

Antragstellerin/Antragsteller:

Lukas Prader

Akademischer Grad, Vorname/n, Nachname

Hocheggstraße 720, 6100 Seefeld in Tirol

Anschrift: Straße, PLZ und Ort

lukas.prader@tutanota.com, +43 650 6533276

E-Mail, Telefonnummer

Matrikelnummer:

12115058

Kennzeichnung des Studiums

UC 033 630

Einreichung der Bachelorarbeit

(gem. § 80 Universitätsgesetz 2002 – UG)

Titel der Bachelorarbeit:

Studying SDM performance throughout a time series:

A case study using the invasive species Harmonia axyridis

Angaben zur Lehrveranstaltung (im Rahmen welcher die Bachelorarbeit verfasst und positiv beurteilt worden ist):

SE Biologisches Seminar: Ökologie

Titel der Lehrveranstaltung

743411

Lehrveranstaltungsnummer

4

Semester

Univ.-Prof. Mag. Dr. Gabriel Singer

Name der Lehrveranstaltungsleiterin / des Lehrveranstaltungsleiters

Ich erkläre hiermit an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Alle Stellen, die wörtlich oder inhaltlich den angegebenen Quellen entnommen wurden, sind als solche kenntlich gemacht.

Ich stimme einer Archivierung der Bachelorarbeit am Institut / an der Fakultät zu:

 Ja Nein

Datum

Unterschrift der/des Studierenden

I, Lukas Prader, agree that my bachelor's thesis ("Studying SDM performance throughout a time series: A case study using the invasive species *Harmonia axyridis*"), mainly supervised by Dr. Lauren Talluto, will be digitally archived in a repository of Leopold Franzens University Innsbruck and made available for further academic use.

Innsbruck, 29th March 2024

Signature

University of Innsbruck

Faculty of Biology

Department of Ecology



Bachelor Thesis
submitted for the degree of
Bachelor of Science

**Studying SDM performance throughout a
time series: A case study using the invasive
species *Harmonia axyridis***

by
Lukas Prader
Matriculation Nr.: 12115058
SE Biological Seminar: Ecology

Submission Date: 29th March 2024
Supervisors: Lauren Talluto, Gabriel Singer

Abstract

In invasive ecology, a lot of research is aimed at predicting the potential impact that a species could have outside its native range. Species Distribution Models (SDMs) have been applied in this context already, but with still ongoing discussion of their actual viability given the dynamic aspect of niches in the process of invasion. To further insight into the applicability of SDMs for predicting invasive species, a SDM ensemble time series modelling the European niche of *Harmonia axyridis* over the time period 2002-2020 was created. In addition, niche dynamic analyses were conducted, computing the niche stability and expansion over time among other parameters. It was shown that SDMs trained on data only from the invaded range provide consistently high TPR (True Positive Rate), and this in the context of an almost completely stable niche over the whole time period. Models with data only from the invaded range, as well as models using native data only are consistent with previous work predicting the range of *H. axyridis*.

Keywords: Species Distribution Model, invasive species, ensemble model, time series, niche dynamic analysis, ecospat

Contents

1	Introduction	1
2	Materials and Methods	4
2.1	Datasets	4
2.2	Data preparation	4
2.3	Model building	5
2.4	Analysis	7
3	Results	8
3.1	Temporal change of data availability	8
3.2	Niche dynamic analysis	8
3.3	SDMs, ensemble and time series	10
4	Discussion	13
5	Conclusion	14
6	Acknowledgements	14
A	Visualization of cleaned vs raw dataset	19
B	Niche analysis	19
C	Species distribution modelling	22
D	Model predictions	25
E	Model response curves	26

1 Introduction

Invasive species are of special interest in ecological research due to their impact on native ecosystems. Main goals in this area are to find out which species have potential to become invasive, what habitat will be susceptible to invasion by those species, how fast the species will invade the new area and what impact its invasion will have on the native ecosystem (Shigesada et al. 1997). To this end, many theories have been created to describe invasion processes. The invasion of a species can generally be described with four stages (Blackburn et al. 2011):

1. Transport: Leaving the native range, arriving at a new location
2. Introduction: Existing in specific locations (captivity / cultivation)
3. Establishment: Existing outside of areas of introduction in the wild
4. Spread: Sustaining establishment and dispersing to new environments

Depending on the current stage there can be significant differences in behaviour and impact of a species. The impacts of invasive alien species can be numerous, ranging from food web changes to reductions in habitat and species richness, hydrology and nutrient cycle changes, enhanced invasion of other species and extinctions (Simberloff et al. 2013). For example intraguild predation, the predation of species using similar resources, can create completely new stable states of an ecological system (Polis et al. 1989). Fully understanding the dynamics at play during the invasion process would open more possibilities to actively influence the invasion of threatening species. For this, creating models which are able to predict the invasion is one current focus of research. Since invasion theory already uses niche theory, it is quite appropriate to think about applying niche models to the problem.

Species Distribution Models (SDMs), are being applied to predict the further development of species occurrences in many contexts, also for invasions. These types of models have been shown to generate substantial insight into the ecological requirements of species and, as niche models, can be used to predict the potential habitat of a species (Araújo et al. 2006). There has been considerable debate on the capabilities and limitations of SDMs, especially when used for prediction outside the data domain. In general, SDMs are made with the (ideal) assumption that the species is in environmental equilibrium (Elith et al. 2009), implying that its ecological niche is not currently changing. If these models are now used to predict new, unsampled areas, there actually is no measure to assess their accuracy, since no data is presently available for that area (Araújo et al. 2006). This means that when trying to predict areas which are potentially outside the calibration range, sufficient validation data is lacking, implying strong uncertainty about the predictive performance of a given

model (Araújo et al. 2006). This issue of model transferability is an ongoing area of research in the SDM community. There is also no guarantee that the biotic interactions sampled in the study area will reflect the final interactions in the new area (Elith et al. 2009). In many cases it has been shown that invasive species suffered less from parasites or other natural enemies in the invaded range, for example making it easier for them to become pests with negative impact (Enemy Release Hypothesis, (Williamson 1996)). All of these issues apply especially to the prediction of invasive species, since there might be limited data in the invaded range, the species is often not currently at equilibrium and interactions with native species are completely new (Mainali et al. 2015). Despite all these challenges, SDMs have been used numerous times to provide insight into the invasive potential and the invasion dynamics of alien species (Zimmermann et al. 2010). One way of gaining more insight into the invasion process is to create models with data from different time periods during the invasion (Briscoe Runquist et al. 2019). For example, data from a time period early in the invasion process can be used to build models which then are evaluated against data from a later time period (Barbet-Massin et al. 2018). With this, SDMs can be used to detect niche shifts, which in turn improves the understanding of the underlying niche dynamics and their impact, which helps to put model performance into perspective, for example when using its results for risk assessment of potential invasions (Pearman et al. 2008).

SDM performance is not only influenced by the underlying data, but also the type of model chosen for the analysis. Models range from regression methods to machine learning and each feature various strengths and weaknesses, possibly leading to vastly different results for the same dataset (Valavi et al. 2022). Due to those differences, a possible approach is to create an ensemble of multiple models (Araújo et al. 2007). The way of combining model predictions can vary, but the goal is to improve total performance by combining the results of all computed models.

In order to conduct an iterative modelling approach, a species with sufficient data over the time span of invasion is necessary. *Harmonia axyridis*, also known as the Harlequin ladybird or multicoloured Asian lady beetle, is of the family of the Coccinellidae and has its native origin in Asia (Roy et al. 2016). At the time of download, the GBIF dataset for *H. axyridis* consisted of 468.462 data points globally, resulting in very sufficient amounts of data (see 3.1). At first widely introduced as a control species against pest aphids, *H. axyridis* has turned out to be a highly invasive species reaching an almost global distribution (P. Brown et al. 2008). In America, the species was introduced as early as 1916 (California) and in 1988, first populations outside intended release were found (Chapin et al. 1991). Usage of *H. axyridis* for biological control in Europe dates back as far as 1990 (France) (Coutanceau 2006). First invasive occurrences were confirmed in multiple

countries during the early 2000s, including Germany (2000), Belgium (2001), the Netherlands (2002) and the United Kingdom (2003) (Roy et al. 2016). The first confirmation in Austria, where it was never used for biological control, was in 2006 (Rabitsch et al. 2006). It has been shown that all established invasive populations outside of North America have their origin in the first established population in eastern North America, with the European populations being significantly influenced by the used biocontrol strain (Lombaert et al. 2010).

The impact of *H. axyridis* on invaded areas is diverse. In some contexts, the ladybird has been shown to have a negative impact on the diversity and abundance of native ladybird species (Roy et al. 2016). Many studies show intra guild predation and direct interspecific competition in favour of *H. axyridis* (Pell et al. 2008). This results in a large potential for *H. axyridis* to be a significant threat for guild diversity and community structure in its introduced ranges. It has also been shown that the species feeds on a variety of damaged fruit crops, for example grapes, apples, stone fruit and berry crops, making it a pest in these scenarios (Koch et al. 2004). The aggregating behaviour of *H. axyridis*, mostly as a strategy for overwintering, is also a cause of disturbance, since private homes and facilities are invaded by large amounts of beetles at a given time (Nalepa et al. 2005). In general, *H. axyridis* can be concluded to be a species with high impact as an invader, and thus of interest for active research questions.

There have been several publications which modelled and predicted the distribution of *H. axyridis*, constrained to certain geographical ranges (i.e. Spain (Ameixa et al. 2019), Chile (Alaniz et al. 2018)) or even on global scales (Bidinger et al. 2012; Poutsma et al. 2008). There has not yet been any model iteration in form of a time series, which is what this thesis aims to add as new insight. Another goal of this thesis is to look into the limitations of models built early in the invasion process of a species. By iterating over the years of the invasion, model performance can be evaluated with consideration to the current state of invasion. In the end, a better understanding of the invasion process of *H. axyridis* in Europe and the performance of models trying to capture it should be the result.

2 Materials and Methods

2.1 Datasets

For occurrence data, all global occurrences of *H. axyridis* were downloaded from the GBIF database (Gbif.org 2023). All traditional 19 bioclim variables, which mostly concern mean temperature, precipitation and different kinds of their variation (see Supplementary Table 3 for full breakdown), were obtained from the CHELSA V2.1 climatologies dataset (Karger et al. 2017; Karger et al. 2021). The 1981-2010 time frame was used for all years from 2002 to 2010, the MPI-ESM 1.2 ssp370 scenario 2011-2040 for all years from 2011 to 2022. As additional information, land cover data was used from the Copernicus Land cover Classification dataset (Copernicus Climate Change Service 2019) with yearly resolution for 2002 up to 2020.

2.2 Data preparation

All bioclim and land cover layers were resampled to a matching resolution of 30 arc seconds and cropped to two spatial extents, Europe and the presumed native range referencing (Orlova-Bienkowskaja, Ukrainsky & Brown, 2015) (Orlova-Bienkowskaja et al. 2015).

The presence-only points from GBIF were checked for missing values for latitude, longitude, year or coordinate uncertainty and then subset to the afore mentioned spatial extents. No occurrences after 2022 were used, also no points with a coordinate uncertainty larger than 1 km. In Europe, the initial cut off year for presences was 1991, since this is the year of invasion according to the EASIN website. Afterwards, using the library CoordinateCleaner (Zizka et al. 2019), all remaining data points were again checked for common errors or biases in the respective subset. The tests used check for occurrences too close to country capitals("capitals"), country centroids("centroids"), duplicates or points with equal absolute coordinates in both dimensions("duplicates", "equal"), as well as points close to biological institutions("institutions"), geographical outliers("outliers"), points located in oceans("seas") and points with equal longitude and latitude or close/equal to the point (0,0)("zeros"). In addition, all occurrences were checked for their land cover class values in their respective year, removing points in the water or with no data. In the end, remaining data points prior to 2002 were deemed insignificant and removed from the dataset. This resulted in a total of 124.746 presence points over all years and areas. To prepare the data for modelling, pseudo-absences were generated for each year, randomly sampling the area, discarding and re-sampling points in the water or with no data.

To correct for sampling bias in the data, the European and native extents were split into sub-extents in order to add additional absences to denser sampled regions, attempting to re-create the sampling bias for the pseudo-absences (Phillips et al. 2009). For this, an algorithm was written which splits a given extent in half and continues to do so with the created subextents until the amount of points in the extents is at most some chosen number. For the first part of absence generation completely random absences are drawn from the original extent in order to ensure at least some coverage of the whole study area. In the second step, additional absences are generated for each subextent separately and in relation to the amount of presences inside the respective subextent. This results in more absences in regions with more presences as well (Fig. 1 A). The subdivision of the dataset was carried out using all presences in Europe over all years and setting the threshold to be $\leq 30\%$, leading to subextents converging around the United Kingdom and the Netherlands, which seem to have been sampled very intensely (Fig. 1 B). This bias probably originates from the nature of the invasion, with *H. axyridis* seemingly having multiple fronts and degrees of severity during establishment. For the first and second step of absence generation, absences equal to and twice the amount of presences were generated respectively, resulting in three absences per presence in total.

2.3 Model building

In order to evaluate the impact of the invasion process on SDM predictive performance for *H. axyridis*, multiple models were compared: General Linear (GLM) (Guisan et al. 2002), General Additive (GAM) (Guisan et al. 2002), Boosted Regression Trees (BRT) (Elith et al. 2008) and Maximum Entropy (MAXENT) (Phillips et al. 2017). A model for a specific year always included all points from past years as well. The iterative models that were built only use data points from Europe, though there was one model created only with native occurrences and predicted for each year in Europe. This is to compare performance between models with and without potential influences of a niche shift through the process of invasion. For all used occurrence points after 2020, the land cover data of 2020 was used, since at the time of this thesis data for 2021 and 2022 was not yet available. The four built models were then used to create a TPR (sensitivity) weighted ensemble (see 2.4) for each year and the native models respectively.

Variance inflation factors were used to select the variables used for model building. For this, a GAM was computed only using Europe data from 2002, using all bioclim and land cover variables. For land cover variables, a PCA was computed on the relative area of all land cover classes in an 18 km radius around 5000 random data

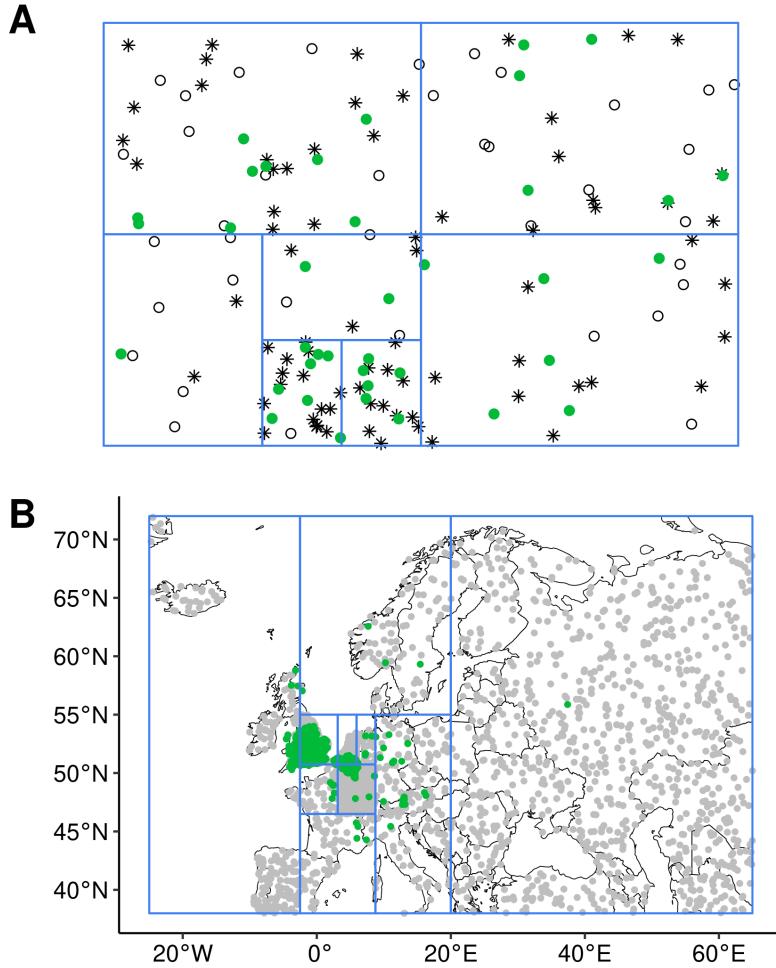


Figure 1: Visualization of the subdivision algorithm and absence generation. Subfigure A shows an example of 30 generated presences (green), subdivided with a threshold of ≤ 10 points per subextent. The generated absences are shown in black, with circles indicating the first 30 completely random absences, and asterisks indicating absences generated relative to the amount of presences in a subextent. Subfigure B shows the calculated subextents for the total presences of the dataset, with presences (green) and absences (grey) for 2008 plotted as an example.

points in Europe, subsequently projecting the occurrence data onto the resulting axes. The 18 km radius was chosen, since it is the average flight distance determined for *H. axyridis* (Jeffries et al. 2013). PCA axes were included in the model until a cut-off of 80% of explained variance was reached. Variance inflation factors were computed for this GAM and the variable with the highest VIF was dropped until none of the remaining variables had a VIF greater than 10 (rule of ten, see (O'brien 2007)).

2.4 Analysis

All SDM models of each year were evaluated for their accuracy on predicting the occurrences of the following year using the Sensitivity or True Positive Rate (TPR). The TPR values were used to create a TPR-weighted ensemble of all model predictions, which was again evaluated for its accuracy. The thresholds used to compute TPR were computed using the library `PresenceAbsence` (Freeman et al. 2008) while maximizing $(\text{sens} + \text{spec}) / 2$, which is equivalent to TSS (Allouche et al. 2006). This thresholding was chosen since just maximizing TPR can be achieved trivially by setting the threshold as low as possible, creating a model which predicts everything as suitable (used thresholds in Supplementary Tab. 6). For each year, the occupied niche was computed by running a PCA analysis on the bioclim variables. The niche was then visualized by plotting a dynamic occurrence density grid for the first two PCA axes (Broennimann et al. 2012). The overlap between each year for Europe and the respective following year was computed, as well as the niche dynamic indices "stability", "expansion" and "unfilling" (Guisan et al. 2014). Niche overlap for the total EU data was also visualized in comparison to the total native niche. All mentioned niche analyses were conducted using the library `ecospat` (Di Cola et al. 2017). The development of TPR over time was tested for correlation with the amount of training data and the niche stability for a given year, using the Pearson correlation test.

3 Results

The analyses mentioned above resulted in findings which shed light on the niche dynamics and predictability of *H. axyridis*.

3.1 Temporal change of data availability

Taking the amount of presence points in Europe and the native range, one can plot the amount of presences available for each year respectively (Fig. 2).

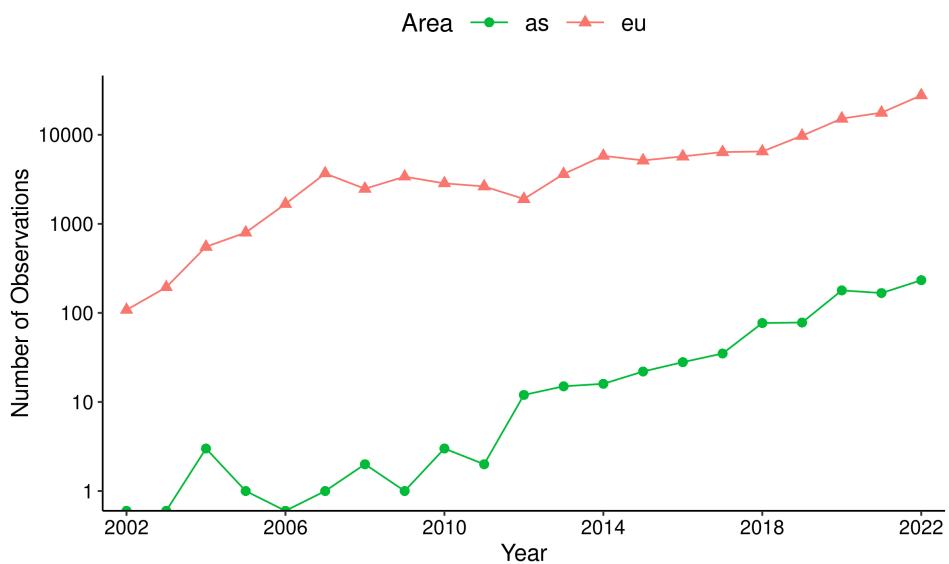


Figure 2: Amount of presence points for *H. axyridis* by year and area, using the cleaned dataset (Supplementary Fig. 8).

The resulting figure shows that the amount of data available in the invaded range greatly surpasses the amount in the native range. With at least 100 presences for any year, the European dataset is definitely sufficient to create SDMs for each year separately, more so if data from previous years is also used. The exponential increase in observations (in later years) is most likely due to an increase in sampling effort, since native observations grow at the same rate. The lack of presence points in early years in the native range was the reason why it was decided to only create one SDM with the native data of all years combined, since it is more likely to provide a complete evaluation of the native niche.

3.2 Niche dynamic analysis

When comparing the total native and invaded niches (so all years included), the niches are clearly different (Fig. 3). A niche similarity test produced a p value of $p = 0.32$, leading to an accepted null hypothesis meaning the two niches are not

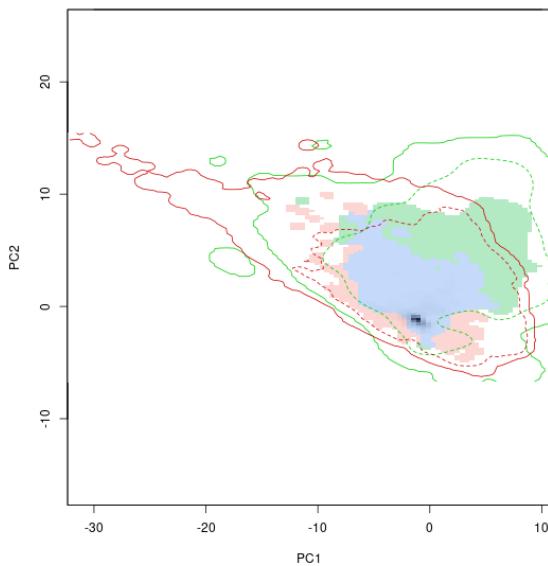


Figure 3: Native (green) and invaded (red) niche of *H. axyridis*, shown along the first two axes of a PCA using all bioclim variables (Supplementary Fig. 10). The blue area indicates overlap between the occupied niches. Dashed and solid lines indicate 50% and 100% of the potentially available environment in each area (from background). Grey shading shows the density distribution of the invaded niche.

more similar than random (Supplementary Fig. 9). Conducting a niche equivalency test, the result was $p = 0.01$ implying highly significant differences between the two niches (Supplementary Fig. 9).

Looking at the niche only in the invaded range, one can visualize the shift and expansion throughout the years (Fig. 4).

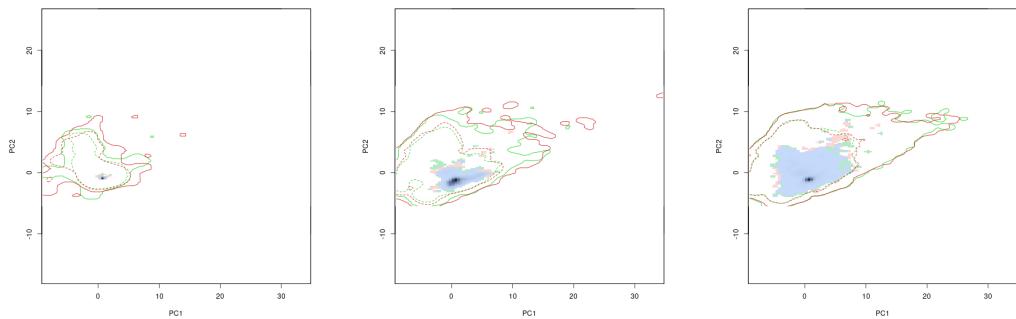


Figure 4: The progression of the invaded niche of *H. axyridis*, shown along the first two axes of a PCA using all bioclim variables (Supplementary Fig. 11). Niches of the years (from left to right) 2002, 2012 and 2021 compared to the niche of their following years respectively. Green shows the first year, red shows the year after and blue indicates overlap. Dashed and solid lines indicate 50% and 100% of the potentially available environment (from background). Grey shading shows the density distribution of the second year.

In addition to a visualization, it makes sense to compute the niche dynamic indices (Guisan et al. 2014), as well as the niche overlap (Schoener’s D) for each year in order to characterize the niche shift further (Fig. 5).

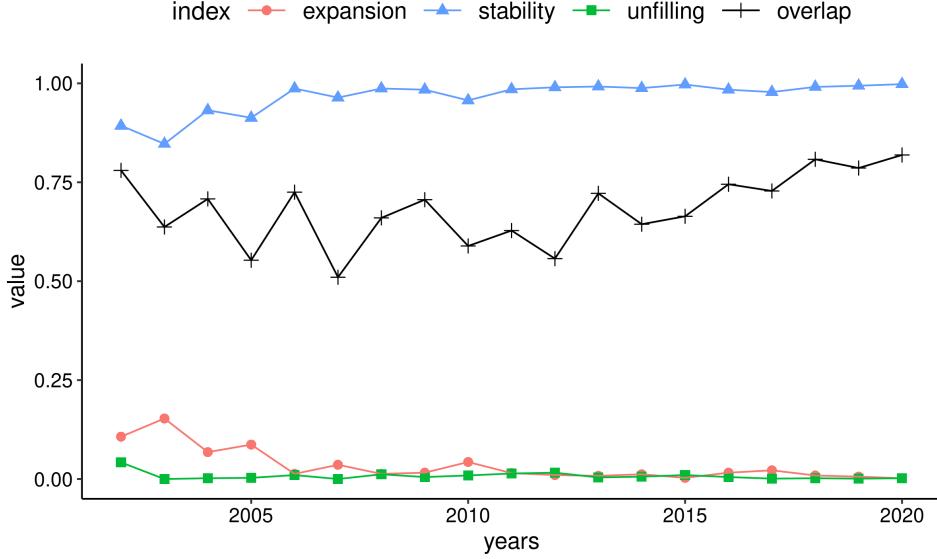


Figure 5: Niche dynamic indices and overlap (Schoener’s D) computed from comparing each year of the invaded niche of *H. axyridis* to the following year.

The results show that in the first five years of the time series, the invaded niche is still expanding, while afterwards it becomes almost completely stable. Once stability is reached, *H. axyridis* seems to just fill out the rest of the acquired niche, as shown by an increase in niche overlap.

3.3 SDMs, ensemble and time series

Variable selection using VIFs resulted in 13 variables, which were used for modelling (Tab. 1).

Table 1: Final 13 variables resulting from variable selection with VIFs, PCA contribution table for all land cover variables in (Supplementary Tab. 4).

Name	lc1	lc2	lc4	lc6	bio10	bio13	bio14	bio15	bio18	bio3	bio6	bio8	bio9
VIF	8.18	5.2	4.93	4.56	4.81	7.96	6.3	7.58	3.34	4.64	5.94	2.63	1.92

After computing all of the mentioned models (response curves in appendix E) and creating a TPR weighted ensemble, one can examine the development over time in more detail (Fig. 6).

The difference in performance between the invaded and native range models is apparent, with the invaded range models achieving a higher TPR on average. Surprisingly, the invasive data models already perform very well in the first years and

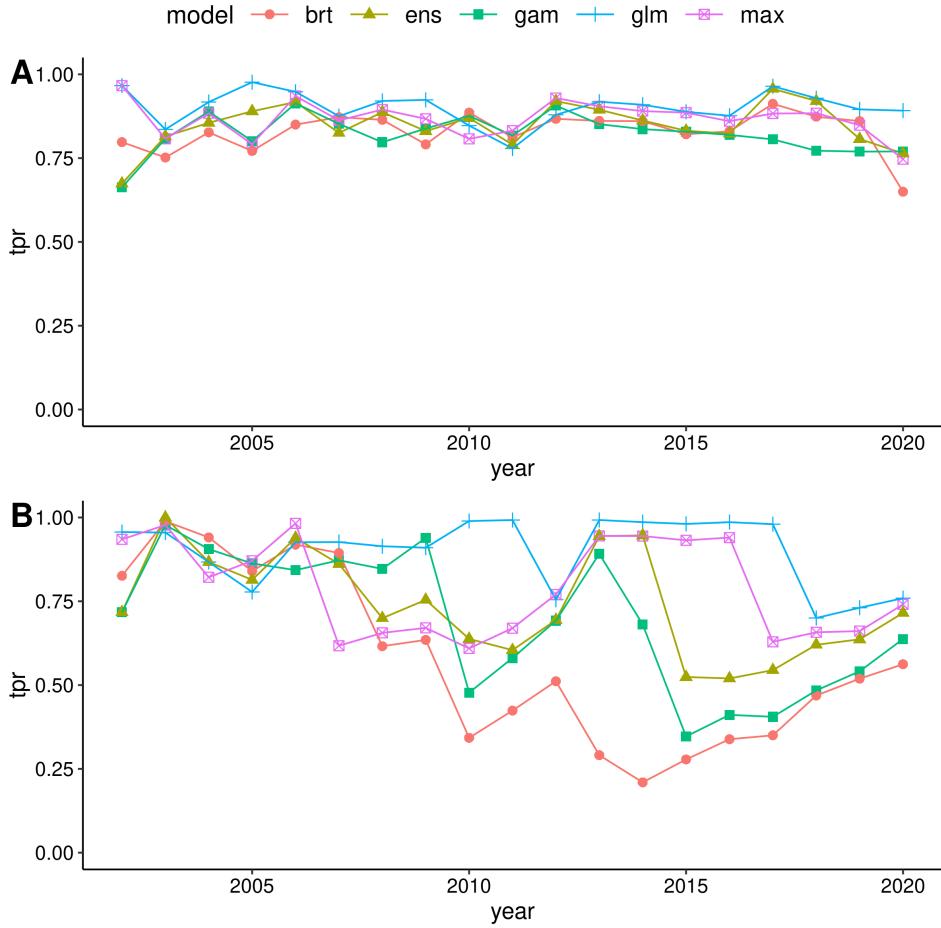


Figure 6: SDM predictive performance over time. Subfigure A shows the performance of models created with invasive data up to the year in question in predicting the following year. Subfigure B shows the performance of models created with all the native data in predicting the year in question.

continue to do so for all further years. The native models vary more in their predictive performance, starting off with a TPR similar to that of the invaded range models, then becoming quite chaotic in performance in the years between 2008 and 2015. After 2015, performance improves again, reaching levels close, but still below the initial performance and that of the final invaded range models. It is notable, that the GLM almost always performs the best out of all native models. The model ensemble performed at least better than 50 % of models used to build the ensemble for almost all years.

Trying to correlate the performance of each model individually to the trend in data amount or the niche stability value lead to statistically insignificant results for all models (Tab. 2).

Table 2: Results when correlating model TPR to either niche stability or amount of data.

Model	glm	gam	brt	max	ens
Corr. niche stability	-0.036	0.278	0.072	0.122	0.253
p-value	0.885	0.249	0.771	0.619	0.296
Corr. data count	-0.044	-0.27	-0.296	-0.32	-0.083
p-value	0.858	0.263	0.219	0.181	0.735

In the end, one can look at the model predictions for the year 2022 and compare them to the observed presences of that year (Fig. 7).

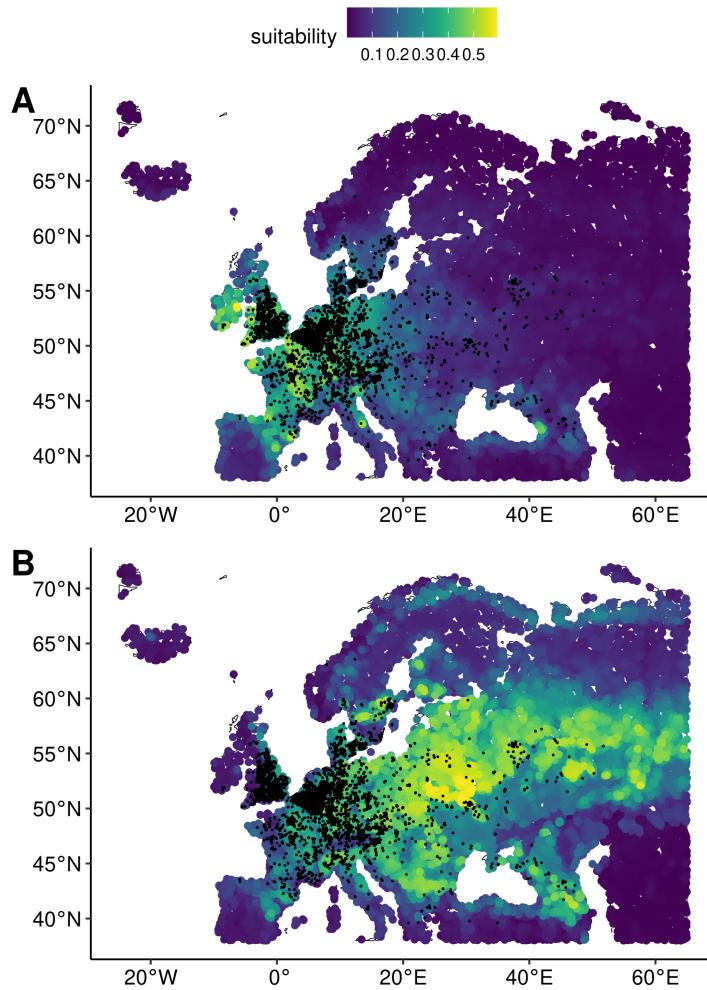


Figure 7: Predictions of the invaded (A) and native data (B) ensemble models for the suitability of *H. axyridis* in 2022. Suitability shown as colour, black dots indicating presences observed in 2022.

The ensemble from the invaded range proves to be more accurate in predicting the presences of 2022, although being rather conservative in its predictions. The native ensemble predicts a much wider range in comparison, reaching much further to the east.

4 Discussion

Using a time series of a well monitored species, it was possible to probe into the change in SDM performance over time. It was shown that all used SDMs consistently performed very well in predicting *H. axyridis* presences of the following year, when trained only on data from the invaded range in Europe.

The better performance of the invaded range models in comparison to the native models is supported by the results of niche analysis, showing that the native and invaded niche differ significantly from each other. It has already been shown that the genetic strains of *H. axyridis* present in Europe have been strongly influenced by American control strains (Lombaert et al. 2010), so a model using American data might also prove insightful.

The consistent performance of all invaded range models might be in part due to the rather consistent niche stability, though also in years with significant niche expansion (early years) the models are able to perform well (also see Supplementary Fig. 12 for a yearly visualization of model spatial predictions). This could also imply that SDMs are able to perform well in environments without total niche conservatism, but the data shown in this thesis is not sufficient to make this claim. The lack of impact of data amount on model performance is unsurprising, since all models had a very sufficient amount of data points (> 15 , (Støa et al. 2019)). The native, as well as the invaded range ensemble predict ranges very similar to a study from 2012 (Bidinger et al. 2012), which speaks for the modelling process of this thesis. Agreement between predictions from 2012 and models created with more recent data also supports the result of niche conservatism over this time period.

The use of TPR as a measure for model performance is not ideal, but was chosen instead of TSS due to better interpretability in the context of an invasive species. For future work, it would be better to look into different evaluation metrics, as there have been numerous proposed in other works. It might also be interesting to look into the period of niche expansion of *H. axyridis* in more detail, for example with a monthly resolution, since data might be sufficient.

In the specific case of *H. axyridis*, SDMs can definitely be used to model the current potential niche in Europe, aided by the very stable invaded niche over the past few years and the very good performance of SDMs as shown in this case study. It would surely be more useful to also take dispersal into account in future models, since with a now seemingly stable niche, the dominating limitations of further spread will be strongly related to its dispersal ability. While currently, *H. axyridis* is not listed as a "Species of Union Concern" by the EU (Commission et al. 2020), new research still keeps showing negative impacts on native environments, especially on native ladybird species (P. M. Brown et al. 2022). This opens possibilities for SDMs to

support future endeavours to quantify the impact of *H. axyridis* on native species, by enhancing sampling strategies and potential spatial correlations of impact severity.

5 Conclusion

The results of this thesis show that SDMs can be a strong tool in predicting the potential niche of an invasive species, also with not completely static niches. Ensembles proved to be a good way to combine different model predictions, resulting in predictions consistent with previous studies. Using additional niche analyses provides valuable context to interpret model performance and gives additional insight into the establishment process of *H. axyridis*. It has been shown that the process used in this thesis can generate new insight into the dynamics and relations at play when dealing with invasive species and predicting their potential impact, with more interesting areas to research further. Also in the specific case of *H. axyridis*, this thesis shows possible future potential for SDM application in research concerning the future impact of this invasive species.

6 Acknowledgements

This thesis would not have been possible in this form without the huge amount of freedom given by my main supervisor Lauren Talluto. I was able to just try out my ideas on my own accord, which is something I value highly. My twin brother Emanuel gave me a lot of advice on good coding practice, which helped me a lot. I also want to thank all my friends and other relatives, who just listened to me rambling about how cool (or sometimes nerve-wracking) my progress was, even if they only understood half of it.

References

- Alaniz, AJ, AA Grez and T Zaviezo (2018). ‘Potential spatial interaction of the invasive species *Harmonia axyridis* (Pallas) with native and endemic coccinellids’. In: *Journal of Applied Entomology* 142.5, pp. 513–524. DOI: 10.1111/jen.12498.
- Allouche, Omri, Asaf Tsoar and Ronen Kadmon (2006). ‘Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS)’. In: *Journal of applied ecology* 43.6, pp. 1223–1232. DOI: 10.1111/j.1365-2664.2006.01214.x.
- Ameixa, Olga MCC et al. (2019). ‘Factors influencing the introduction and spread of *Harmonia axyridis* in the Iberian Peninsula’. In: *Biological Invasions* 21, pp. 323–331. DOI: 10.1007/s10530-018-1841-x.
- Araújo, Miguel B and Antoine Guisan (2006). ‘Five (or so) challenges for species distribution modelling’. In: *Journal of biogeography* 33.10, pp. 1677–1688. DOI: 10.1111/j.1365-2699.2006.01584.x.
- Araújo, Miguel B and Mark New (2007). ‘Ensemble forecasting of species distributions’. In: *Trends in ecology & evolution* 22.1, pp. 42–47. DOI: 10.1016/j.tree.2006.09.010.
- Barbet-Massin, Morgane et al. (2018). ‘Can species distribution models really predict the expansion of invasive species?’ In: *PLoS one* 13.3, e0193085. DOI: 10.1371/journal.pone.0193085.
- Bidinger, K et al. (2012). ‘Species distribution models for the alien invasive Asian Harlequin ladybird (*Harmonia axyridis*)’. In: *Journal of Applied Entomology* 136.1-2, pp. 109–123. DOI: 10.1111/j.1439-0418.2010.01598.x.
- Blackburn, Tim M et al. (2011). ‘A proposed unified framework for biological invasions’. In: *Trends in ecology & evolution* 26.7, pp. 333–339. DOI: 10.1016/j.tree.2011.03.023.
- Briscoe Runquist, Ryan D et al. (2019). ‘Species distribution models throughout the invasion history of Palmer amaranth predict regions at risk of future invasion and reveal challenges with modeling rapidly shifting geographic ranges’. In: *Scientific Reports* 9.1, p. 2426. DOI: 10.1038/s41598-018-38054-9.
- Broennimann, Olivier et al. (2012). ‘Measuring ecological niche overlap from occurrence and spatial environmental data’. In: *Global ecology and biogeography* 21.4, pp. 481–497. DOI: 10.1111/j.1466-8238.2011.00698.x.
- Brown, Peter MJ et al. (2022). ‘Invasive intraguild predators: Evidence of their effects, not assumptions’. In: *Ecological Entomology* 3, pp. 249–252.
- Brown, PMJ et al. (2008). ‘*Harmonia axyridis* in Europe: spread and distribution of a non-native coccinellid’. In: *BioControl* 53, pp. 5–21. DOI: 10.1007/s10526-007-9132-y.
- Chapin, Joan B, VA Brou et al. (1991). ‘*Harmonia axyridis* (Pallas), the third species of the genus to be found in the United States (Coleoptera: Coccinellidae)’. In: *Proc. Entomol.*

- Soc. Wash* 93.3, pp. 630–635. URL: <https://www.biodiversitylibrary.org/partpdf/55539>.
- Commission, European and Directorate-General for Environment (2020). *Invasive alien species of Union concern – Version 2020*. Publications Office. DOI: doi/10.2779/791940.
- Copernicus Climate Change Service (2019). *Land cover classification gridded maps from 1992 to present derived from satellite observation*. (Last accessed 3 August 2023). DOI: 10.24381/cds.006f2c9a.
- Coutanceau, Jean (Jan. 2006). ‘Harmonia axyridis (Pallas, 1773): Une coccinelle asiatique introduite, acclimatée et en extension en France’. In: *Bulletin de la Société Entomologique de France* 111, pp. 395–401. DOI: 10.3406/bsef.2006.16343.
- Di Cola, Valeria et al. (2017). ‘ecospat: an R package to support spatial analyses and modeling of species niches and distributions’. In: *Ecography* 40.6, pp. 774–787. DOI: 10.1111/ecog.02671.
- Elith, Jane and John R Leathwick (2009). ‘Species distribution models: ecological explanation and prediction across space and time’. In: *Annual review of ecology, evolution, and systematics* 40, pp. 677–697. DOI: 10.1146/annurev.ecolsys.110308.120159.
- Elith, Jane, John R Leathwick and Trevor Hastie (2008). ‘A working guide to boosted regression trees’. In: *Journal of animal ecology* 77.4, pp. 802–813. DOI: 10.1111/j.1365-2656.2008.01390.x.
- Freeman, Elizabeth A. and Gretchen Moisen (2008). ‘PresenceAbsence: An R Package for Presence Absence Analysis’. In: *Journal of Statistical Software* 23.11, pp. 1–31. DOI: 10.18637/jss.v023.i11. URL: <https://www.jstatsoft.org/article/view/v023i11>.
- Gbif.org (15th July 2023). *GBIF Occurrence Download*. DOI: 10.15468/dl.y4y3s9.
- Guisan, Antoine, Thomas C Edwards Jr and Trevor Hastie (2002). ‘Generalized linear and generalized additive models in studies of species distributions: setting the scene’. In: *Ecological modelling* 157.2-3, pp. 89–100. DOI: 10.1016/S0304-3800(02)00204-1.
- Guisan, Antoine et al. (2014). ‘Unifying niche shift studies: insights from biological invasions’. In: *Trends in ecology & evolution* 29.5, pp. 260–269. DOI: 10.1016/j.tree.2014.02.009.
- Jeffries, Daniel L et al. (2013). ‘Characteristics and drivers of high-altitude ladybird flight: insights from vertical-looking entomological radar’. In: *PloS one* 8.12, e82278. DOI: 10.1371/journal.pone.0082278.
- Karger, Dirk Nikolaus et al. (2017). ‘Climatologies at high resolution for the earth’s land surface areas’. In: *Scientific data* 4.1, pp. 1–20. DOI: 10.1038/sdata.2017.122.
- (2021). *Climatologies at high resolution for the earth’s land surface areas*. (Last accessed 3 August 2023). DOI: 10.16904/envidat.228.

- Koch, RL et al. (2004). ‘Phytophagous preferences of the multicolored Asian lady beetle (Coleoptera: Coccinellidae) for autumn-ripening fruit’. In: *Journal of Economic Entomology* 97.2, pp. 539–544. DOI: 10.1093/jee/97.2.539.
- Lombaert, Eric et al. (2010). ‘Bridgehead effect in the worldwide invasion of the biocontrol harlequin ladybird’. In: *PloS one* 5.3, e9743. DOI: 10.1371/journal.pone.0009743.
- Mainali, Kumar P et al. (2015). ‘Projecting future expansion of invasive species: comparing and improving methodologies for species distribution modeling’. In: *Global change biology* 21.12, pp. 4464–4480. DOI: 10.1111/gcb.13038.
- Nalepa, CA, GG Kennedy and C Brownie (2005). ‘Role of visual contrast in the alighting behavior of *Harmonia axyridis* (Coleoptera: Coccinellidae) at overwintering sites’. In: *Environmental Entomology* 34.2, pp. 425–431. DOI: 10.1603/0046-225X-34.2.425.
- O’brien, Robert M (2007). ‘A caution regarding rules of thumb for variance inflation factors’. In: *Quality & quantity* 41, pp. 673–690. DOI: 10.1007/s11135-006-9018-6.
- Orlova-Bienkowskaja, Marina J, Andrey S Ukrainsky and Peter MJ Brown (2015). ‘*Harmonia axyridis* (Coleoptera: Coccinellidae) in Asia: a re-examination of the native range and invasion to southeastern Kazakhstan and Kyrgyzstan’. In: *Biological Invasions* 17, pp. 1941–1948. DOI: 10.1007/s10530-015-0848-9.
- Pearman, Peter B et al. (2008). ‘Niche dynamics in space and time’. In: *Trends in ecology & evolution* 23.3, pp. 149–158. DOI: 10.1016/j.tree.2007.11.005.
- Pell, Judith K et al. (2008). ‘Intraguild predation involving *Harmonia axyridis*: a review of current knowledge and future perspectives’. In: *BioControl* 53, pp. 147–168. DOI: 10.1007/978-1-4020-6939-0_10.
- Phillips, Steven J et al. (2009). ‘Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data’. In: *Ecological applications* 19.1, pp. 181–197. DOI: 10.1890/07-2153.1.
- Phillips, Steven J et al. (2017). ‘Opening the black box: An open-source release of Maxent’. In: *Ecography* 40.7, pp. 887–893. DOI: 10.1111/ecog.03049.
- Polis, Gary A, Christopher A Myers and Robert D Holt (1989). ‘The ecology and evolution of intraguild predation: potential competitors that eat each other’. In: *Annual review of ecology and systematics* 20.1, pp. 297–330. DOI: 10.1146/annurev.es.20.110189.001501.
- Poutsma, J et al. (2008). ‘Predicting the potential geographical distribution of the harlequin ladybird, *Harmonia axyridis*, using the CLIMEX model’. In: *From Biological Control to Invasion: the ladybird *Harmonia axyridis* as a model species*, pp. 103–125. DOI: 10.1007/s10526-007-9140-y.
- Rabitsch, W and R Schuh (2006). ‘First record of the multicoloured Asian ladybird *Harmonia axyridis* (Pallas, 1773) in Austria’. In: *Beiträge zur Entomofaunistik* 7, pp. 161–164. URL: https://www.zobodat.at/pdf/BEF_7_0161-0164.pdf.

- Roy, Helen E et al. (2016). ‘The harlequin ladybird, *Harmonia axyridis*: global perspectives on invasion history and ecology’. In: *Biological invasions* 18, pp. 997–1044. DOI: 10.1007/s10530-016-1077-6.
- Shigesada, Nanako and Kohkichi Kawasaki (1997). *Biological invasions: theory and practice*. Oxford University Press, UK. URL: https://books.google.at/books/about/Biological_Invasions_Theory_and_Practice.html?id=Ri-hle_zdpsc&redir_esc=y.
- Simberloff, Daniel et al. (2013). ‘Impacts of biological invasions: what’s what and the way forward’. In: *Trends in ecology & evolution* 28.1, pp. 58–66. DOI: 10.1016/j.tree.2012.07.013.
- Støa, Bente et al. (2019). ‘How much is enough? Influence of number of presence observations on the performance of species distribution models’. In: *Sommerfeltia* 39.1, pp. 1–28. DOI: 10.2478/som-2019-0001.
- Valavi, Roozbeh et al. (2022). ‘Predictive performance of presence-only species distribution models: a benchmark study with reproducible code’. In: *Ecological Monographs* 92.1, e01486. DOI: 10.1002/ecm.1486.
- Williamson, Mark (1996). *Biological invasions*. Springer Science & Business Media. URL: <https://books.google.at/books?id=eWUDzI6j3V8C&printsec=frontcover#v=onepage&q&f=false>.
- Zimmermann, Niklaus E et al. (2010). ‘New trends in species distribution modelling’. In: *Ecography* 33.6, pp. 985–989. DOI: 10.1111/j.1600-0587.2010.06953.x.
- Zizka, Alexander et al. (2019). ‘CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases’. In: *Methods in Ecology and Evolution* 10.5, pp. 744–751. DOI: 10.1111/2041-210X.13152.

A Visualization of cleaned vs raw dataset

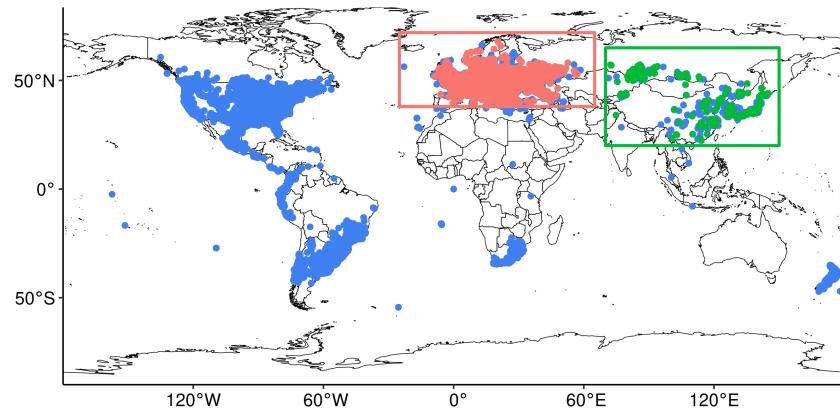


Figure 8: Visualization of the cleaned dataset for *H. axyridis* in comparison to the total raw dataset. The red and green boxes show the used extents for Europe and the native range respectively, red and green points show the cleaned presence points in their respective extents, while blue points show all points of the raw dataset.

B Niche analysis

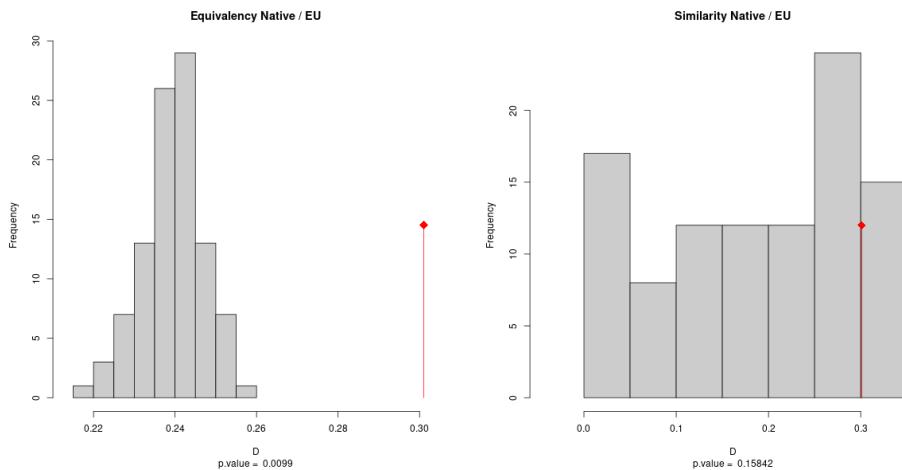


Figure 9: Results of the niche equivalency (left) and niche similarity (right) test comparing the native and invaded niche of *H. axyridis*. Histograms of the simulated niche overlaps, the observed overlap shown as a red bar with a diamond.

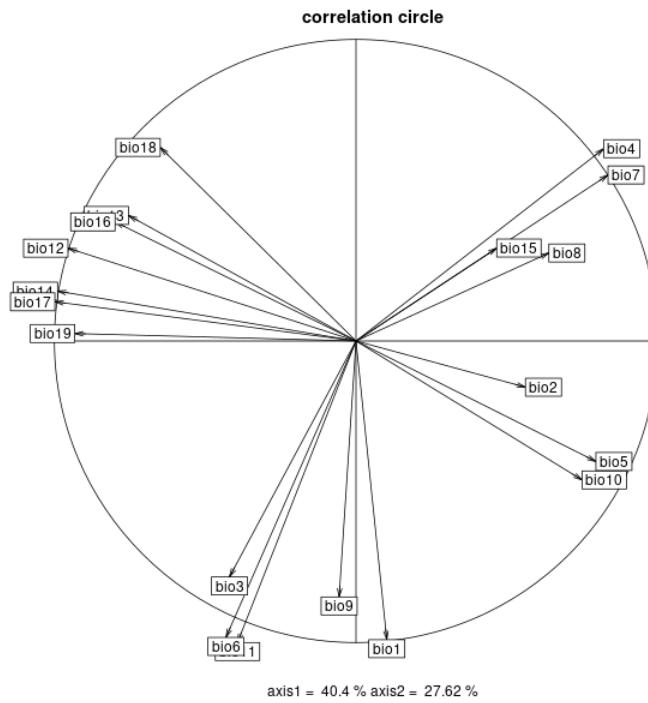


Figure 10: Component contributions for the PCA used to conduct the niche analyses comparing the total native and invaded niches. For detail on the bioclim variables, see (Supplementary Tab. 3).

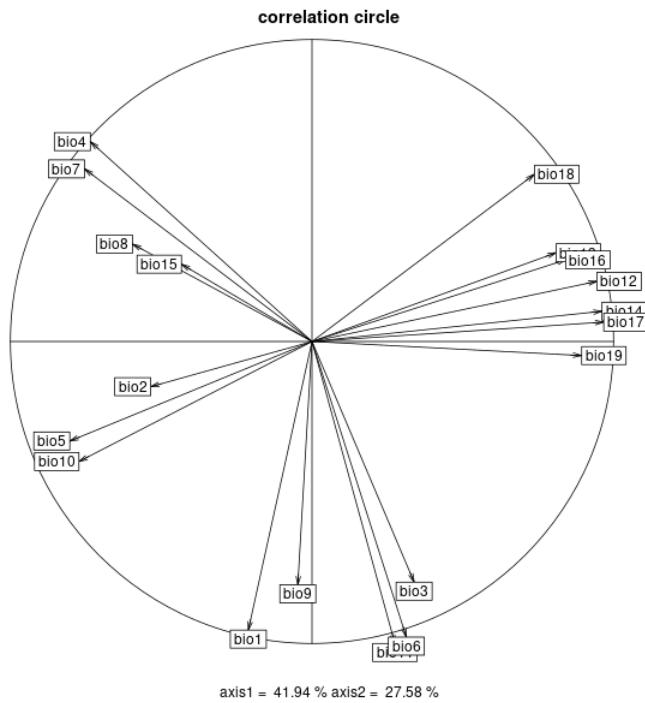


Figure 11: Component contributions for the PCA used to conduct the niche analyses comparing each year of the invaded niche to its following year. For detail on the bioclim variables, see (Supplementary Tab. 3).

Table 3: Explanation of the bioclim variables (from CHELSA 2.x technical specifications).

Variable	Explanation
bio1	mean annual air temperature
bio2	mean diurnal air temperature range
bio3	isothermality
bio4	temperature seasonality
bio5	mean daily maximum air temperature of the warmest month
bio6	mean daily minimum air temperature of the coldest month
bio7	annual range of air temperature
bio8	mean daily mean air temperatures of the wettest quarter
bio9	mean daily mean air temperatures of the driest quarter
bio10	mean daily mean air temperatures of the warmest quarter
bio11	mean daily mean air temperatures of the coldest quarter
bio12	annual precipitation amount
bio13	precipitation amount of the wettest month
bio14	precipitation amount of the driest month
bio15	precipitation seasonality
bio16	mean monthly precipitation amount of the wettest quarter
bio17	mean monthly precipitation amount of the driest quarter
bio18	mean monthly precipitation amount of the warmest quarter
bio19	mean monthly precipitation amount of the coldest quarter

C Species distribution modelling

Table 4: Table showing PCA contributions for each Copernicus land cover class, Copernicus land cover class explanation in (Supplementary Tab. 5).

Class	lc1	lc2	lc3	lc4	lc5	lc6
0	7.67×10^{-34}	2.90×10^{-34}	9.09×10^{-34}	6.84×10^{-35}	3.13×10^{-34}	1.35×10^{-36}
10	4.34×10^1	1.71×10^1	2.27×10^1	5.36×10^{-1}	7.00	4.36×10^{-1}
11	6.41	7.20×10^{-1}	5.86	8.10	6.54×10^1	4.62×10^{-1}
12	1.67×10^{-3}	6.40×10^{-6}	1.53×10^{-2}	6.85×10^{-5}	6.96×10^{-2}	2.52×10^{-2}
20	2.95×10^{-3}	1.32×10^{-1}	2.58×10^{-4}	2.28×10^{-2}	6.36×10^{-2}	2.97×10^{-3}
30	3.54×10^{-1}	4.06×10^{-2}	4.35×10^{-1}	5.83×10^{-1}	3.13×10^{-2}	3.80×10^{-1}
40	8.36×10^{-2}	5.08×10^{-3}	3.07×10^{-1}	9.58×10^{-5}	2.46×10^{-1}	1.05×10^{-1}
50	0.00	3.01×10^{-34}	4.81×10^{-33}	2.77×10^{-30}	7.70×10^{-32}	1.23×10^{-30}
60	1.84×10^{-1}	4.91×10^{-1}	6.49	6.12×10^1	3.45×10^{-3}	1.31
61	1.21×10^{-4}	6.93×10^{-5}	4.19×10^{-6}	1.43×10^{-3}	9.89×10^{-5}	5.24×10^{-5}
62	4.04×10^{-41}	1.84×10^{-38}	4.23×10^{-35}	0.00	0.00	1.23×10^{-30}
70	4.59×10^1	2.44×10^1	6.62	3.54	5.85×10^{-3}	2.85
71	5.64×10^{-11}	9.17×10^{-11}	1.12×10^{-11}	1.25×10^{-11}	4.80×10^{-11}	4.79×10^{-10}
72	0.00	2.74×10^{-46}	0.00	1.84×10^{-38}	4.59×10^{-39}	3.01×10^{-34}
80	1.41×10^{-5}	6.90×10^{-7}	9.66×10^{-10}	1.76×10^{-6}	2.09×10^{-7}	7.78×10^{-6}
81	3.98×10^{-57}	2.61×10^{-52}	4.28×10^{-48}	0.00	0.00	7.35×10^{-38}
82	0.00	0.00	4.18×10^{-51}	6.84×10^{-47}	6.84×10^{-47}	7.17×10^{-41}
90	1.05	6.90×10^{-1}	3.31×10^{-1}	8.27	2.01×10^{-2}	7.52×10^{-1}
100	8.89×10^{-2}	2.08×10^{-2}	9.54×10^{-2}	3.51×10^{-2}	2.48×10^{-2}	4.78×10^{-2}
110	1.82×10^{-2}	4.76×10^{-3}	1.07×10^{-2}	4.45×10^{-3}	1.82×10^{-2}	2.46×10^{-1}
120	1.10×10^{-2}	3.53×10^{-1}	6.47×10^{-2}	2.27×10^{-4}	7.52×10^{-3}	1.53×10^{-1}
121	1.32×10^{-80}	3.45×10^{-75}	0.00	2.37×10^{-64}	9.50×10^{-64}	1.02×10^{-54}
122	1.56×10^{-2}	1.31×10^{-2}	3.05×10^{-3}	3.08×10^{-3}	4.34×10^{-3}	6.71×10^{-1}
130	2.88×10^{-1}	3.07	2.67×10^1	1.60×10^1	2.21×10^1	1.94×10^1
140	1.12×10^{-4}	5.70×10^{-3}	6.71×10^{-4}	1.30×10^{-3}	3.27×10^{-3}	4.87×10^{-2}
150	1.31×10^{-1}	4.39	2.93×10^{-1}	1.30	3.45	5.20×10^1
151	3.00×10^{-93}	1.23×10^{-89}	1.81×10^{-84}	6.81×10^{-81}	1.08×10^{-80}	5.53×10^{-74}
152	2.53×10^{-4}	3.69×10^{-5}	1.65×10^{-5}	2.25×10^{-5}	5.64×10^{-6}	2.87×10^{-4}
153	2.57×10^{-5}	1.70×10^{-5}	1.49×10^{-5}	8.04×10^{-5}	2.01×10^{-5}	1.54×10^{-4}
160	5.97×10^{-5}	1.07×10^{-4}	1.04×10^{-4}	2.47×10^{-5}	4.30×10^{-5}	4.22×10^{-4}
170	0.00	0.00	0.00	0.00	0.00	0.00
180	1.28	2.45×10^{-1}	3.66×10^{-2}	1.77×10^{-2}	3.78×10^{-2}	2.67×10^{-1}
190	3.76×10^{-2}	7.01×10^{-3}	2.54×10^{-2}	2.14×10^{-2}	1.18×10^{-1}	7.56×10^{-3}
200	5.52×10^{-1}	4.83×10^1	3.00×10^1	1.53×10^{-1}	1.39	9.21
201	4.02×10^{-5}	9.95×10^{-6}	1.26×10^{-4}	6.16×10^{-8}	1.47×10^{-4}	6.12×10^{-5}
202	1.90×10^{-4}	1.61×10^{-2}	8.83×10^{-3}	8.67×10^{-5}	5.88×10^{-4}	1.87×10^{-3}
210	2.50×10^{-1}	1.19×10^{-2}	1.47×10^{-2}	2.43×10^{-1}	8.46×10^{-6}	1.16×10^1
220	1.87×10^{-4}	4.15×10^{-3}	2.22×10^{-7}	1.12×10^{-5}	5.05×10^{-4}	1.48×10^{-2}

Table 5: Explanation of the Copernicus land cover classes (from dataset legend).

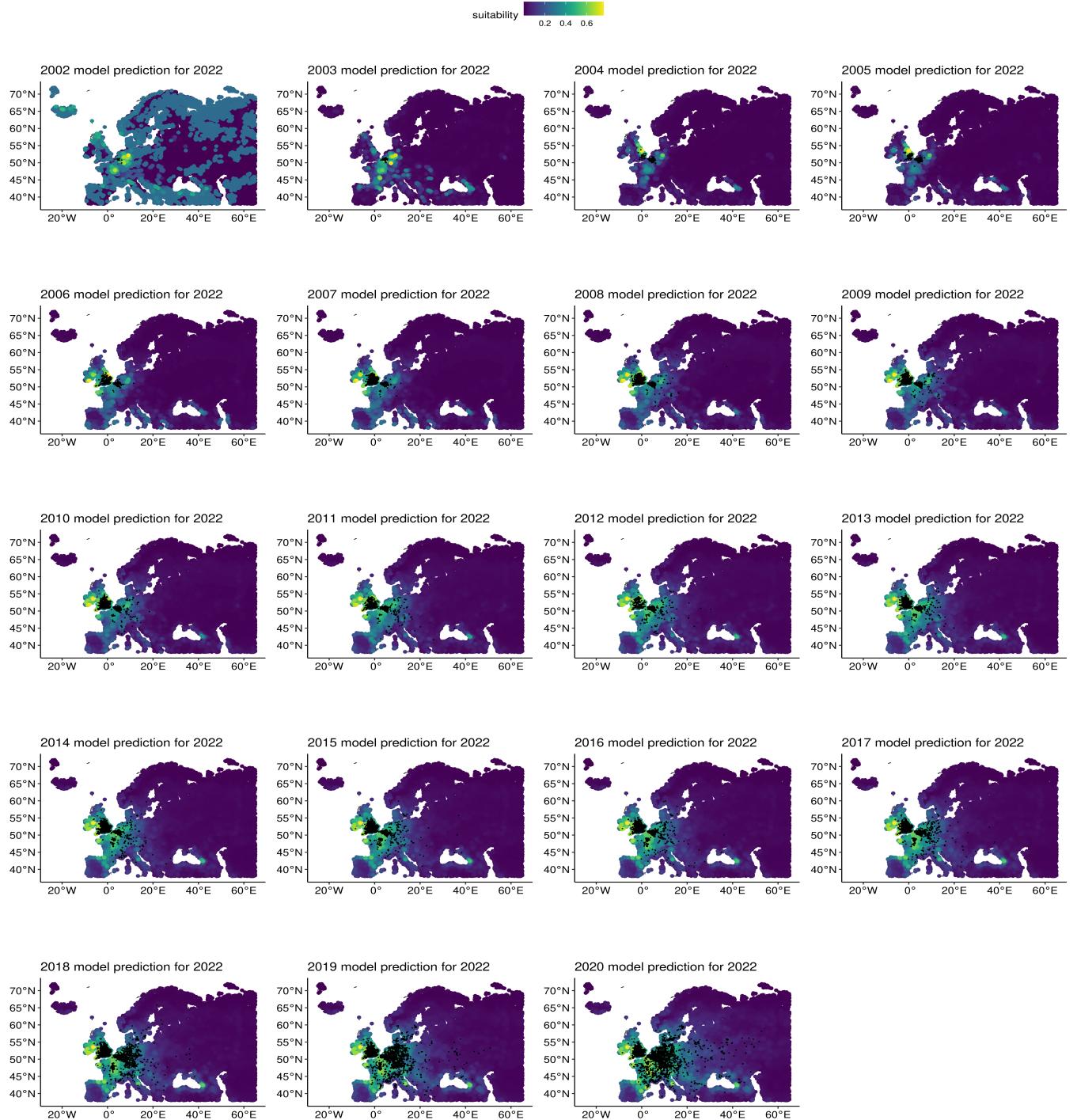
Class	Explanation
0	no data
10	cropland rainfed
11	cropland rainfed herbaceous cover
12	cropland rainfed tree or shrub cover
20	cropland irrigated
30	mosaic cropland
40	mosaic natural vegetation
50	tree broadleaved evergreen closed to open
60	tree broadleaved deciduous closed to open
61	tree broadleaved deciduous closed
62	tree broadleaved deciduous open
70	tree needleleaved evergreen closed to open
71	tree needleleaved evergreen closed
72	tree needleleaved evergreen open
80	tree needleleaved deciduous closed to open
81	tree needleleaved deciduous closed
82	tree needleleaved deciduous open
90	tree mixed
100	mosaic tree and shrub
110	mosaic herbaceous
120	shrubland
121	shrubland evergreen
122	shrubland deciduous
130	grassland
140	lichens and mosses
150	sparse vegetation
151	sparse tree
152	sparse shrub
153	sparse herbaceous
160	tree cover flooded fresh or brakish water
170	tree cover flooded saline water
180	shrub or herbaceous cover flooded
190	urban
200	bare areas
201	bare areas consolidated
202	bare areas unconsolidated
210	water
220	snow and ice

Table 6: Thresholds used to predict the models for the given years (inv = invaded range model (of the previous year), nat = native range model).

Year	glm inv	gam inv	brt inv	max inv	ens inv	glm nat	gam nat	brt nat	max nat	ens nat
2002	-	-	-	-	-	0.12	0.4	0.27	0.93	0.44
2003	0.03	0.5	0.04	0.15	0.28	0.11	0.13	0.15	0.9	0.32
2004	0.21	0.17	0.21	0.32	0.22	0.11	0.125	0.14	0.92	0.32
2005	0.13	0.135	0.11	0.27	0.14	0.1	0.08	0.14	0.81	0.31
2006	0.02	0.14	0.17	0.26	0.12	0.02	0.04	0.03	0.17	0.08
2007	0.22	0.15	0.18	0.27	0.22	0.02	0.03	0.03	0.83	0.09
2008	0.24	0.23	0.2	0.37	0.3	0.02	0.03	0.14	0.79	0.21
2009	0.22	0.26	0.21	0.35	0.27	0.02	0.01	0.14	0.81	0.18
2010	0.21	0.25	0.27	0.38	0.3	0.01	0.13	0.24	0.87	0.27
2011	0.21	0.14	0.13	0.34	0.22	0.01	0.11	0.29	0.53	0.28
2012	0.29	0.28	0.29	0.41	0.34	0.05	0.11	0.295	0.53	0.32
2013	0.24	0.27	0.28	0.38	0.3	0.01	0.01	0.29	0.16	0.06
2014	0.23	0.31	0.28	0.42	0.33	0.01	0.04	0.3	0.15	0.06
2015	0.23	0.3	0.28	0.41	0.33	0.01	0.18	0.29	0.15	0.32
2016	0.26	0.3	0.28	0.41	0.34	0.01	0.18	0.3	0.15	0.34
2017	0.26	0.3	0.28	0.42	0.34	0.01	0.19	0.3	0.53	0.29
2018	0.07	0.22	0.1	0.29	0.11	0.05	0.21	0.25	0.53	0.3
2019	0.1	0.24	0.13	0.29	0.15	0.05	0.19	0.25	0.53	0.29
2020	0.12	0.24	0.14	0.32	0.245	0.06	0.17	0.28	0.52	0.3
2021	0.16	0.27	0.31	0.44	0.31	0.06	0.17	0.28	0.53	0.3
2022	-	-	-	-	-	0.06	0.17	0.3	0.53	0.3

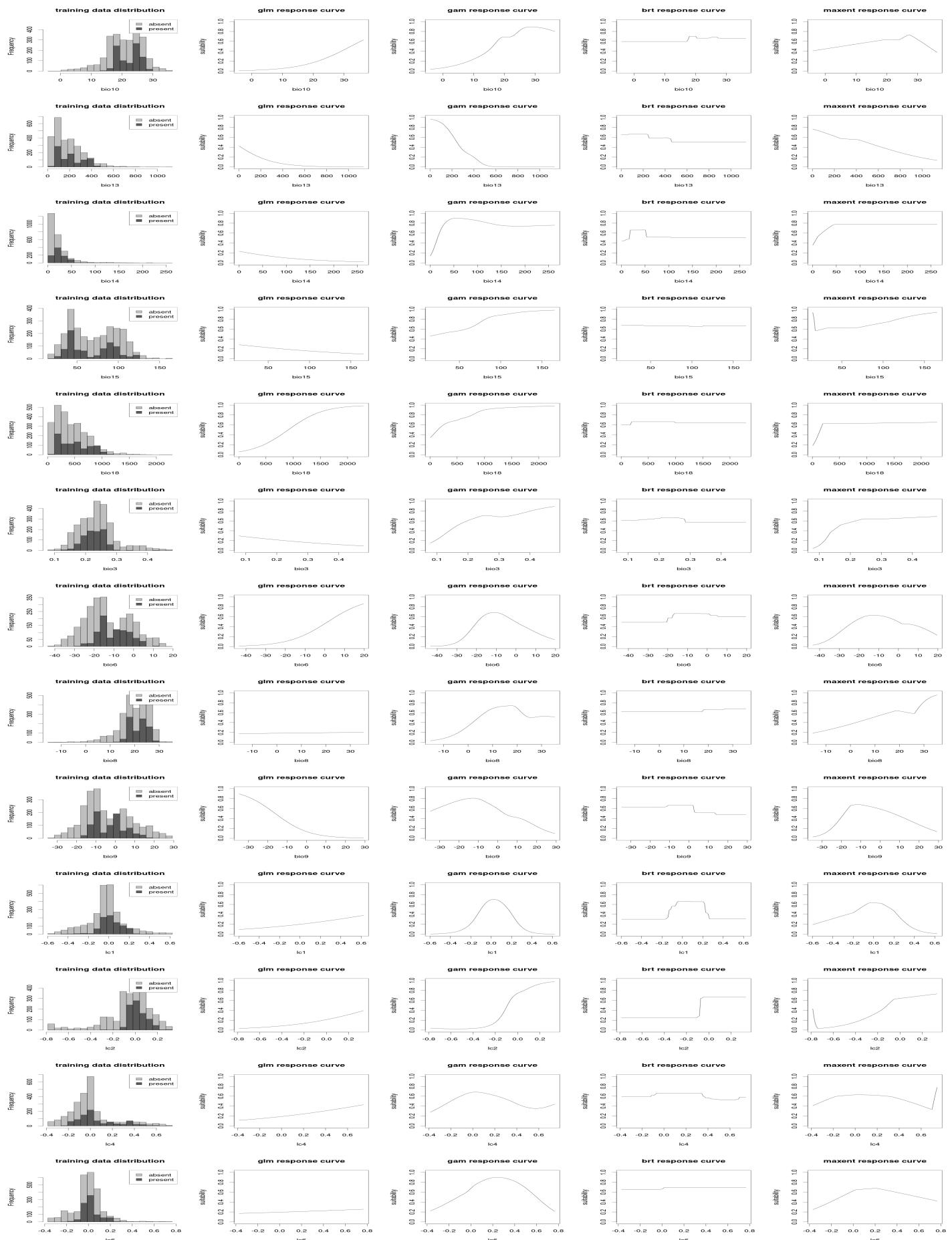
D Model predictions

Figure 12: Predictions of each invaded range ensemble model for the year 2022, points in black showing the occurrences of the respective year, so i.e. in 2003 model it shows the occurrences of 2003, but the suitability prediction for 2022.

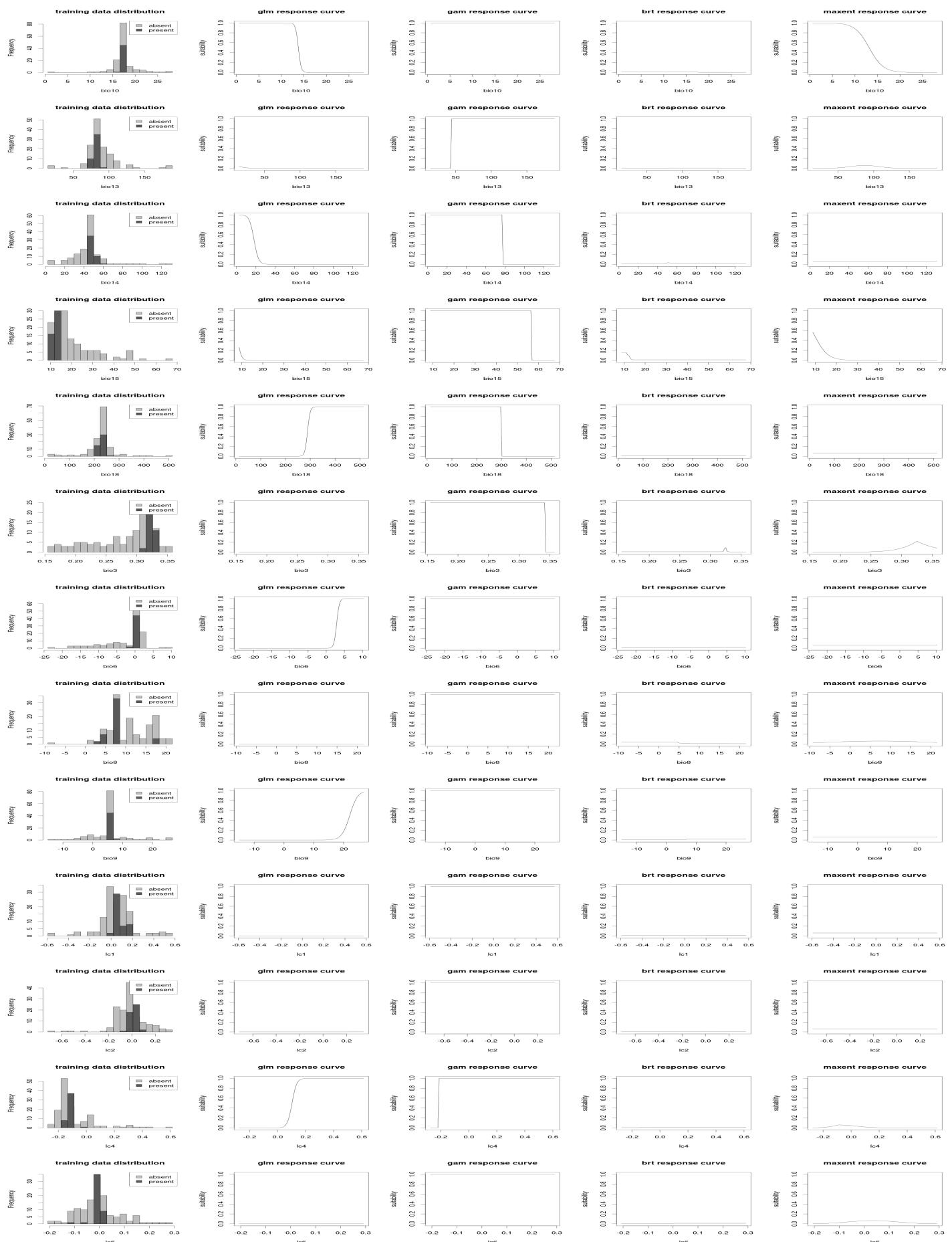


E Model response curves

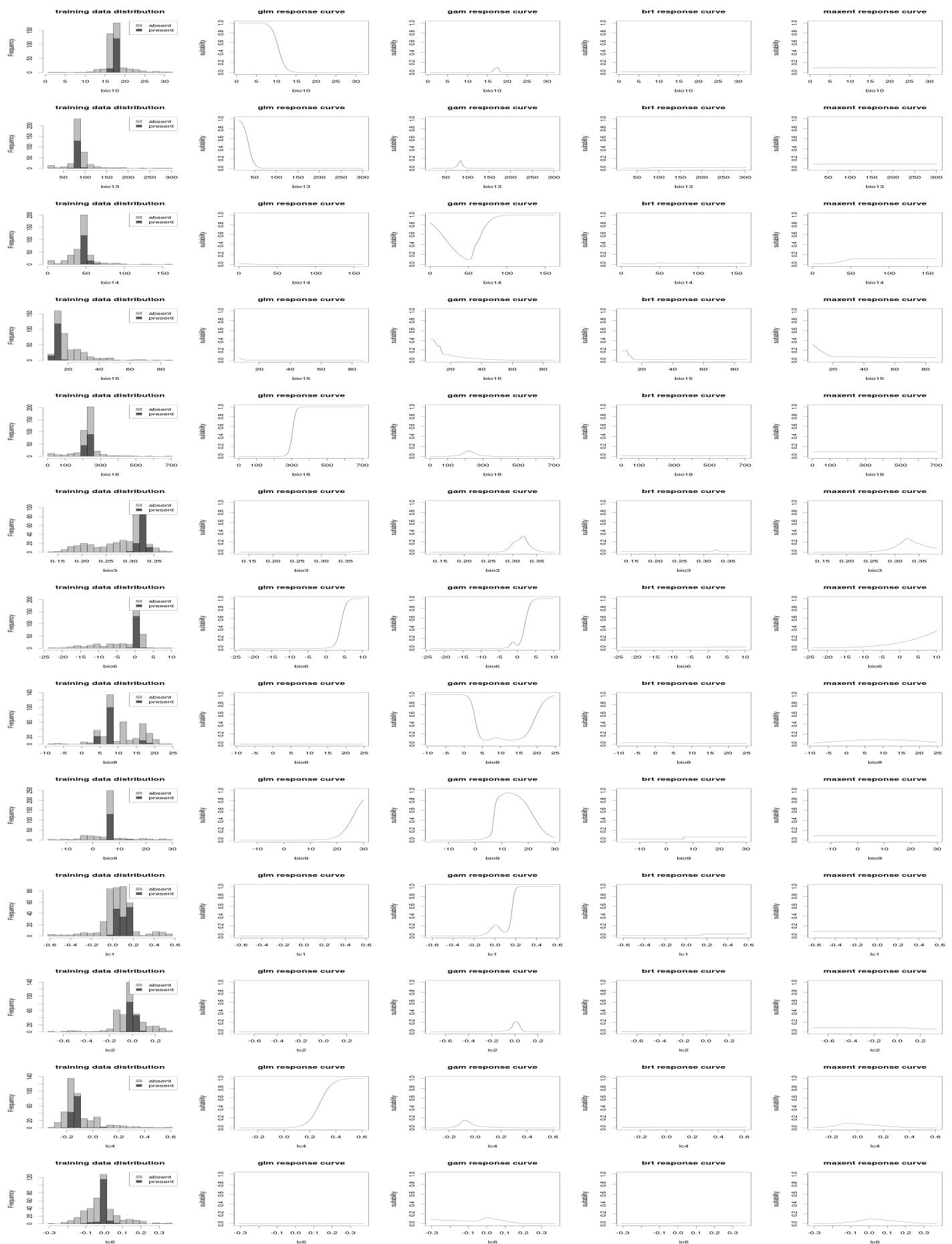
Native model response curves



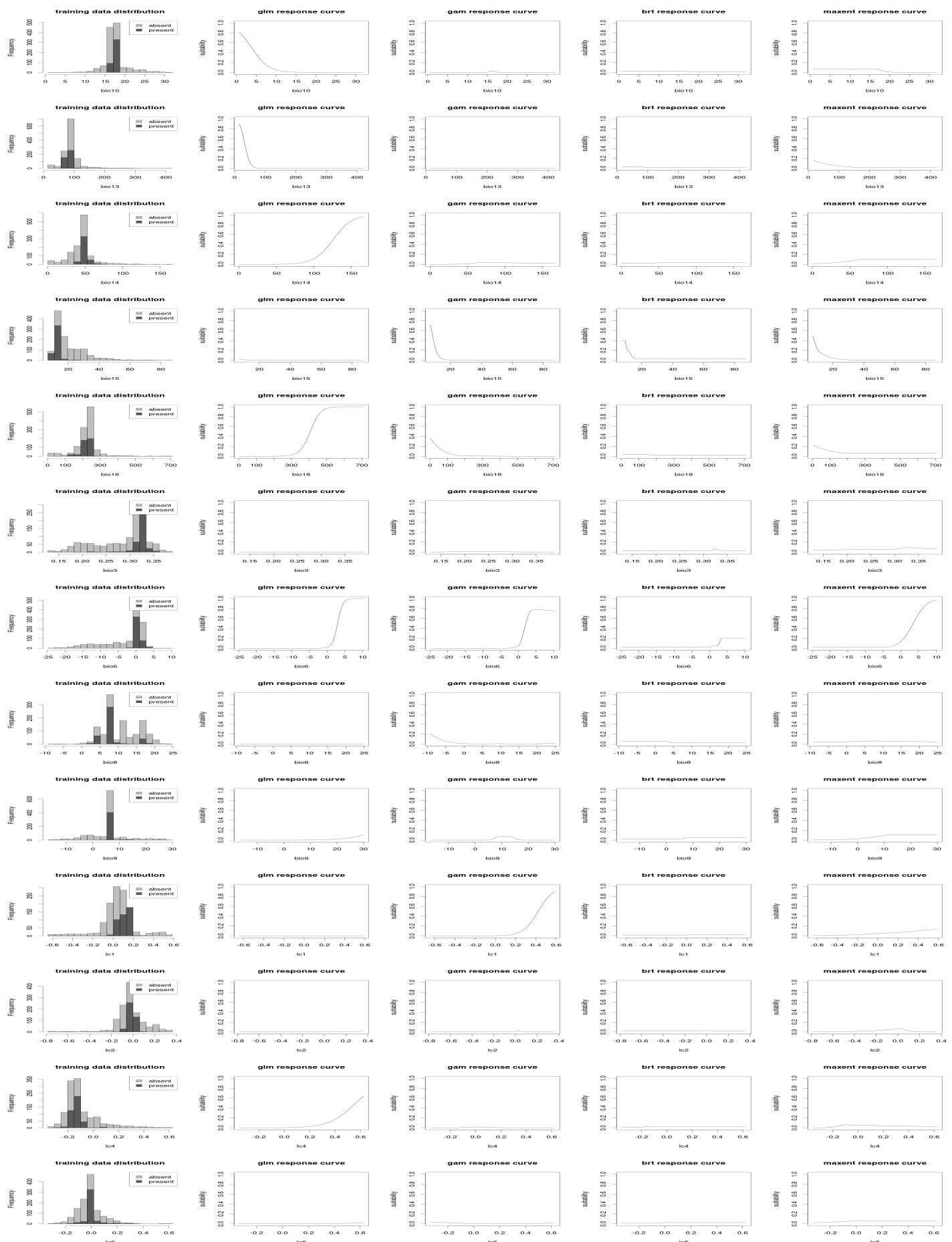
2002 model response curves



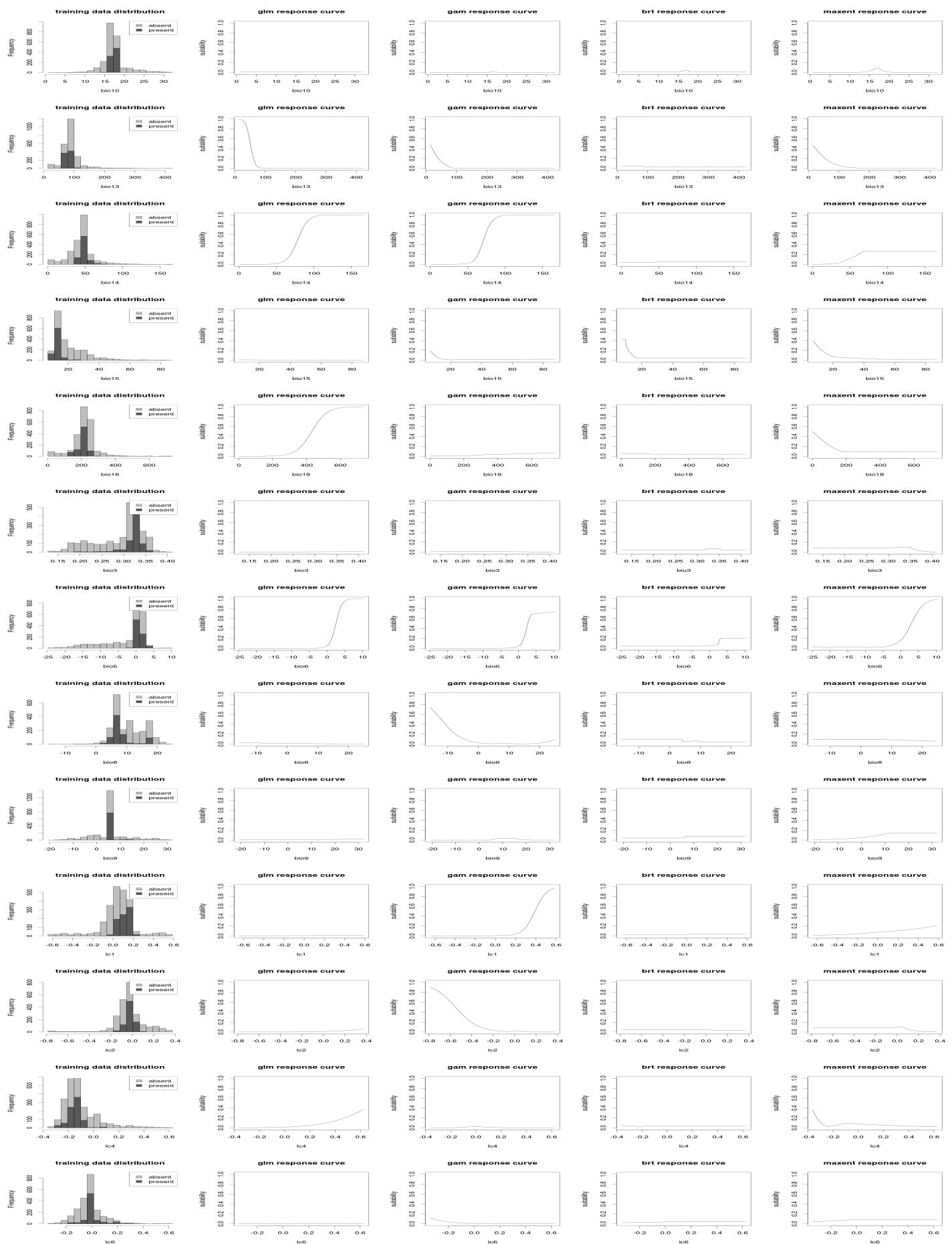
2003 model response curves



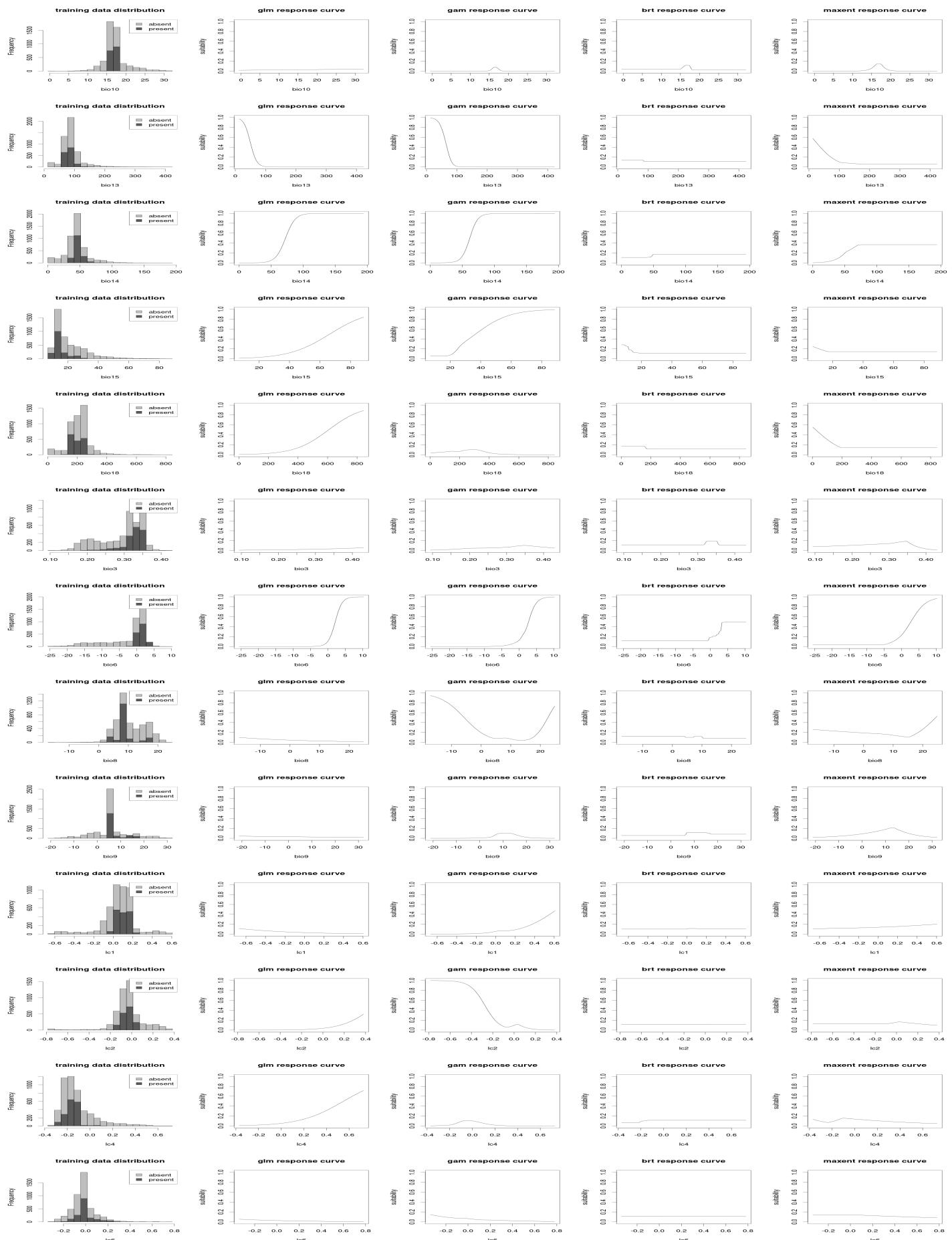
2004 model response curves



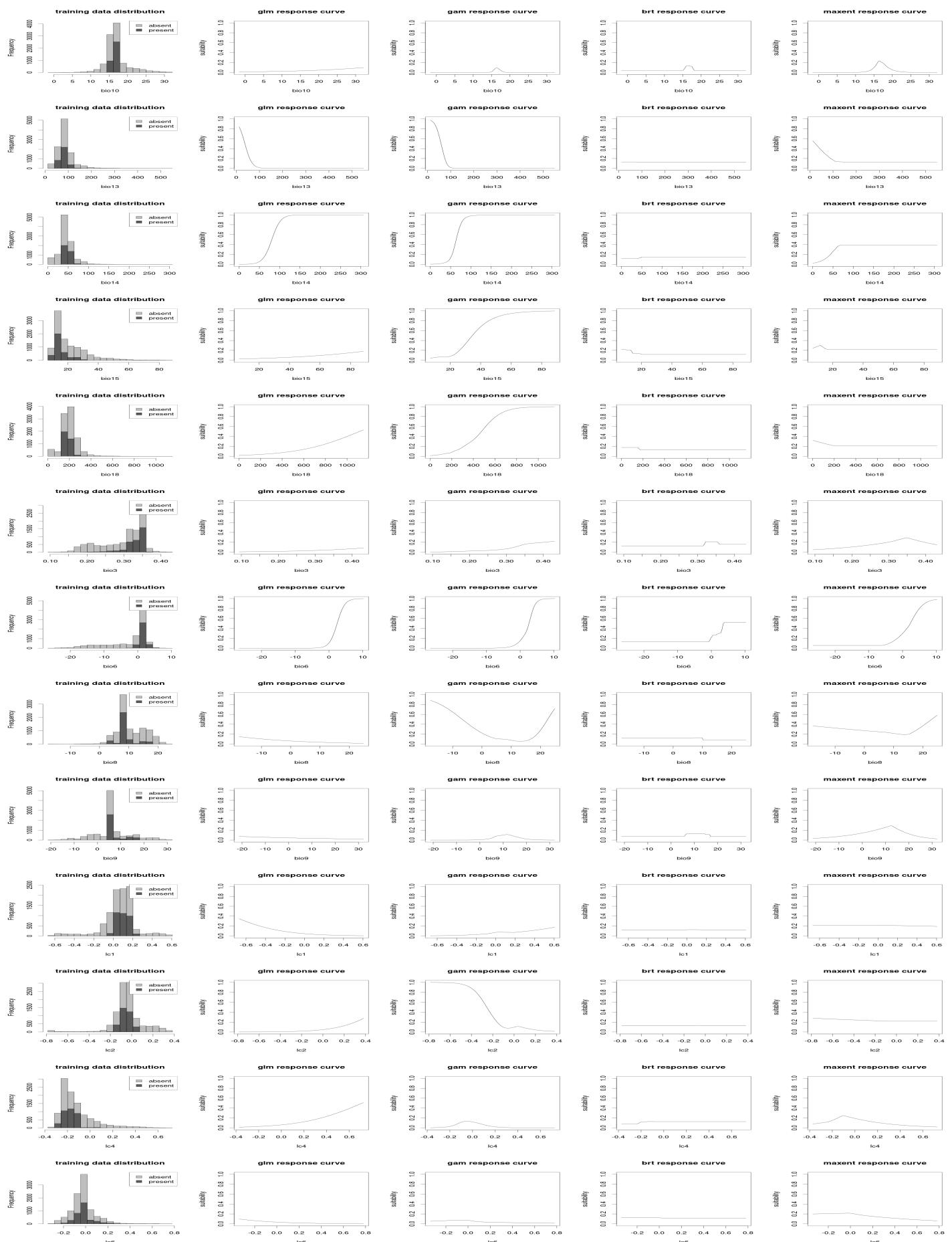
2005 model response curves



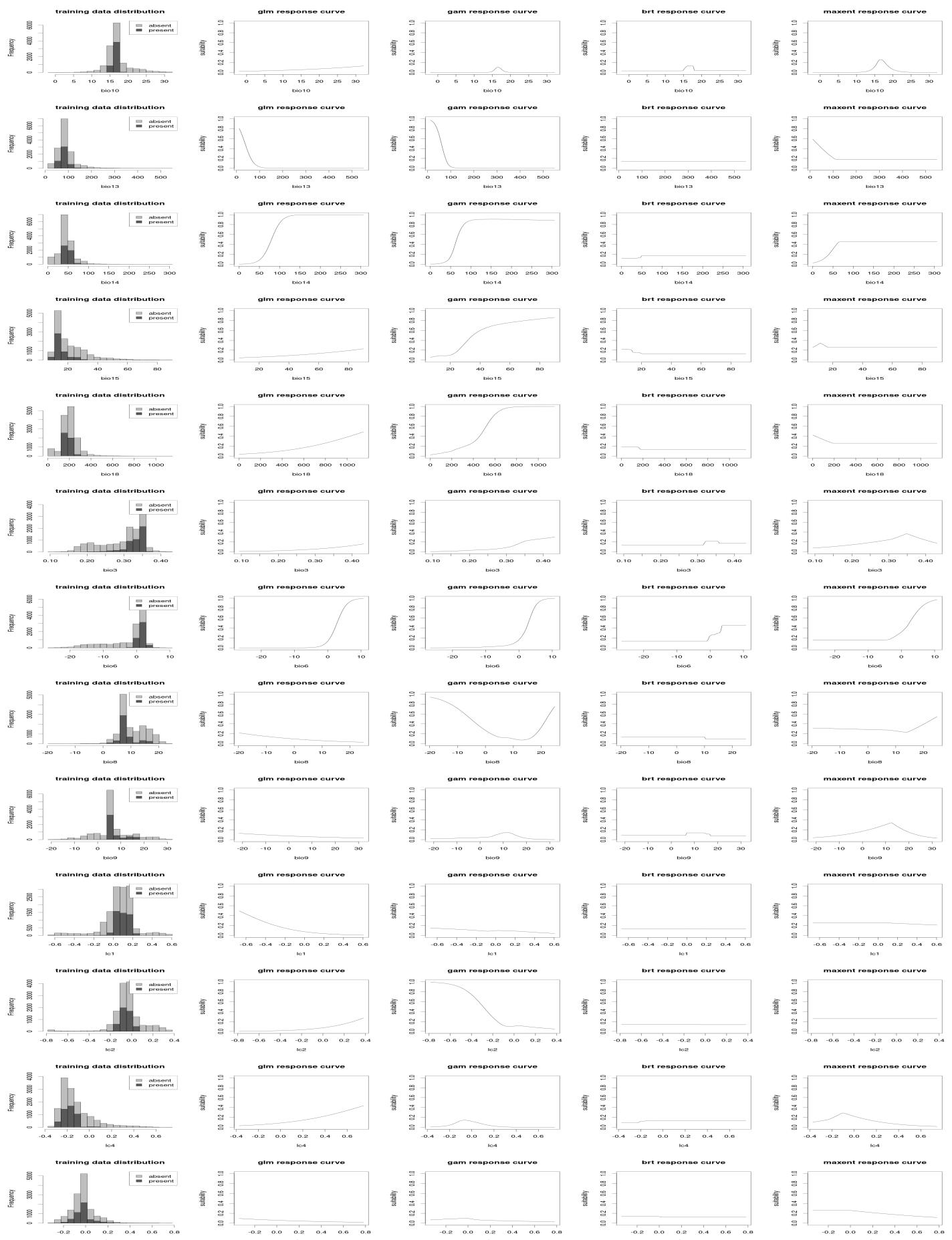
2006 model response curves



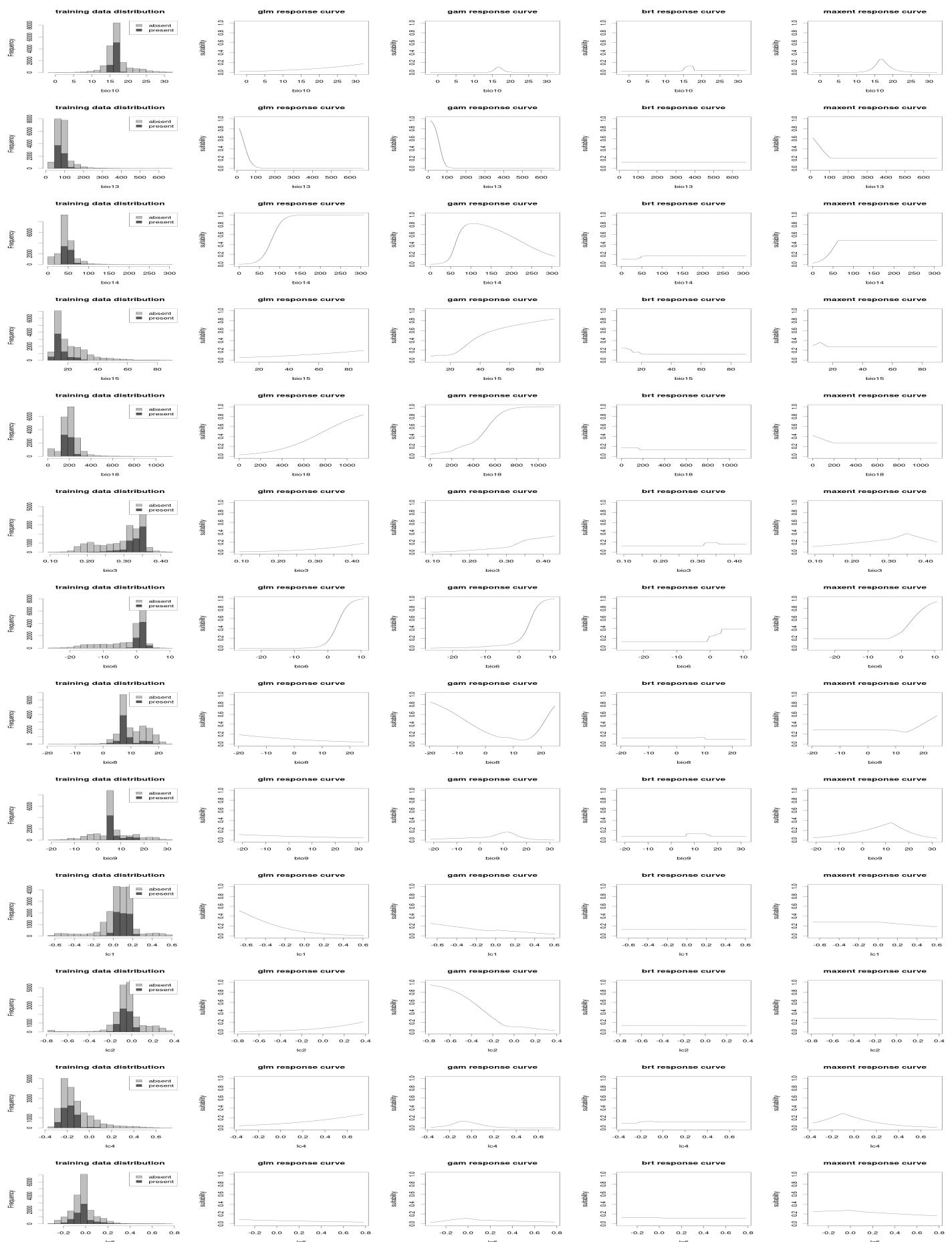
2007 model response curves



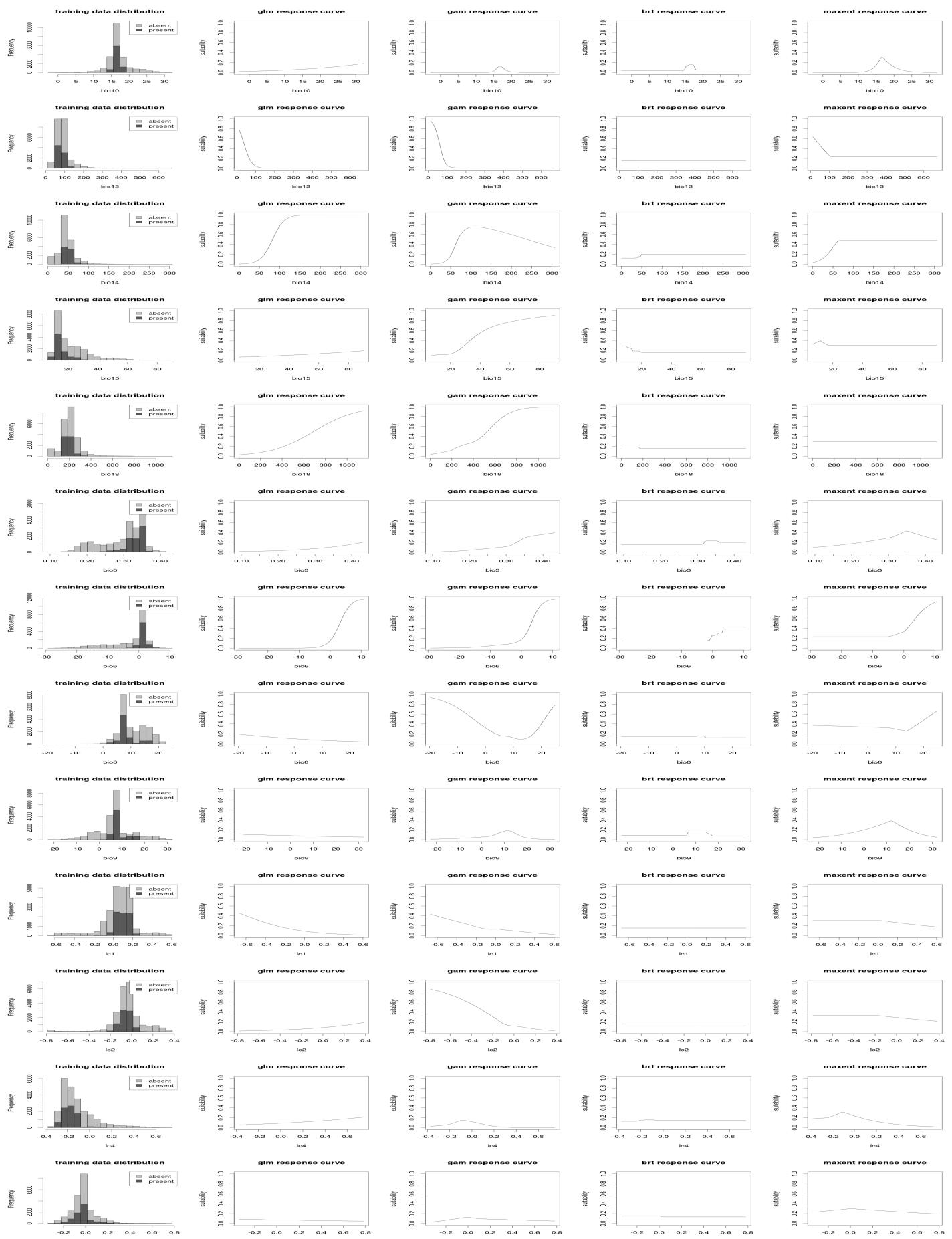
2008 model response curves



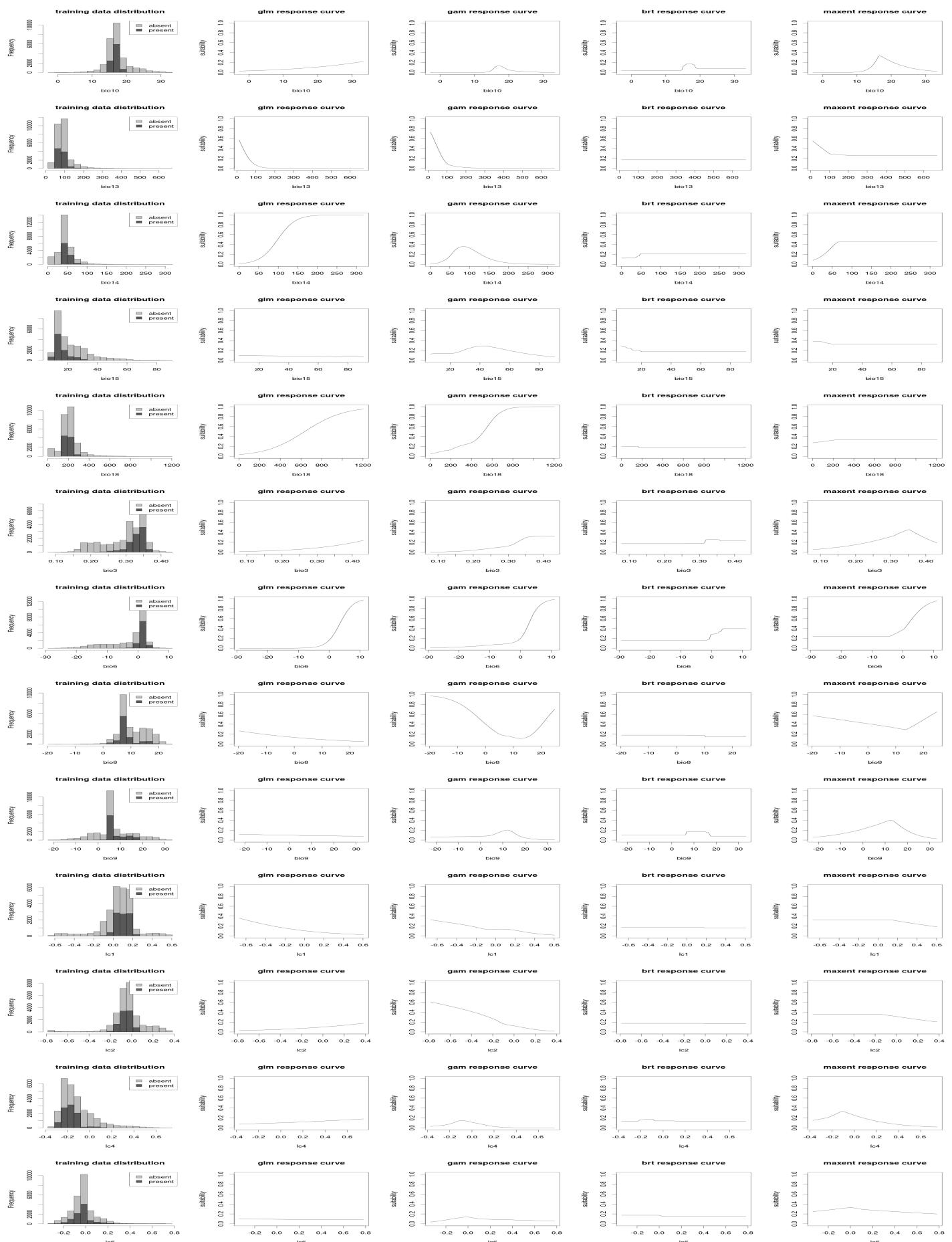
2009 model response curves



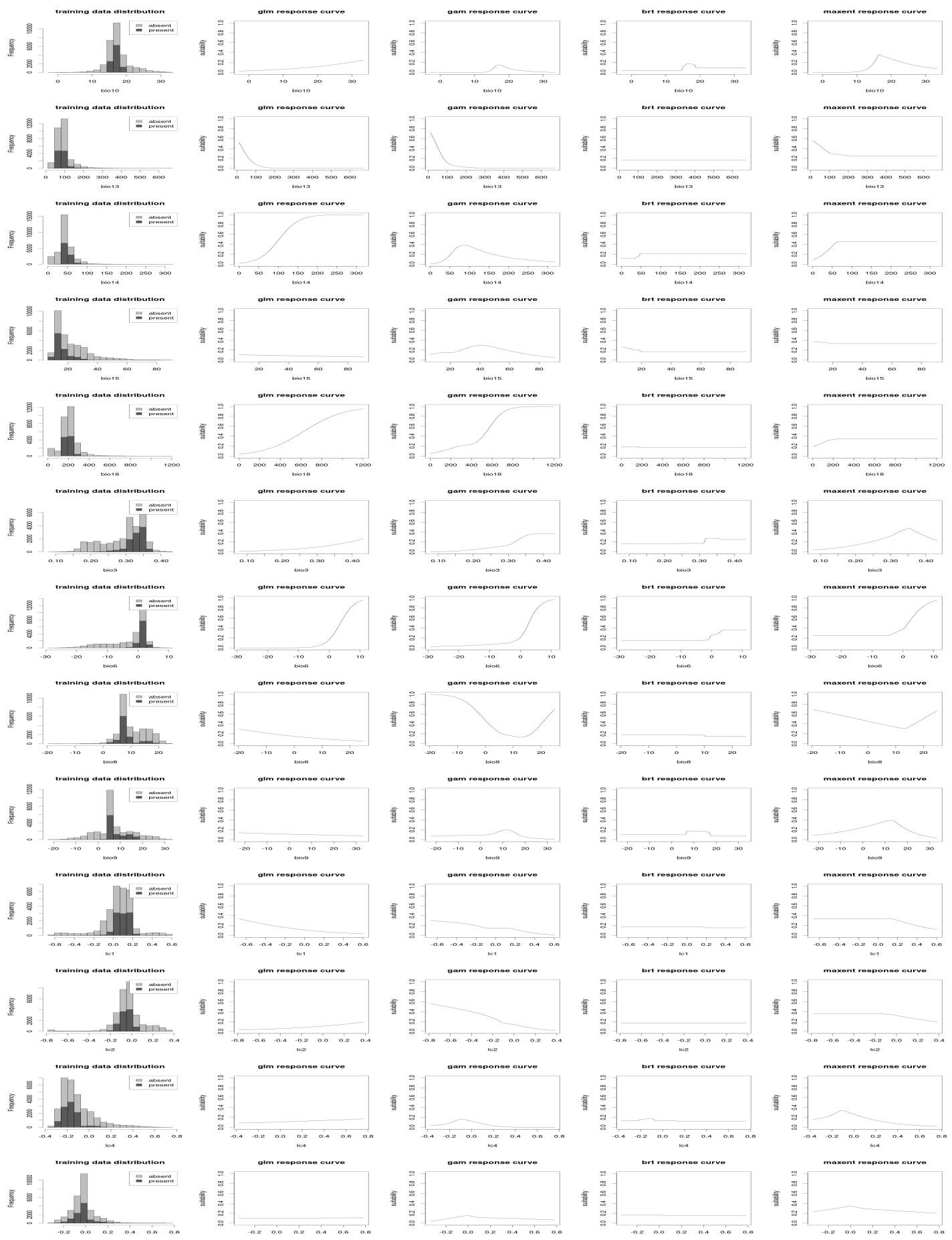
2010 model response curves



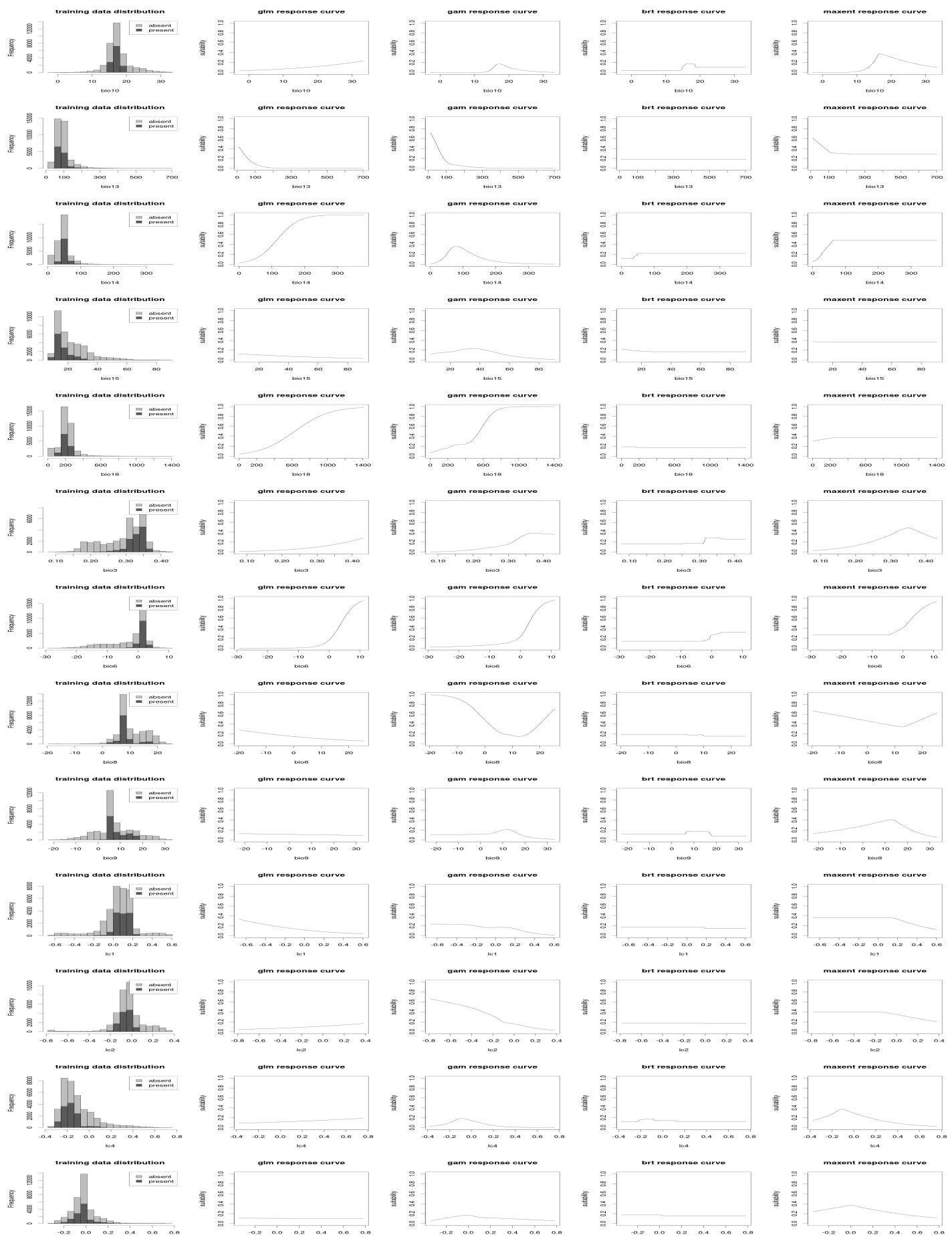
2011 model response curves



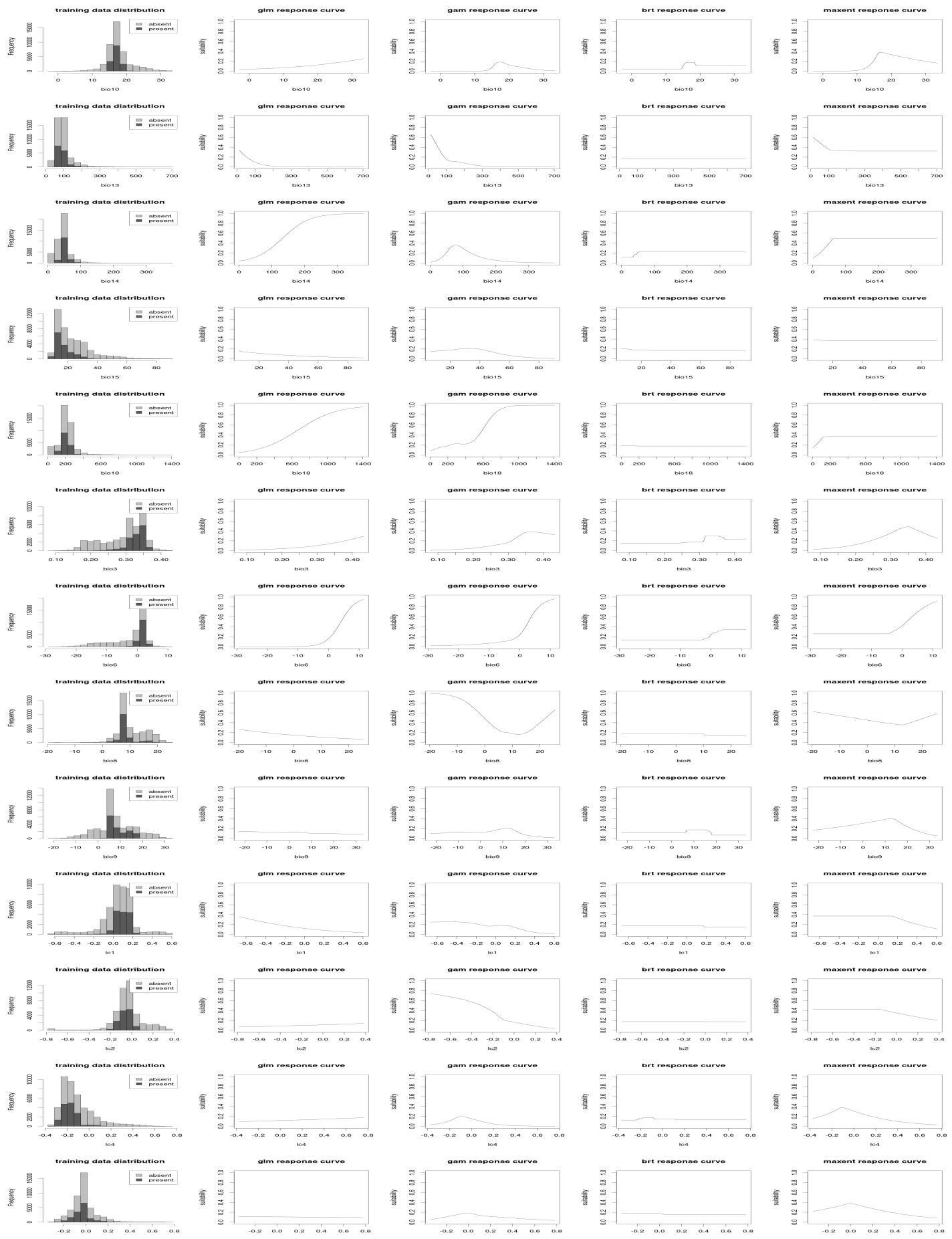
2012 model response curves



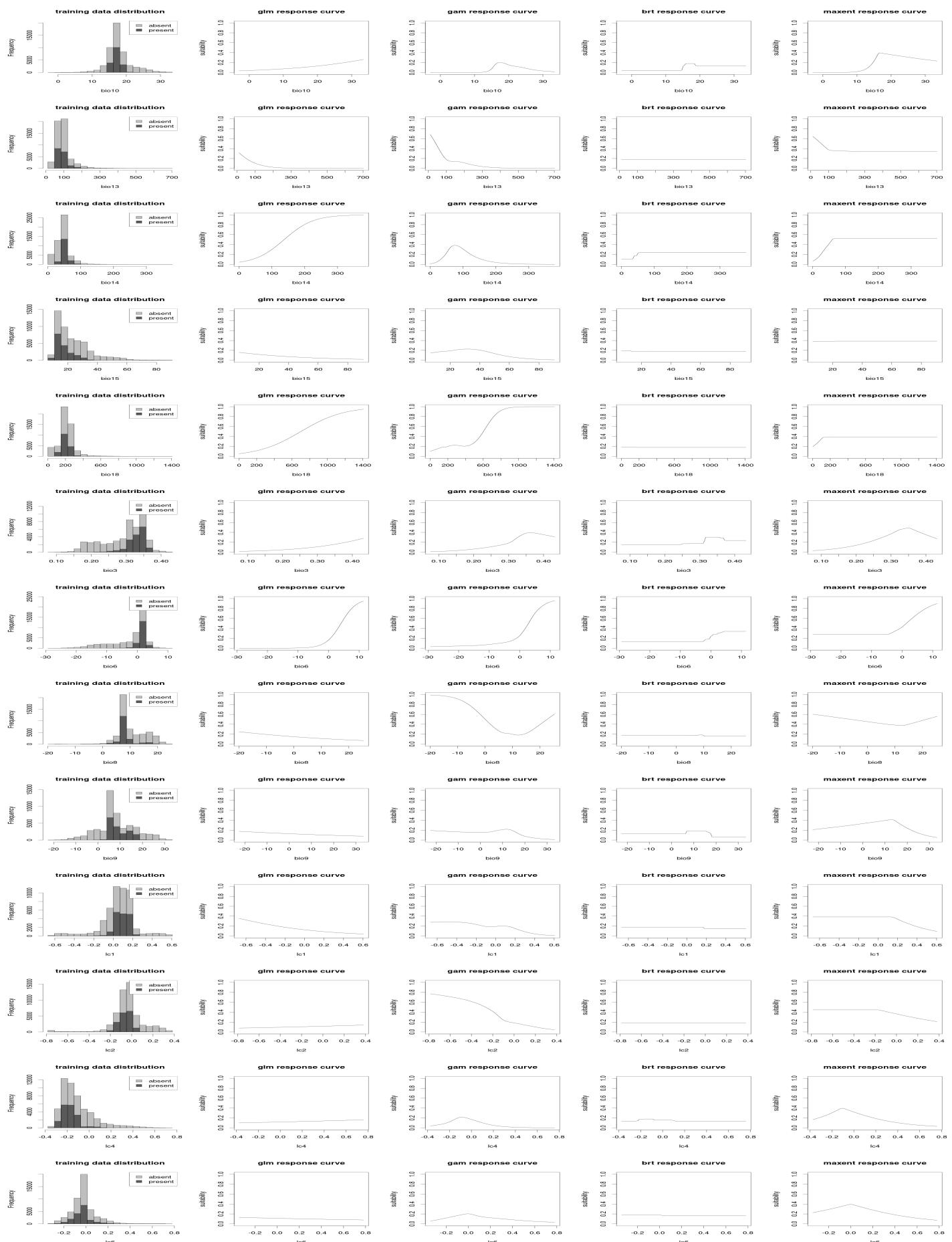
2013 model response curves



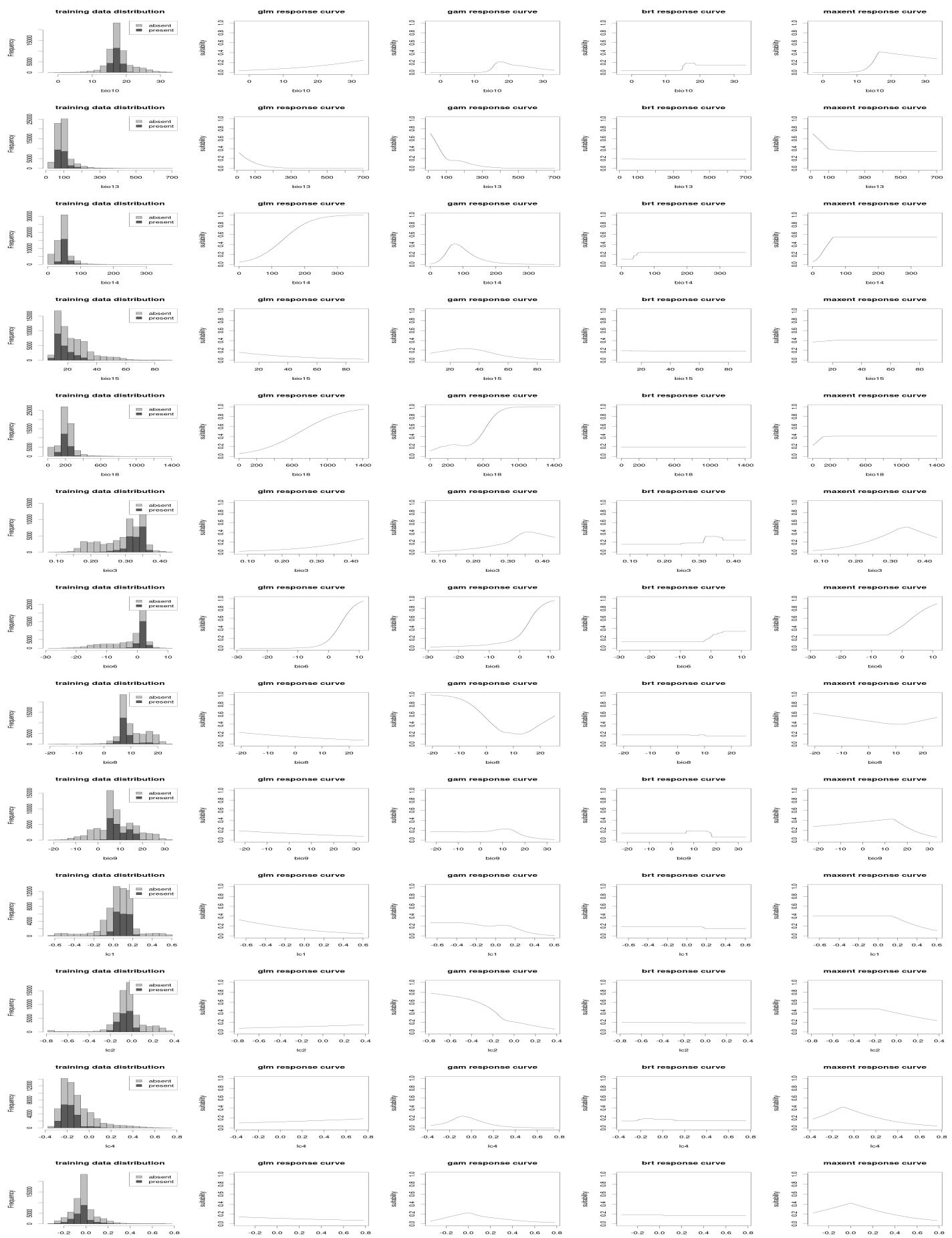
2014 model response curves



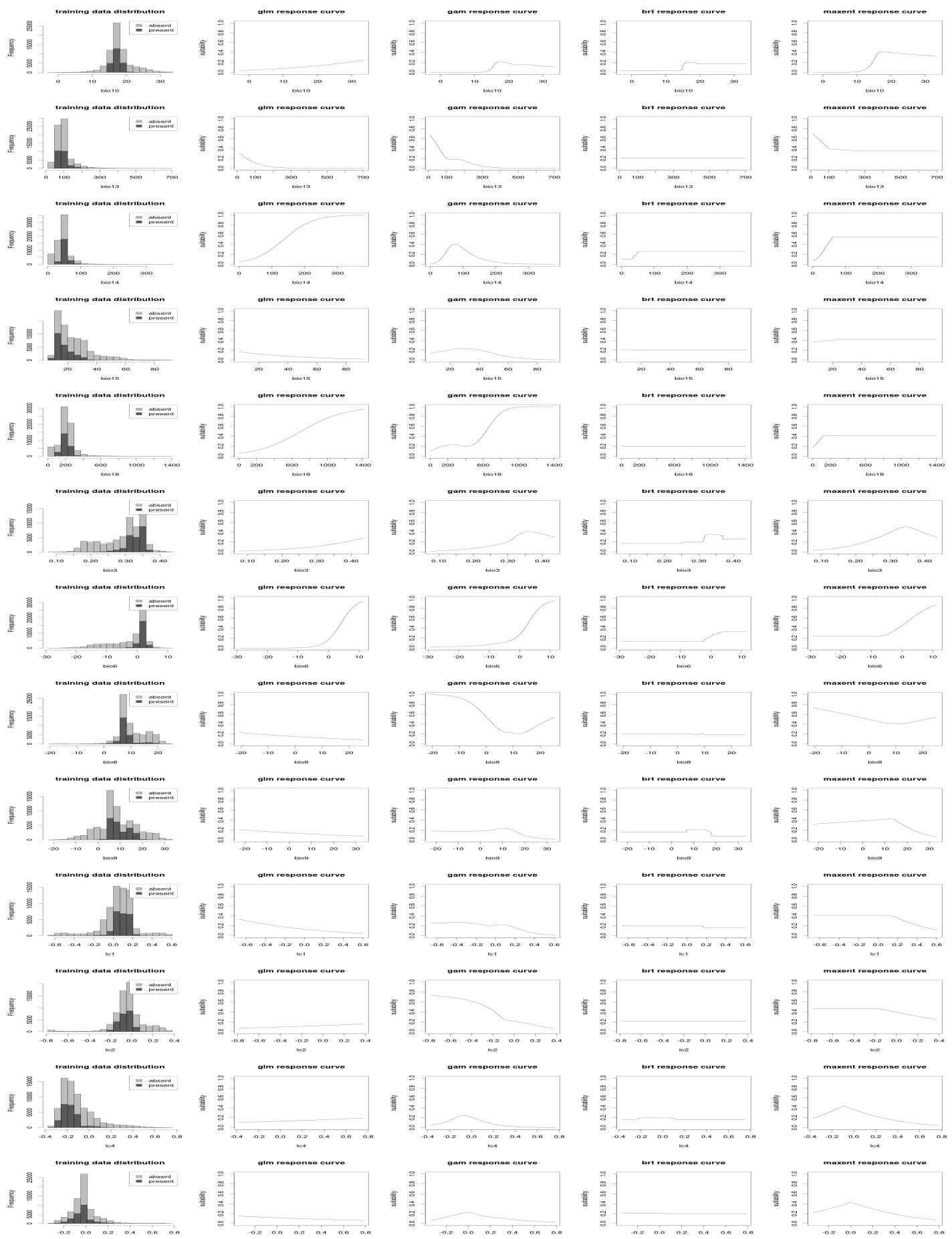
2015 model response curves



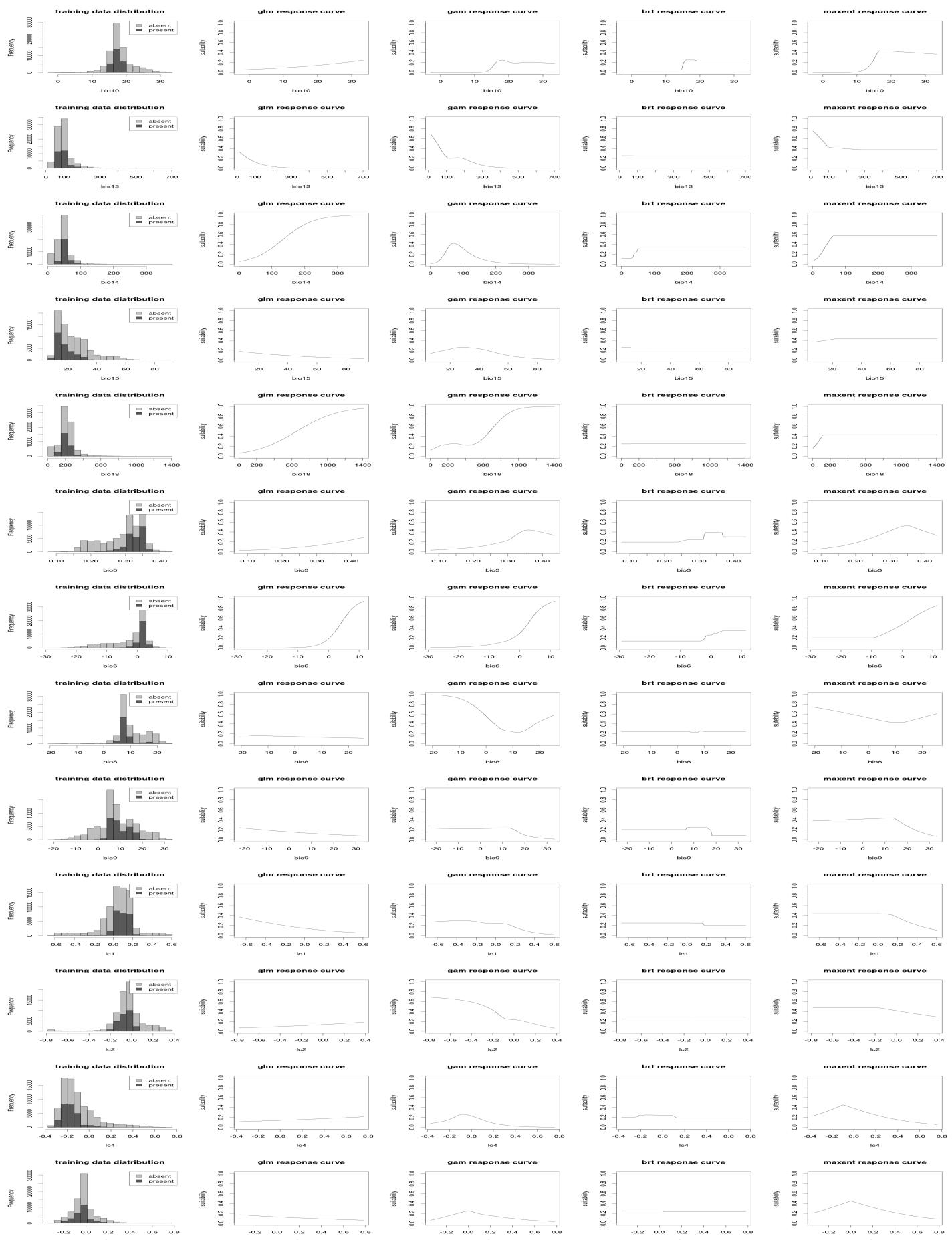
2016 model response curves



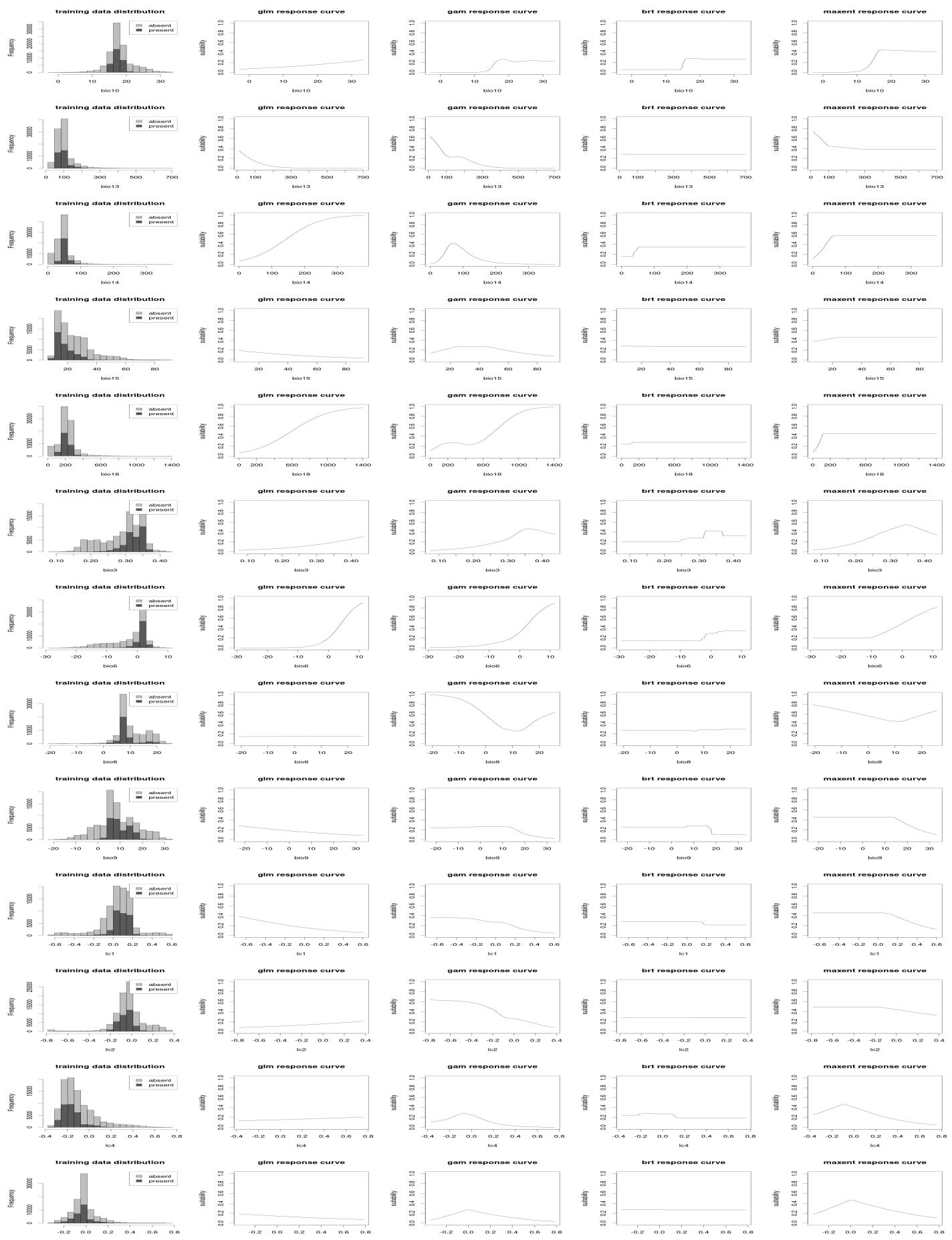
2017 model response curves



2018 model response curves



2019 model response curves



2020 model response curves

