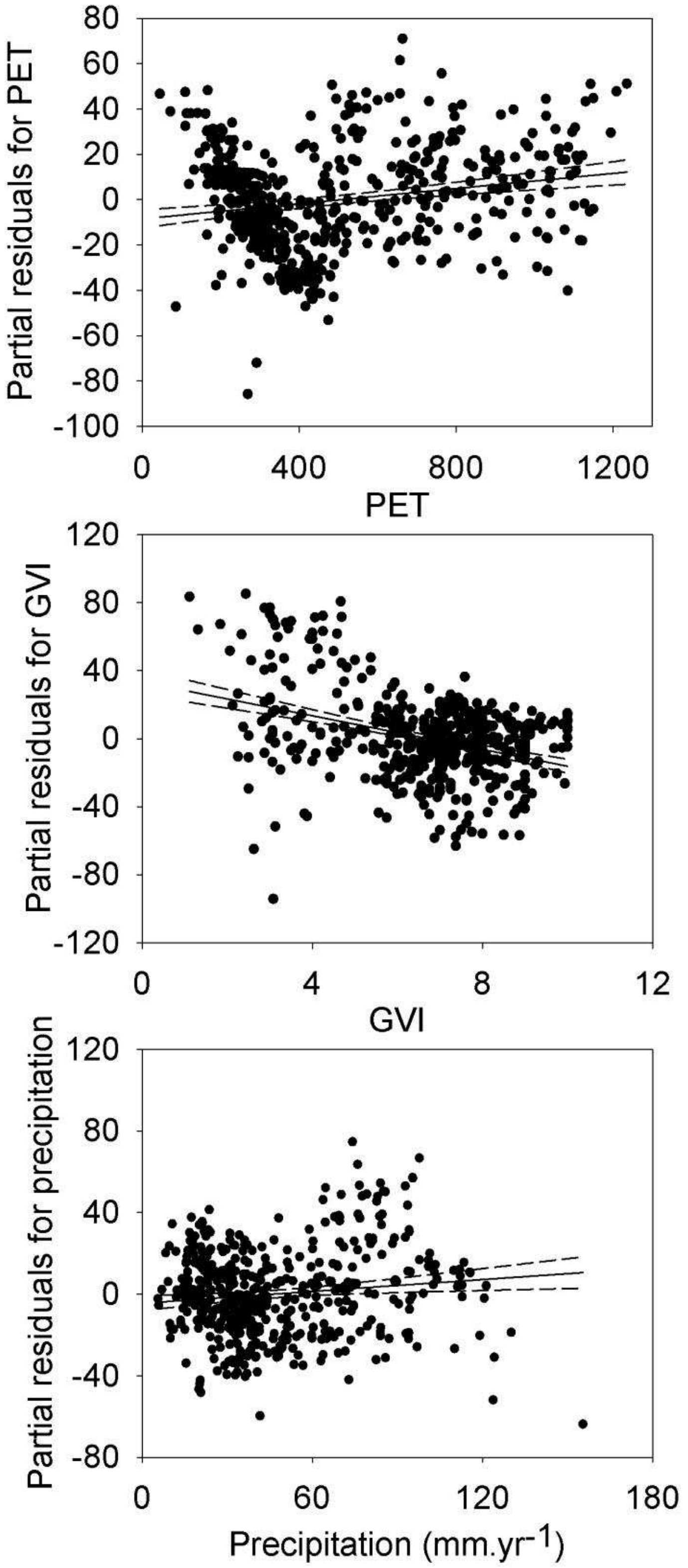

Fig. 2 – Partial correlations of odonate species richness in North America against potential evapotranspiration (PET, top), global vegetation index (GVI, middle) and precipitation (bottom).

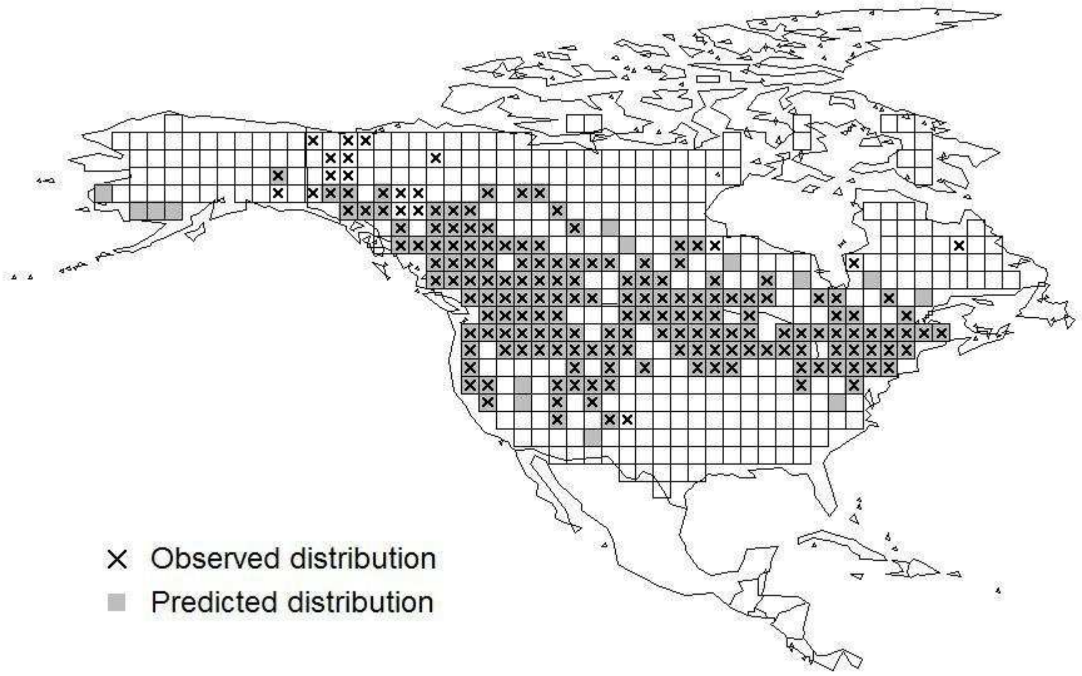

Fig 3 – Example of a climate space map showing predicted and observed distribution for *Coenagrion resolutum* (Hagen in Selys, 1876). Maps for all 176 modelled species will be made available online at http://www.odonatacentral.org/.

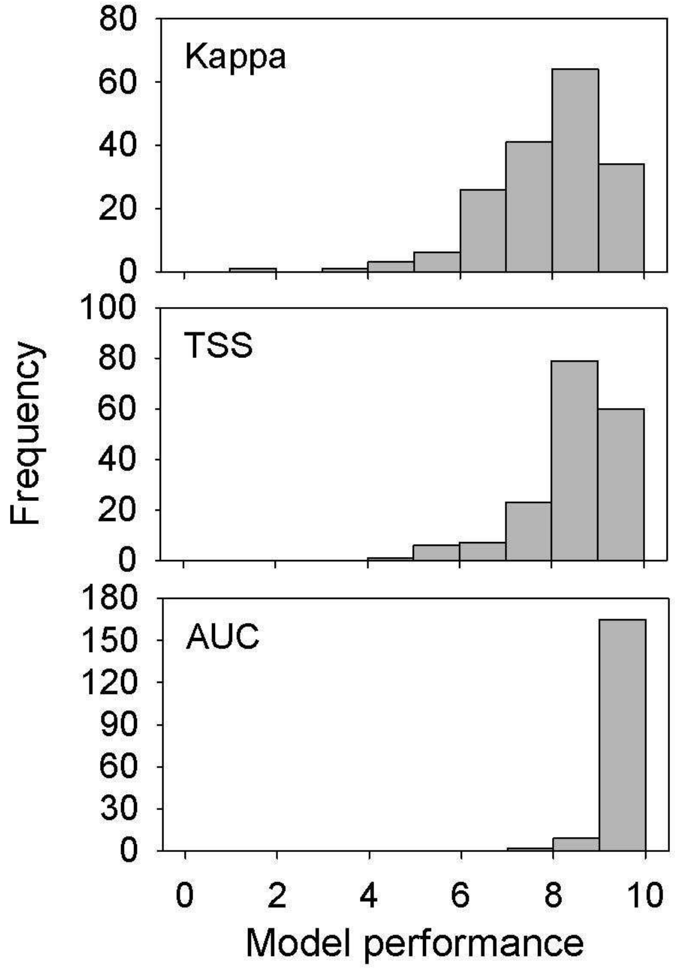

Fig 4 – Goodness-of-fit statistics for species distribution models describing the ranges of 176 species of North American Odonata.

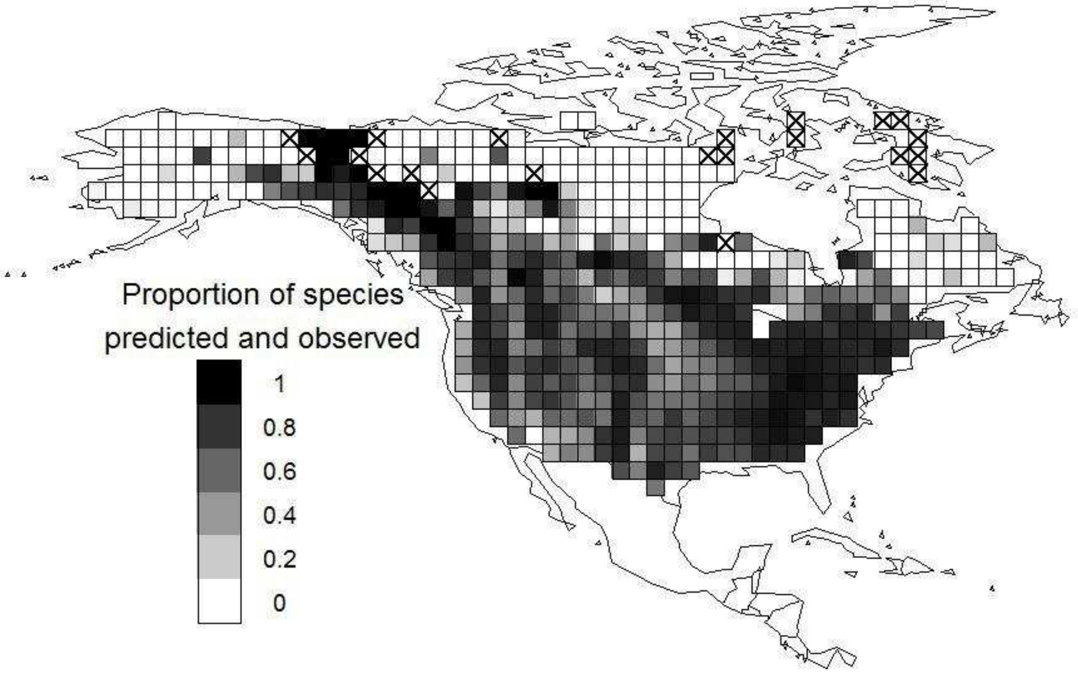

Fig 5 – Geographical variation in the proportion of predicted species observed in each grid cell – a value of "1" suggests that all species that are predicted by climate models have been recorded while a value of "0" suggests that no species that were predicted have been recorded. "X" indicates no species predicted.

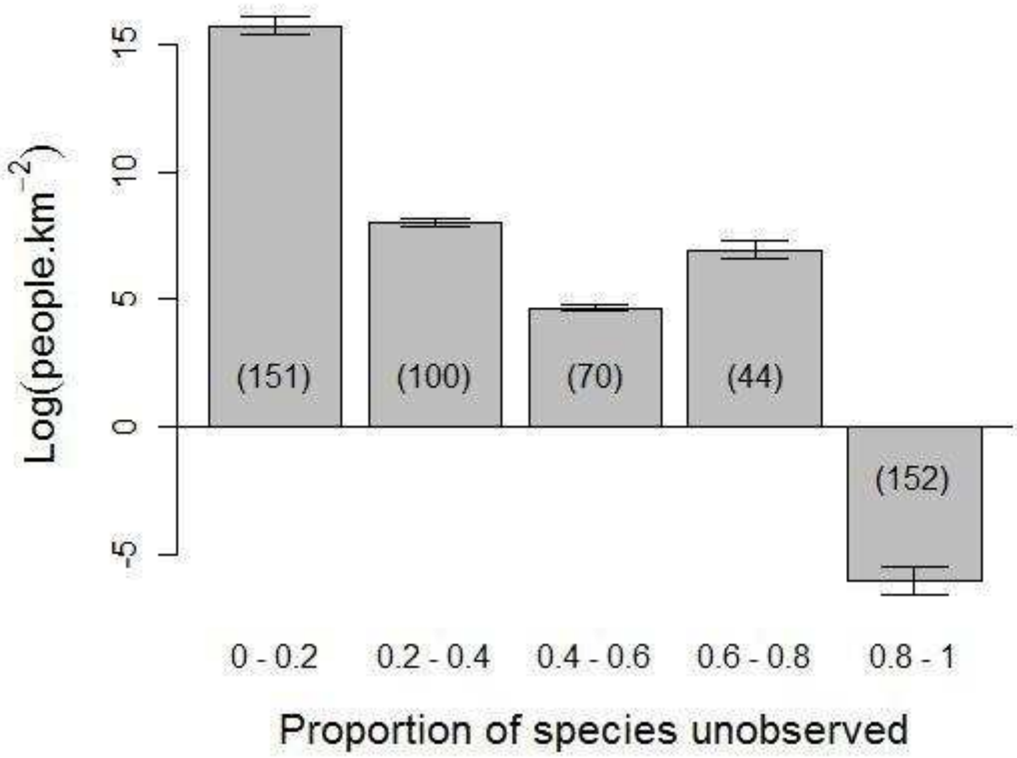

Fig 6 – The relationship between the number of species of Odonata predicted but unobserved and population density across North America. Numbers in brackets are numbers of cells in each category (total=517).