Exploring improvements to models used in risk assessment of the Scottish monitoring programme for marine biotoxins in shellfish harvested from classified production areas

FSS/2017/020

Grietje Holtrop

Biomathematics and Statistics Scotland, Aberdeen, AB25 2ZD, UK

Address for correspondence:

Dr G. Holtrop

Biomathematics & Statistics Scotland

Foresterhill, Aberdeen AB25 2ZD

UK

Tel: +44 (0)1224 438608

Email: g.holtrop@bioss.ac.uk

Report date: May 2018

Executive summary

Food Standards Scotland has previously[1] commissioned risk assessments of the Scottish monitoring programme for marine biotoxins in shellfish harvested from classified inshore production areas in Scotland. Initially only three to six years of biotoxin test results were available, which necessitated the risk assessments being based on relatively simple model assumptions. In the current project, we look at whether some of the limiting assumptions are still necessary. Based on mussel test results from 2001-15 we looked at two aspects in particular, namely i) can the timescale be refined so that the actual date of collection is used as opposed to aggregating data by month; and ii) can the models be refined to allow for smooth progression of estimated biotoxin prevalence over time.

Smooth models were successfully fitted to the mussel biotoxin test results, providing predicted toxin prevalences that show a smooth progression from day to day throughout the year, thereby successfully addressing both aspects mentioned above. We also looked at how these predictions might affect suggested monitoring frequencies, based on Paralytic Shellfish Toxin (PST) > 400 µg/kg, Lipophilic Toxins (LT) > Maximum Permitted Level (MPL), and Domoic Acid (DA) > 5mg/kg. These were compared against the suggested monitoring frequencies obtained from the simple models employed previously. It was found, that, on the whole, there was good agreement between the suggested frequencies derived from the smooth models and those derived from the simple models.

For the LT toxins, test results until 2011 comprised of whether or not the LT level exceeded the MPL. From mid-2011 onwards test results provide actual levels of the various LT toxins, and this would potentially allow for developing models and, subsequently, monitoring schemes based on LT exceeding half the MPL , which is a more precautionary approach and which is also employed for PST. Simple models, however, could not be successfully fitted to these data, whereas the smooth modelling approach employed in the current report, was capable of successfully fitting models to these data.

The main drawback of the smooth modelling approach is that it is time consuming; the fitting routine takes several hours to complete, as compared to minutes for fitting simple models. Furthermore, for both modelling approaches it was found that occasionally the observed biotoxin trends in the data are not well captured by the model. This tends to happen in particular for groups of pods where very few positive toxic events have been observed. We therefore propose a visual display that not only shows the model predictions and observed prevalence of a given biotoxin level (at half the MPL, say), but in addition also gives an indication of the actual level observed (below limit of detection, between limit of detection and 0.5 MPL, between 0.5 MPL and MPL, exceeding MPL), and when it was observed (between 2001-5, 2006-10, 2011-15). This allows for more comprehensive integration of all the information available when developing monitoring schemes.

[1]Holtrop, G., Swan, S., Duff, B., Wilding, T, Naryanaswamy, B. & Davidson, K. (2016) Risk assessment of the Scottish monitoring programme for marine biotoxins in shellfish harvested from classified production areas: review of the current sampling scheme to develop an improved programme based on evidence of risk. Report to Food Standards Scotland, Project code FSS/2015/021. September 2016.
Holtrop, G. (2008) Risk assessment of the FSA Scotland inshore shellfish monitoring programme based on historical toxin data from 2004-2006. Report to Food Standards Agency Scotland, Project code S14036. February 2008.
Holtrop, G., & Horgan, G.W. (2004) Risk assessment of the FSA Scotland monitoring programme for biotoxins in shellfish harvested from classified inshore areas in Scotland: evaluation of the current scheme and development of improved alternatives based on historical data. Report to Food Standards Agency Scotland, Project code S01026. December 2004.

In conclusion, more refined models allow for more realistic modelling of biotoxin prevalence, in particular smooth progression of prevalence from day to day, as opposed to monthly estimates obtained from simple models employed previously. It comes at a cost though, which is that fitting these models is time consuming. It has also shown, however, that the simple models, despite their crude monthly time scale, generally capture the general behaviour of biotoxin prevalence well, albeit on a much cruder time scale. These findings suggest that for future risk assessments the simple models employed previously continue to be adequate. In addition, for key outcomes of interest (such as biotoxin test results exceeding half the MPL in indicator shellfish species) more refined models such as the smooth models presented here, should also be considered.

Contents

Glossary

Abbreviation	Description
AZA	Azaspiracid
BioSS	Biomathematics & Statistics Scotland
DA	Domoic Acid
FSA	Food Standards Agency
FSS	Food Standards Scotland
GAMM	Generalised Additive Mixed Model – used for modelling the biotoxin data
HGLM	Hierarchical Generalised Linear Model – used for modelling the biotoxin data
LCMS	Liquid Chromatography Mass Spectrometry
LT	Lipophilic Toxins
MPL	Maximum Permitted Level
N	North (used in names for pod groups)
NWC	North West Coast (used in names for pod groups)
OA	Okadaic Acid
PST	Paralytic Shellfish Toxins
SE	South east (used in names for pod groups)
SW	South West (used in names for pod groups)
W	West (used in names for pod groups)
WC	West Coast (used in names for pod groups)
YTX	Yessotoxin

1 Introduction

Food Standards Scotland (FSS) has previously (Holtrop & Horgan 2004, Holtrop 2008, Holtrop et al. 2016) commissioned risk assessments of the Scottish monitoring programme for marine biotoxins in shellfish harvested from classified production areas. Shellfish toxin test results were summarised by month for each (group of) classified production area(s), models were fitted, and findings from models and data summaries were used to assess the current sampling scheme and to develop improved schemes. Initially only three to six years of test results were available. This necessitated the risk assessments being based on model assumptions that may be too simple and may therefore be unrealistic. For example, it had to be assumed that biotoxin levels are constant for a month and then, overnight, change to a new level for the next month. The current monitoring programme has now been running for several years with test results now available for 15+ years. With this large amount of data, especially for mussels, the question arises whether some of the limiting assumptions are still necessary. In particular, can we develop models that allow for smooth progression of toxin levels over time as opposed to levels that are fixed per month.

The aim of the current study is to investigate whether the 2001-2015 time series of mussel test results is sufficiently informative to refine the current risk assessment models. We will look at two aspects in particular, namely i) can the timescale be refined so that the actual date of collection is used (as opposed to aggregating data by month); and ii) can the models be refined to allow for smooth progression of estimated biotoxin prevalence over time. Not only are such models more realistic, it will also ensure that the methodology underlying any future risk assessments will stand up to scrutiny by the scientific community. This will help to ensure that the development of future monitoring programmes and risk assessments will continue be appropriate, and optimal both in terms of timely detection of biotoxins as well as being cost-effective.

2 Materials and Methods

2.1 Data

Mussel test results were extracted from the database used in Holtrop et al. (2016). These data were thoroughly cleaned and checked, and pods (a collection of similar shellfish harvesting sites) with limited test results were carefully grouped based on the similarity in biotoxin and phytoplankton profiles, proximity and similarity in hydrographical and environmental conditions. This resulted in 105 pods being combined into 37 groups. Full details are given in Holtrop et al. 2016.

Samples were analysed for three types of toxin, namely Paralytic Shellfish Toxin (PST), Domoic Acid (DA) and Lipophilic Toxins (LT). Since the middle of 2011 the latter has been subdivided into Okadaic Acid (OA), Azaspiracid (AZA) and Yessotoxin (YTX), using Liquid Chromatography Mass Spectrometry (LCMS) methods. When any of these toxins exceed their Maximum Permitted Level (MPL) the shellfish field is closed for harvesting. The MPL for PST is 800 µg/kg shellfish flesh and for DA it is 20 mg/kg. Until 2011 the LT test result was given as ‘absent’ or ‘present’. From 2011 onwards LCMS methods have been used, and the MPL for the three LT toxins is 160 mg/kg for OA and AZA, and 3.75 mg/kg for YTX.

Risk assessment models were previously developed based on samples exceeding 0.5 MPL for PST, 5 mg/kg for DA, and exceeding MPL for LT. Since 2011 LT test results provide more detailed information on actual biotoxin levels measured, offering the possibility of working with 0.5 MPL as a cut-off. To facilitate this, the LT-LCMS data for the three biotoxins OA, AZA and YTX were combined and classified into one biotoxin test result, as follows:

- Classified as 0: OA = 0 and AZA=0 and YTX = 0 mg/kg.
- Classified as 0 – 0.5 MPL: At least one of OA, AZA or YTX testing positive but all three biotoxin test results are less than 0.5 MPL.
- Classified as 0.5 MPL – MPL: At least one of OA, AZA or YTX exceeding 0.5 MPL but less than MPL, and with all three biotoxin test results less than MPL.
- Classified as ≥ MPL: OA ≥ 160 or AZA ≥ 160 or YTX ≥ 3.75 mg/kg.

The mussel sample test results are summarised in Table B1[1]. Throughout, measured values that were below the limit of detection are denoted as 0.

2.2 Model formulation

Test results were formulated as 0 (below a given limit of interest, such as 0.5 MPL) or 1 (exceeding this given limit). Models were fitted to the proportion of mussel samples exceeding this limit, as follows. For a given biotoxin, let p be the probability that a sample is positive (i.e. the toxin level exceeds a given limit). This probability is likely to depend on the time of year (e.g. high values are more likely to occur in summer than in winter) and the location the sample was taken from. There may also be year to year fluctuations with some years showing higher prevalence than others.

2.2.1 Previous models

Previously (Holtrop et al. (2016), Holtrop (2008), Holtrop & Horgan (2004)), and an in-house risk assessment by Food Standards Agency (FSA) statisticians was conducted in 2012, unpublished) such relationships were investigated using a Hierarchical Generalised Linear Model (HGLM, see Lee & Nelder (1996, 2001)), such that estimated prevalence was obtained for each month of the year for each group of pods. Let y_{mgt} be the number of samples exceeding a given limit and let n_{mgt} be the total number of samples, for month m at pod group g in year t. Then y is assumed to follow a binomial distribution:

$$y_{mgt} \sim \text{Binomial}(n_{mgt}, p)$$

where the probability p of a sample exceeding a given limit is modelled as a function of month, group (of pods) and year. Let the odds be defined as $p/(1\text{-}p)$. The following linear model was formulated for the log-odds:

$$\ln\left(\frac{p}{1\text{-}p}\right) = \text{constant} + \text{Month}_m + \text{Group}_g + \text{Year}_t + p_g.\text{Month}_m \qquad (1)$$

with $\ln(.)$ denoting the natural logarithm. Month was regarded as a fixed effect and Group and Year as random effects, i.e. on the log-odds scale, Group and Year effects were assumed to have Normal distributions with a mean of zero and unknown between-group or between-year variances of σ_g^2 and

[1]Presented in Appendix B.

σ_t^2, respectively. In addition, a term reflecting the interaction between Month and Group was also included (i.e. the prevalence over months of the year is group-specific). These models resulted in an estimated prevalence that is fixed during a given month, and then changes to a new level for the next month.

2.2.2 Revised models

In the current report we investigate whether biotoxin prevalence can be estimated in a more realistic manner, in particular,

- Refinement of the time scale through using the date of collection as opposed to aggregating data by month, and
- allowing for smooth progression of estimated biotoxin prevalence over time.

Generalised Additive Mixed Models (GAMM) (Wood, (2006)) were chosen as these models allow for smooth estimated curves, can easily handle irregularly spaced data (in our case: irregular sampling frequencies), and do not require any a priori knowledge about prevalence patterns.

The date of collection was translated into day of year (day 1, 2, ….365, ignoring the extra day in leap years) and the estimated curve was allowed to change smoothly from day to day, with the extra condition that progression from 31 December to 1 January was also continuous. Each (group of) pod(s) was allowed its own smooth toxin profile. Furthermore, year was regarded as a random effect. Model (1) was replaced with

$$\ln\left(\frac{\mathrm{p}}{1\text{-}\mathrm{p}}\right) = \mathrm{smooth}(\mathrm{Day}, \mathrm{by}{=}\mathrm{Group}) + \mathrm{Group_g} + \mathrm{Year_t} \qquad (2)$$

These models were fitted in R (R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/), using the R library gamm4 (gamm4: Generalized additive mixed models using mgcv and lme4, S. Wood, F. Scheipl – R package version 0.2-3, 2014). This routine aims to fit a smooth curve to the data such that goodness of fit (how well does the model describe the data) is balanced against overfitting, based on the noise in the data. Technical details are provided in Appendix A.

Throughout this report, the generalized additive model described above will be referred to as the 'smooth model', and the model employed in Holtrop et al. (2016) where for each group the prevalence within a month is assumed constant, will be referred to as the 'simple model'. Furthermore, all predicted prevalences are based on an average toxin year, for both simple and smooth models, unless mentioned otherwise.

2.2.3 Risk assessment

We will follow the methodology developed previously (Holtrop et al. (2016)) and aim to keep the risk of not detecting a toxic event[1] to a minimum. The risk of non-detection can be defined as the chance

[1]Toxic event: sample exceeds biotoxin limit of interest.

that biotoxin levels unknowingly exceed a given limit in a particular week. In other words, it looks at the probability that the pod is not sampled while toxin levels exceed a given limit (such as MPL).

The risk of non-detection depends on two factors, namely

a) the chance that the field is toxic (i.e. probability that toxin levels exceed a given limit), and
b) the sampling frequency.

An increase in biotoxin prevalence or a decrease in the sampling frequency lead to an increased risk of non-detection.

For simplicity it is assumed that

- the sample tested is representative of the entire harvesting area of interest,
- the test result is accurate (i.e. its reading reflects the true toxin level),
- the test result is valid for one week.

These three assumptions imply that when a sample gives a test result below the limit of interest then the biotoxin levels of shellfish in that particular harvesting area will remain below this limit for the entire week. As a consequence, weekly sampling results in a risk of non-detection of 0%. If samples were taken every fortnight, the risk is 0.5p (for every two weeks there was one week that the risk of non-detection was zero and one week that the risk of non-detection was equal to the prevalence p, so is (0+p)/2 = 0.5p on average). If samples were taken every four weeks, the risk of non-detection is 0.75p (for every four weeks there was one week with zero risk of non-detection and three weeks with risk of p, which gives (0 + p + p + p)/4 = 0.75p on average). To summarise:

- Weekly sampling: risk of non-detection is zero.
- Fortnightly sampling: risk of non-detection is 0.5p.
- Monthly sampling: risk of non-detection is 0.75p.

To keep the risk of non-detection below 1%, monthly sampling would be acceptable when the estimated prevalence p is less than 1.33%, weekly sampling would be required when p exceeds 2%, and fortnightly sampling would be required for p between 1.33 and 2%.

2.2.4 Biotoxin levels considered

Smooth models were successfully fitted to the mussel biotoxin data from 2001-15 for LT > MPL, PST > 0, 400, 800 µg/kg and DA > 0 and > 5 mg/kg. The LT-LCMS samples from 2011-15 were analysed for AZA, OA and YTX toxins and these were summarised according to whether any of the observed levels exceeded (half) their corresponding MPL. Although the simple models employed in Holtrop et al. (2016) failed to achieve a fit to these data the smooth modelling approach was more successful, and therefore results from LT-LCMS > 0.5 MPL and > MPL are also presented.

We focus on the biotoxin levels used for main risk assessment in Holtrop et al. (2016), namely LT > MPL, PST > 0.5 MPL, DA > 5 mg/kg (all based on 2001-15 data). In addition, results for LT LCMS > 0.5 MPL (based on 2011-15 data) are also presented. For brevity, examples that illustrate key points will be shown in the main text. The full model fits for each of theses toxin levels and each group are given in Appendix C. The results for the remaining toxin levels, namely PST > 0 µg/kg, PST > MPL, DA > 0 mg/kg (all based on 2001-15 data) and for LT LCMS > MPL (2011-15 data) are presented in Appendix E.

3 Results

For each toxin level of interest, a smooth model was fitted to the mussel data available from all 37 groups such that each group was allowed its own smooth curve. The running time for fitting the smooth models is noticeably longer than for the simple models used previously; hours as compared to minutes. Akaike's Information Criterion (AIC) was lower for the smooth models compared to the simple models (AIC values presented in Appendix A), suggesting that the smooth models are more appropriate.

3.1 Examples of fitted smooth curves

Figure 1 shows examples of how the smooth model predicts a smooth progression in prevalence. Along the x-axis the 365 days of the year are shown, and light blue ticks just below the x-axis indicate when a test result was below the limit of interest. Blue ticks just above the x-axis indicate when a sample exceeded the limit of interest. The black ticks just below the x-axis indicate the start of each month. The data are shown as the percentage of samples exceeding the limit of interest for a given month (blue circles). The black curve shows the predicted prevalence for an average biotoxin prevalence year based on smooth models.

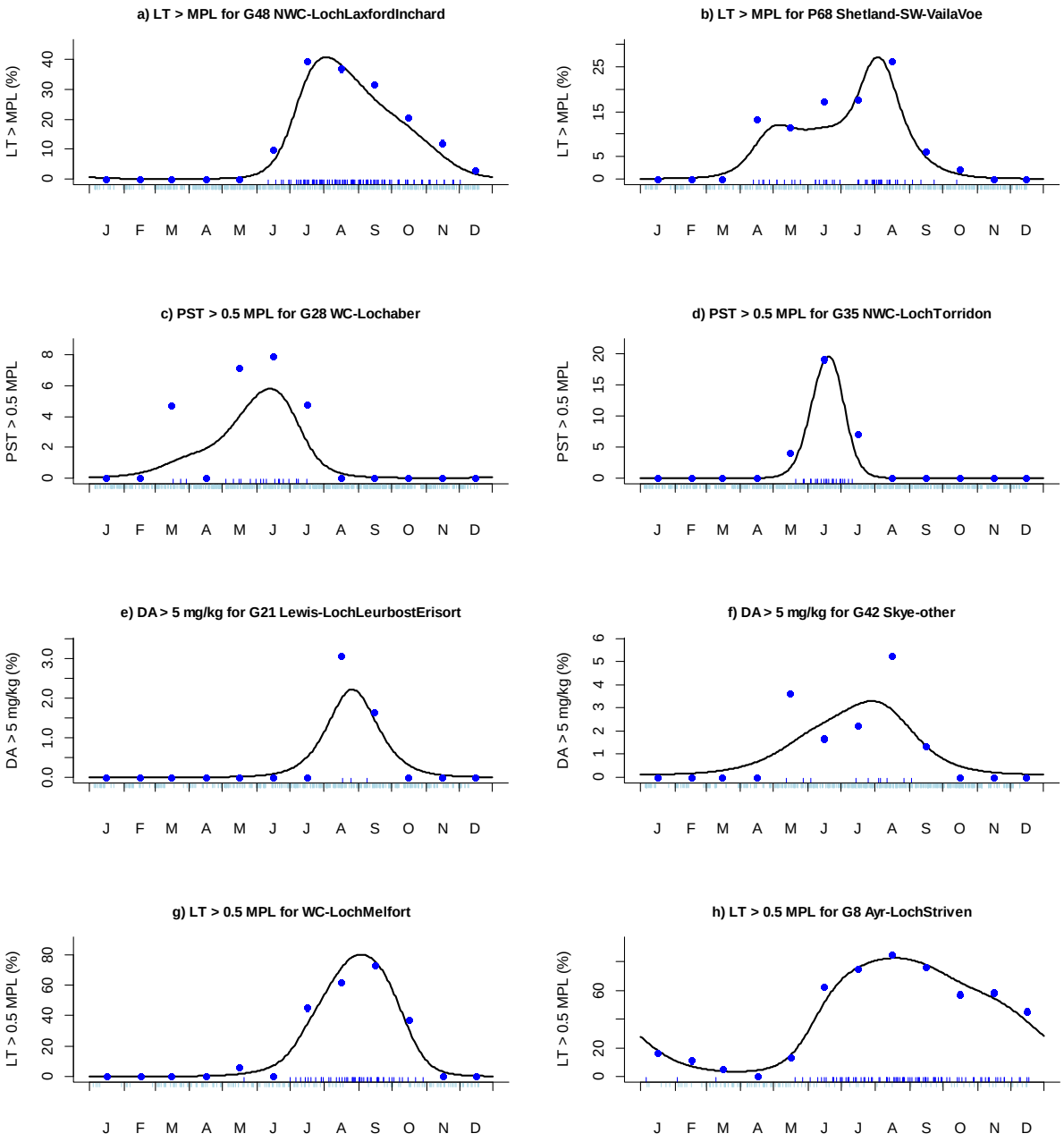


Figure 1: Examples of smooth model fits for a selection of groups and biotoxin levels of interest. Figures a-f are based on data from 2001-15. Figures g-h are based on LT LCMS data from 2011-16. For each (group of) pod(s) the prevalence is shown, based on data (blue circles), and predicted prevalence for an average toxin year from smooth models (black curve). The x-axis shows the days of the year with ticks indicating when samples were obtained and whether their test results was below (light blue) or above (blue) the limit of interest.

The fitted smooth curve attempts to capture the 'denseness' of positive samples, as indicated by the blue ticks above the x-axis. This can be a single peak (see for example Figure 1a and 1d) but can also be a more prolonged prevalence pattern; Figure 1b shows an early plateau followed with a peak later in summer. Figure 1e shows an example of only a small number of positive samples and the model captures these successfully. The smooth model assumes continuation of prevalence levels between the end of Dec and the start of January and this is clearly shown in Figure 1h.

When all the samples tested below the limit of interest the smooth curve estimated the prevalence to be zero throughout the year (illustrated in Figure 2a). When only a few samples exceeded the limit of interest and these samples were spread across the year, the smooth model predicts a low, constant prevalence throughout (Figure 2b). On the other hand, when only a few samples exceeded the limit of interest and these samples occurred close together, then the smooth model predicts a peak in prevalence (illustrated in Figure 1e).

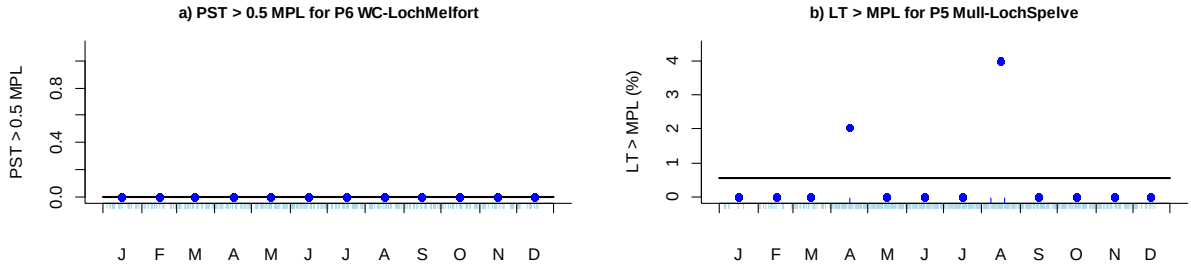


Figure 2: More examples of smooth model fits. Figure a shows an example where all the data from a pod tested below the limit of interest, and Figure b shows an example where very few samples exceeded the limit of interest. The prevalence is shown based on data (blue circles), and predicted prevalence for an average toxin year from smooth models (black curve). The x-axis shows the days of the year with ticks indicating when samples were obtained and whether their test results was below (light blue) or above (blue) the limit of interest.

3.2 Comparison against simple models

Figure 3 shows some examples comparing the fit from simple models (where prevalence is assumed constant throughout a month and then changes to a new level at the start of the next month, Holtrop et al. 2016) against those of the smooth models. These examples were chosen to highlight aspects of interest, with the complete results given in Appendices C and E. Generally, there is good agreement between the proportion of samples exceeding a limit of interest as observed in the data, the predictions from the simple model and those from the smooth model. See for example Figure 3a and 3d. When all test results are below the limit of interest, the smooth model predicts prevalence to be zero, the simple model assumes a slightly elevated, but low, prevalence (Figure 3e). It should be noted however that in the majority of these cases this prevalence is still below 1.33% so that monthly sampling would be regarded safe (more on this in the next Section). Figure 3c shows an example where one sample exceeded the limit of interest, and the simple model follows this pattern correctly. On the other hand, Figure 3b shows an example of two data points exceeding the limit of interest, one in May and one in November, and here the simple model fails to reproduce this pattern; it predicts a peak in August with low prevalence in both May and November.

An example of the simple model showing a large overnight change in biotoxin prevalence is shown in Figure 3f. The estimated prevalence of DA> 5mg/kg is 7% in September, and drops overnight to 0.5% in October. The prediction from the smooth model is more realistic here.

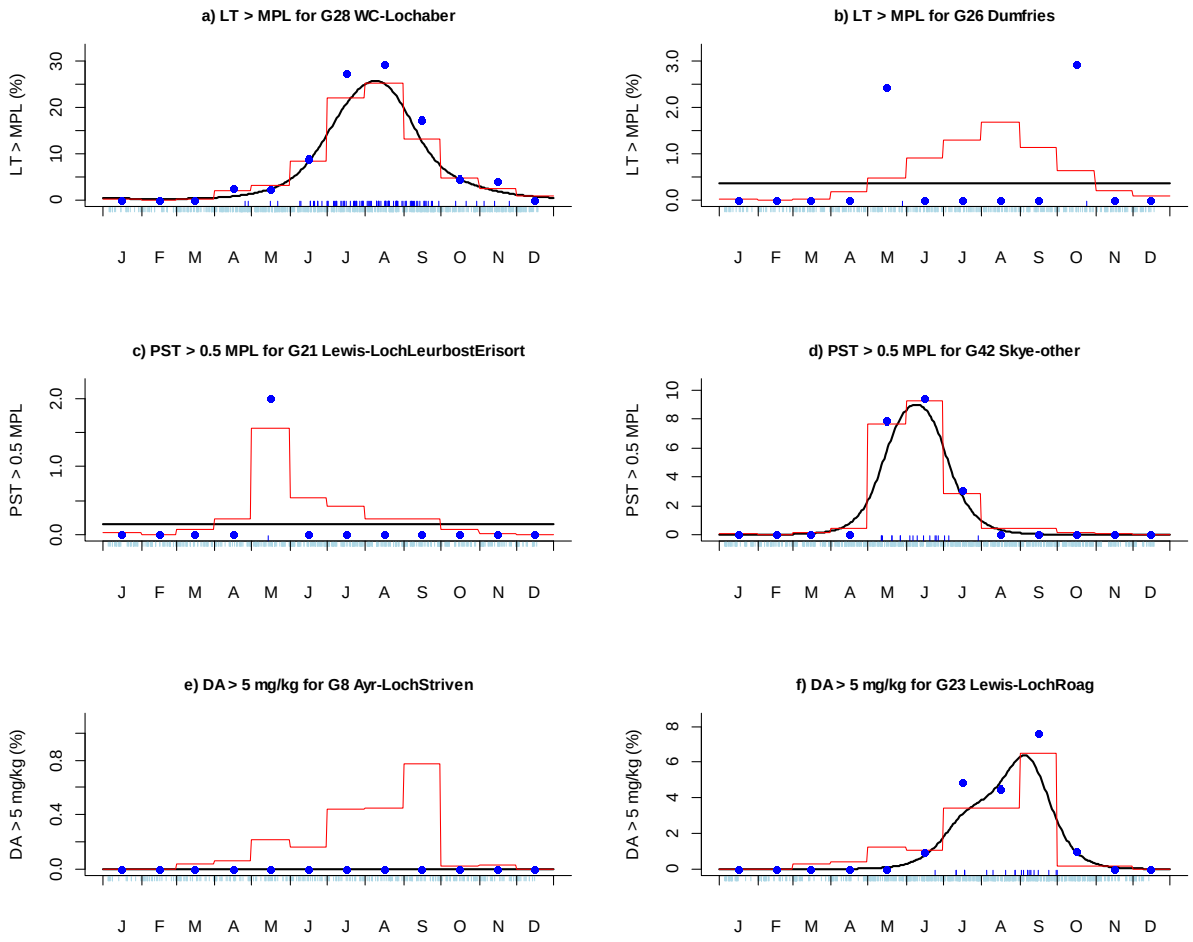


Figure 3: Examples of comparing model fit from smooth model against fit obtained from simple model. For each group of pods the prevalence is shown, based on data (blue circles), predicted prevalence for an average toxin year from smooth models (black curve), and from simple models (red lines, based on Holtrop et al. (2016)). The x-axis shows the days of the year with ticks indicating when samples were obtained and whether their test results was below (light blue) or above (blue) the limit of interest.

3.3 Implications for monitoring schemes

As explained in the Materials and Methods, when the predicted prevalence is less than 1.33% monthly sampling is deemed to be safe (i.e. the risk of not detecting a toxic event is less than 1%). When the predicted prevalence exceeds 2% weekly monitoring would be required, whilst fortnightly monitoring would be required when the predicted prevalence lies between 1.33and 2%. Tables 1-3 compare the required monitoring frequency based on predictions from the smooth model, the simple model (as per Holtrop et al. (2016)), and based on the data. In addition, possible monitoring frequencies based on the results from the smooth model fitted to the LT LCMS data exceeding 0.5 MPL are presented in Table 4.

Table 1: Sampling frequency required to keep the risk of non-detection of LT in mussels exceeding the MPL below 1% for an average year. Left hand side: based on smooth models. Middle section: based on simple models (as presented in Holtrop et al. (2016)). Right hand section: based on data. The minimum sampling frequency required to keep risk of non-detection less than 1% is indicated by red = weekly, yellow = fortnightly, white = monthly. Horizontal lines divide the groups in sets of five, to guide the eye.

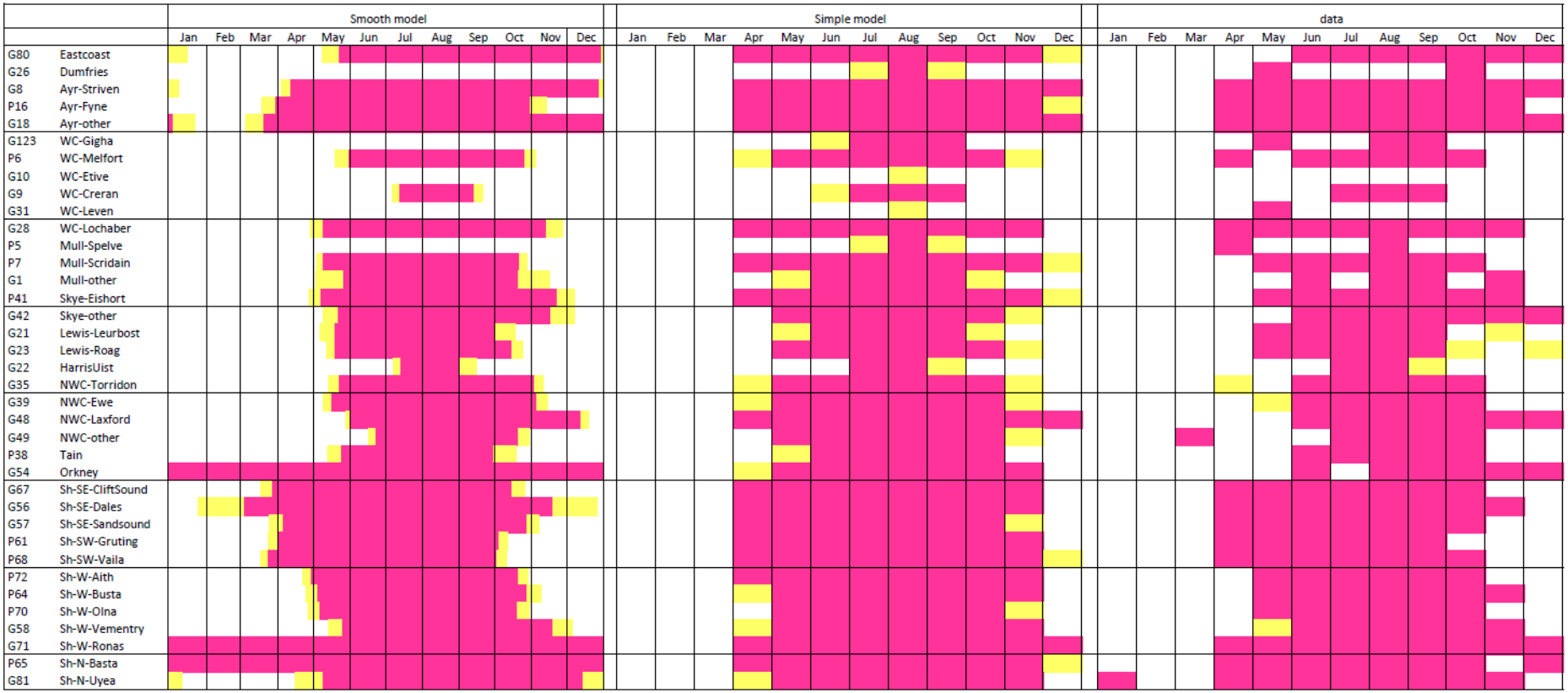


Table 2: Sampling frequency required to keep the risk of non-detection of PST in mussels exceeding 400 µg/kg below 1% for an average year. Left hand side: based on smooth models. Middle section: based on simple models (as presented in Holtrop et al. (2016)). Right hand section: based on data. The minimum sampling frequency required to keep risk of non-detection less than 1% is indicated by red = weekly, yellow = fortnightly, white = monthly.

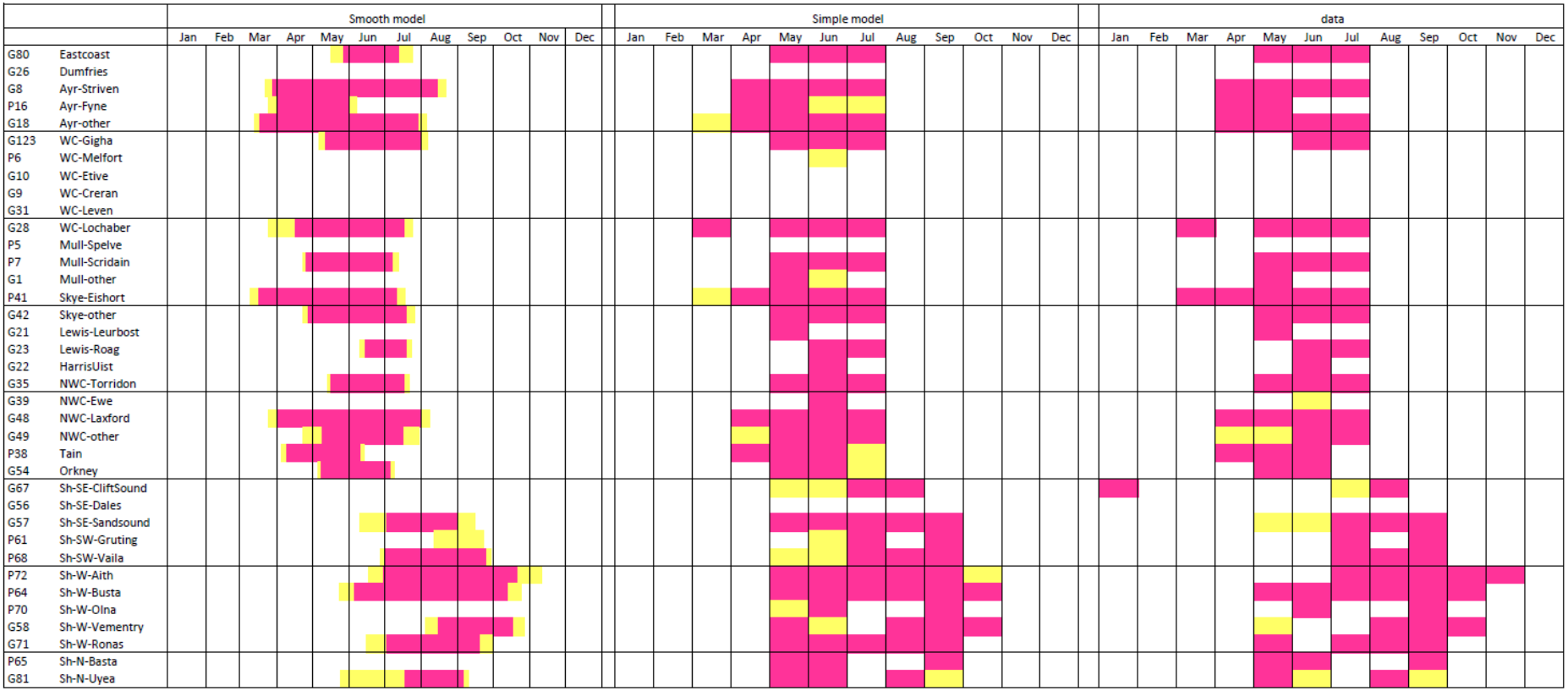


Table 3: Sampling frequency required to keep the risk of non-detection of DA in mussels exceeding 5 mg/kg below 1% for an average year. Left hand side: based on smooth models. Middle section: based on simple models (as presented in Holtrop et al. (2016)). Right hand section: based on data. The minimum sampling frequency required to keep risk of non-detection less than 1% is indicated by red = weekly, yellow = fortnightly, white = monthly.

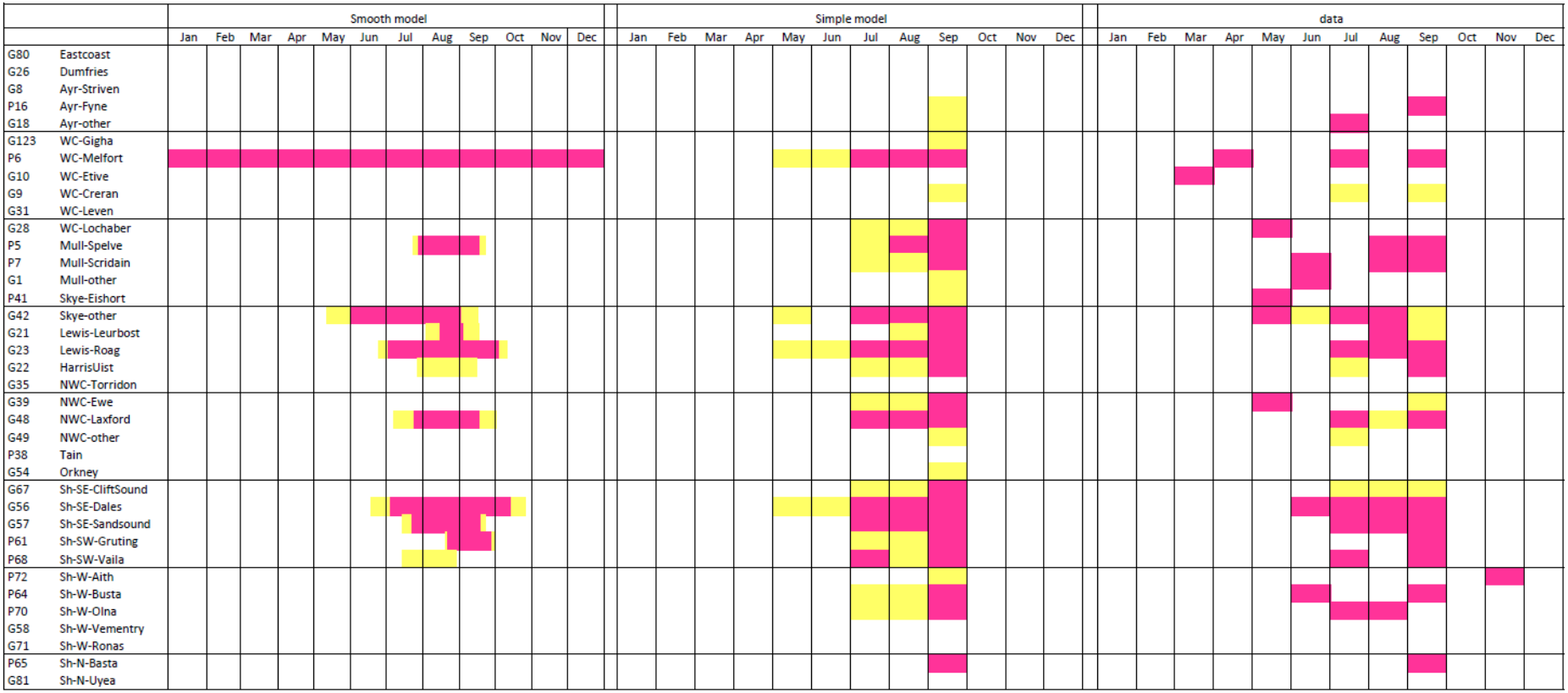


Table 4: Sampling frequency required to keep the risk of non-detection of LT in mussels exceeding 0.5 MPL below 1% for an average year. Left hand side: based on smooth models based on LT data from 2011-15. Right hand section: based on LT data from 2011-15. The minimum sampling frequency required to keep risk of non-detection less than 1% is indicated by red = weekly, yellow = fortnightly, white = monthly.

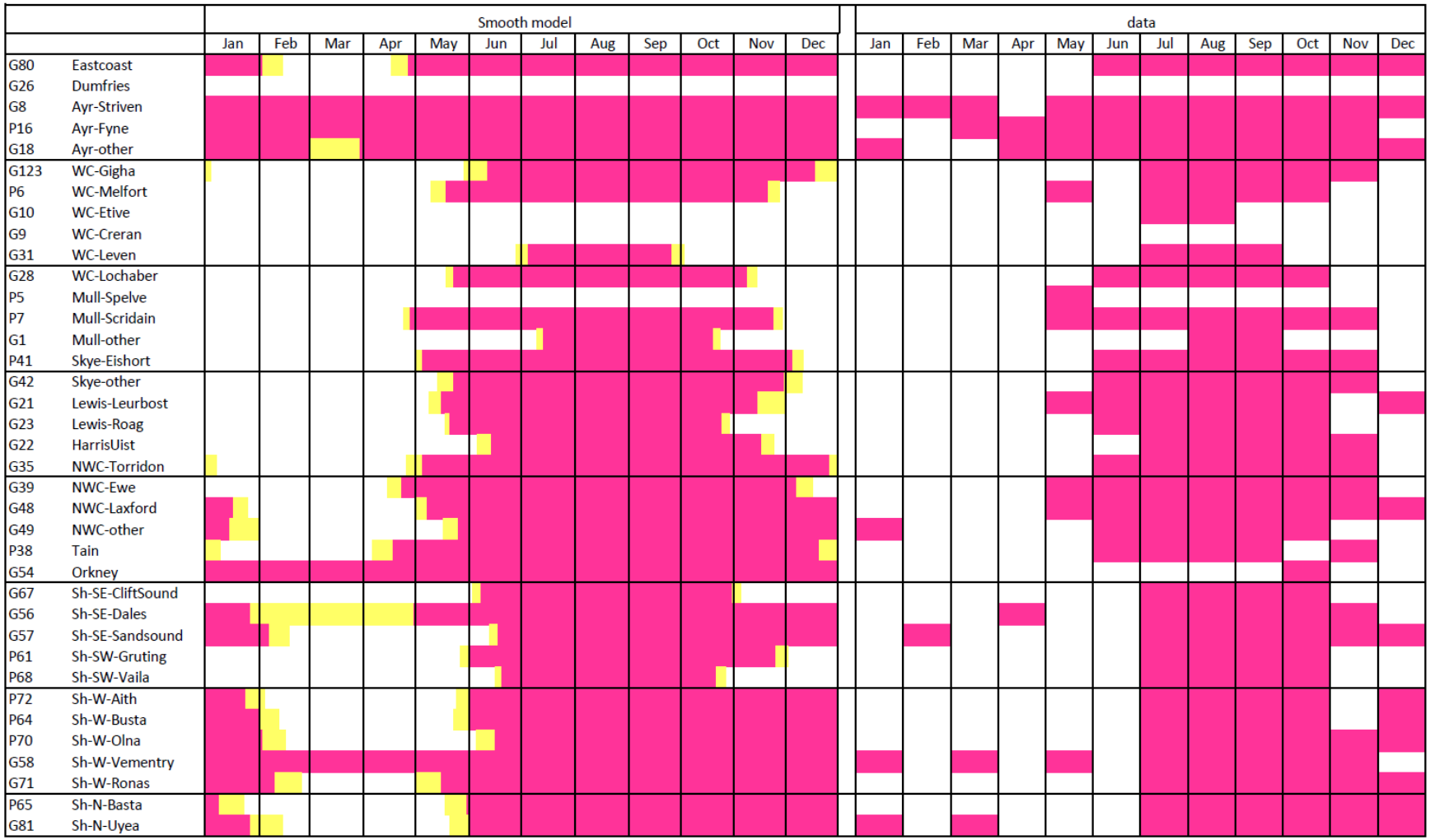


Table 1 shows the results based on LT > MPL. Note that for the smooth model we are no longer limited to having a fixed monitoring frequency throughout the entire month (although in practice we may want to adhere to fixed frequencies from month to month). Due to the nature of the smooth model, suggested frequencies always progress smoothly from monthly to fortnightly to weekly sampling and vice versa. For several Groups the suggested monitoring frequency is fortnightly or weekly in December and this continues into January. The suggested frequencies for G54 Orkney is weekly all year round, but this is a consequence of the smooth model giving a poor fit for this group (see also plot 25 in Figure C1[4]).

For PST >400 µg/kg (Table 2), the smooth model generally shows good agreement with the data. On the whole, both the smooth and simple models suggest similar sampling frequencies, although the smooth model tends to suggest slightly shorter time windows during which more frequent sampling is necessary.

The results based on DA > 5 mg/kg are shown in Table 3. Agreement between the data, simple model and smooth model is a bit patchy. For example, the smooth model suggests weekly monitoring for Group P6 WC-LochMelfort, which is not supported by data (plot 7 in Figure C3). On the other hand, the simple model tends to prefer increased monitoring during September which is not always supported by the data for the majority of groups. For G67 Sheltand-SE-Cliftsound and P65 Shetland-N-Basta the smooth model suggests monthly sampling all year round. The data (plots 26 and 36 in Figure C3) however show some positive samples albeit widely spread throughout the year. The simple model appears to be better able to capture the toxin patterns here.

Table 4 shows proposed frequencies based on LT-LCMS samples from 2011-15 exceeding half the MPL. Simple models could not be fitted to these data and therefore only results based on the smooth model are shown. As with the LT data from 2001-15 exceeding the MPL, for several groups continuation of increased monitoring is suggested for January. The smooth model appropriately suggests weekly monitoring for G8 Ayr-LochStriven, P16 Ayr-LochFyneArdkinglas, G18 Ayr-other and G58 Shetland-W-VementryVoe throughout the year and this reflects how samples exceeding 0.5 MPL tend to occur throughout the year (see Figure C4, plots 3, 4, 5 and 34). The suggested weekly monitoring all year round for G54 Orkney does not agree with the data however, which showed only one sample exceeding 0.5 MPL (Figure C4 plot 25).

There are also occasions where the simple model falls short. Figure 4 demonstrates how the simple model ignores the timing of samples exceeding the limit of interest at the beginning or the end of the month. Prevalence of samples for which PST > 0.5 MPL started early April and lasted until the very end of July. The simple model however assumed negligible prevalence in the months preceding this sequence and suggests monthly monitoring in March and switches to weekly monitoring in April (Table 2). At the end of the sequence of toxic events the reverse is happening. Here the simple model suggests monthly monitoring in August (Table 2) ignoring that toxic events occurred right until the end of July. The smooth model, on the other hand, suggests a more realistic monitoring scheme with intensive sampling starting halfway through March and continuing into the middle of August. A similar example is shown in Figure 3d, while Figure 3f shows the presence of samples exceeding the limit of interest until the very end of September / beginning of October, with the smooth model suggesting increased monitoring to continue into October (Table 3). The simple

[4]Figure numbers beginning with 'C' are presented in Appendix C.

model simply switches weekly sampling in September to monthly sampling October (Table 3), ignoring the presence of toxic samples until nearly the beginning of October.

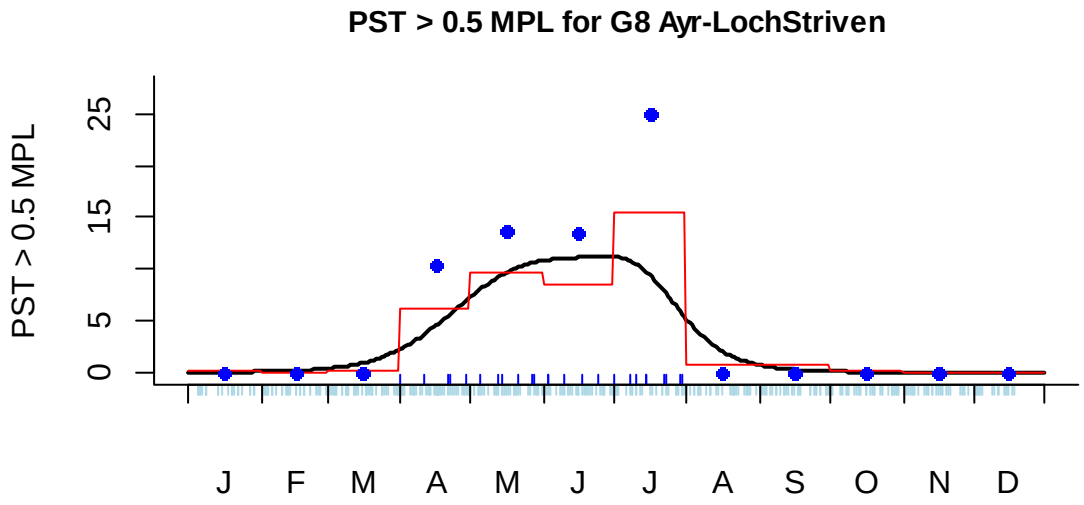


Figure 4: Example of comparing model fit from smooth model against fit obtained from simple model, where simple model ignores that samples were obtained early or late in the month. The prevalence is shown, based on data (blue circles), predicted prevalence for an average toxin year from smooth models (black curve), and from simple models (red lines, based on Holtrop et al. (2016)). The x-axis shows the days of the year with ticks indicating when samples were obtained and whether their test results was below (light blue) or above (blue) the limit of interest.

3.4 Integrated display of model and data

From the above it is clear that both the simple and the smooth models occasionally miss the mark. For example, when monitored for long enough, sooner or later positive samples will occur at unusual times. How much importance we give to these samples depends on the context. A sample taken long time ago that was analysed with outdated methods may be regarded less relevant. Also important to consider is not only did the sample exceed half the MPL (or whatever limit is used for the risk assessment and developing alternative monitoring schemes), but did it exceed the actual MPL. Likewise, were there several samples during that particular time window that tested positive (but not exceeding our limit of interest). Such information might strengthen support more frequent sampling. In an attempt to address these issues we have developed a graphical display that shows the data and model predictions such that at least some of the above concerns are highlighted.

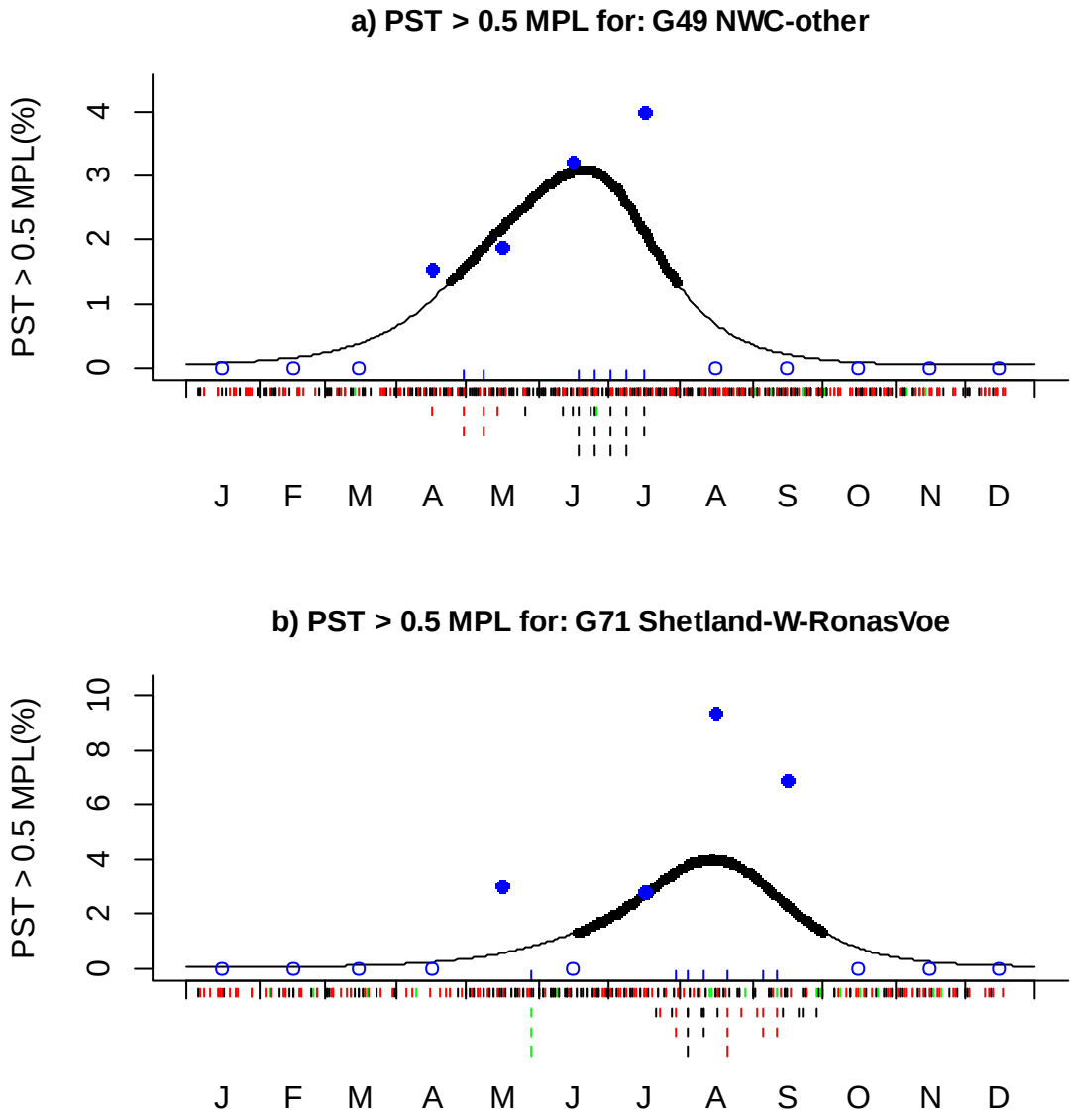


Figure 5: For two groups of pods the prevalence of PST >0.5 MPL is shown, based on data (blue circles) from 2001-15, and predicted prevalence for an average toxin year from smooth models (black curve). Bold lines and closed symbols indicate where monthly sampling is insufficient (prevalence > cut-off and fortnightly or weekly sampling would be required to keep the risk of non-detection below 1%). The x-axis shows the days of the year. Ticks above the x-axis are samples for which PST > 400 µg/kg. Ticks below the x-axis indicate the following: first row of ticks: samples for which PST = 0 µg/kg, second row of ticks: 0 < PST < 400 µg/kg, the third row of ticks: 400 ≤ PST < 800 µg/kg, and the fourth row of ticks: PST ≥ 800 µg/kg. The colouring of these ticks is as follows: green 2001-5, red 2006-10, black 2011-15.

Figure 5 shows examples for two locations for PST exceeding half the MPL. The graph shows the data as % exceeding 0.5 MPL, by month. When the data suggest that monthly sampling is insufficient, then this is indicated with a closed circle (i.e. observed prevalence > 1.33%). The black curve shows the predicted prevalence of PST exceeding 0.5 MPL, based on fitting smooth models to the data. Where the predicted prevalence exceeds 1.33%, i.e. monthly monitoring would be insufficient, this is indicated by using a bold line segment. The blue ticks above the x-axis indicate occurrences where the test result exceeded 0.5 MPL. The rows of ticks below the x-axis reflect the following: The first

row of ticks shows samples that tested zero (below the limit of detection). The second row of ticks shows samples that tested positive (i.e. > 0 µg/kg). The third row of ticks shows samples for which the test result exceeded 400 µg/kg, and the fourth row of ticks indicates samples exceeding the MPL of 800 µg/kg. Note that a sample that exceeded 800 µg/kg will show a tick in rows 2, 3 and 4, whereas a sample testing between 400 and 800 µg/kg will show a tick in rows 2 and 3. To indicate the year span during which the sample was observed, three different colours are used for the ticks. Green, red and black refer to samples taken during 2001-5, 2006-10 and 2011-15, respectively.

For group G49 NWC-other (Figure 5a) we see that 2 samples exceeded 400 µg/kg early in summer (late April / early May), and these were taken during 2006-10 (shown in red). There is another cluster of samples that exceeded 400 µg/kg in June and July, taken during 2011-15 (shown in black). For this second cluster, as well as that the test results were observed during more recent years, several samples actually exceeded 800 µg/kg (as indicated by ticks in the fourth row below the x-axis), suggesting that we should take these observations seriously. For Group G71 Shetland-W-RonasVoe (Figure 5b) we see how one test result exceed 0.5 MPL in May, and that in actual fact it exceeded 800 µg/kg (tick in fourth row below x-axis). However, this sample was taken during 2001-5 and so perhaps should not be given too much importance.

The above approach has been worked out in detail for all groups in Figures D1[5] (LT), D2 (PST), D3 (DA) and D4 (LT LCMS). For the latter, as data cover more recent years only, a year range has not been indicated. Instead, the three types of toxin (OA, AZA, YTX) have been highlighted.

4 Discussion

We investigated the use of smooth models that have a more detailed time scale than the simple models used previously and that, unlike the simple models, allow for smooth progression of biotoxin prevalence patterns over time. Generalised additive models were chosen because they allow for smooth description of the data without imposing restrictions on the shape of the curve, that is, the data drive the shape of the curve. These models were successfully fitted to the mussel test results from 2001-15 and gave predicted prevalences that change smoothly from day to day throughout the year.

The main advantage of the smooth models is that they provide a detailed day to day summary of biotoxin prevalences, with a smooth progression over time. It is obvious that, compared to the simple models used previously (where toxin levels were assumed constant throughout the month and change to a new level overnight at the start of the next month), this provides a more accurate indication of when the monitoring scheme should switch to a higher sampling frequency.

The two main disadvantages to fitting smooth models are that they are time consuming to fit and that the optimisation routine sometimes runs into problems when the model specification is not 'quite right'. The latter meant that if a term was included in the model but the data did not fully support this term, then the optimisation routine has a tendency to crash. Memory problems were also experienced when certain models were fitted (using a standard PC). With regards to the time taken to fit a model, it took several hours to fit a smooth model to one mussel biotoxin data set,

[5] Figure numbers beginning with 'D' are presented in Appendix D.

whereas the simple models only took a couple of minutes to run. This is not a problem if the most suitable model formulation that describes the data best is known in advance. In practice however, often several model formulations need to be explored and tested before settling on a model that is most suitable for the data at hand. When combined with the various biotoxins of interest, each at various cut-off levels, for several shellfish species, then the number of models to be fitted rapidly multiplies and running time does become important.

Comparing the fits from simple models to those from the smooth models for the mussel data, we found that generally the simple model captured the main trends quite well. In general, toxin levels observed in the data change sufficiently slowly that the simple month-based models quite adequately capture the main trends.

For both the simple and the smooth modelling approaches it was found that occasionally the observed biotoxin trends in the data were not captured well by the model. This tended to happen in particular for groups of pods where only a few or no toxic events had been observed. We therefore propose a visual display that not only shows the model predictions and observed prevalence of a given biotoxin level (at half the MPL, say), but in addition also gives an indication of the actual level observed (below limit of detection, between limit of detection and 0.5 MPL, between 0.5 MPL and MPL, exceeding MPL), and when it was observed (between 2001-5, 2006-10, 2011-15). This allows for a comprehensive integration of all the biotoxin information available when developing monitoring schemes.

Changes in test results over the years

There are several possible reasons why prevalence varies over years. In addition to changes in biotoxin profiles and prevalence due to natural causes such as climate change, other factors may also play a role:

- Biotoxin test methodologies have changed, such as moving from the mouse bioassay to HPLC and LCMS techniques, and this may have resulted in changes in reported biotoxin levels or its composition over the years.
- The organisation and running of the monitoring programme has become much more streamlined in the last 10 years or so, making these more recent data more consistent than data from the early years of monitoring.
- The average biotoxin profile of a Pod may change over time with some shellfish farms being taken out of production and others being added.

Despite the above pointing towards perhaps giving test results from earlier years less importance, they nevertheless act as a reminder that potentially biotoxin levels can increase at any time of the year.

Limitations of monitoring schemes

As explained in detail in Holtrop et al. (2016), monitoring schemes will always be limited due to, among others, the assumption that the test result from a small shellfish sample is representative of

the entire shellfish field in question, and the assumption that biotoxin levels change sufficiently slowly such that test results are valid for one week, implying that weekly monitoring is safe.

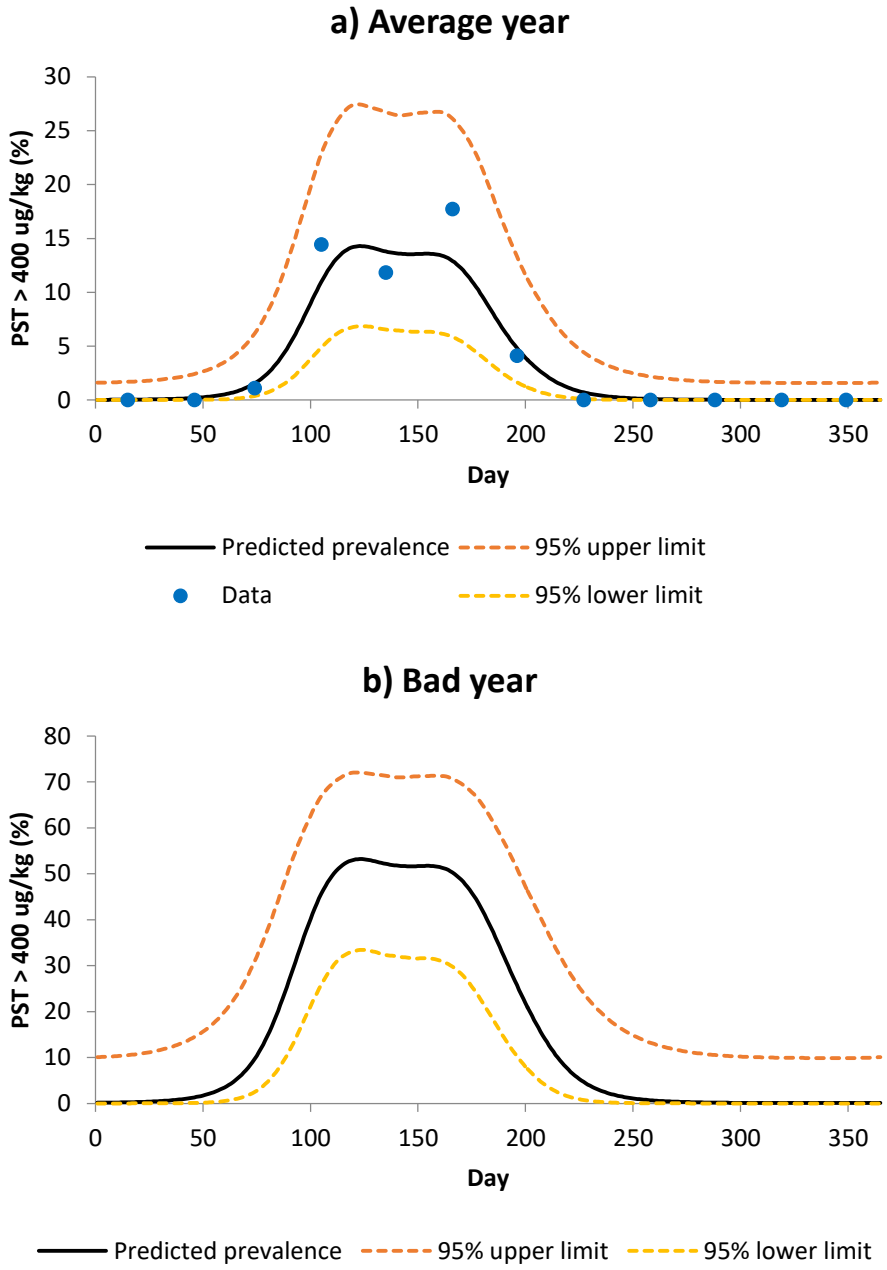


Figure 6: Predicted prevalence of PST > 400 µg/kg for an average year (top panel), for G18 Ayr-other. Also shown is the 95% confidence interval for the predicted prevalence. The bottom panel shows the predicted prevalence and its 95% confidence interval for a bad toxin year (that belongs to the top 5% of years with high toxin prevalence).

Another limitation is the large uncertainty in the model predictions. This is illustrated in Figure 6a, which shows the predicted prevalence of PST > 0.5 MPL for a group of pods that show an 'average' prevalence pattern in the data. The predicted prevalence for an average biotoxin year is shown, as well as the limits of its corresponding 95% confidence interval. Its lower limit exceeds 1.33% (when

monthly sampling would no longer be regarded sufficient) from day 87 through to day 200, so that we can be confident that the true (but unknown) prevalence is sufficiently high to warrant frequent monitoring between the end of March and the middle of July. The upper limit of the 95% confidence interval exceeds 2% even when both model and data show zero prevalence, and if this were to be taken as our guide for developing monitoring schemes then weekly sampling all year round would be needed. This would be the case (data not shown) for nearly all groups for all three toxin levels of interest (LT > MPL, PST > 0.5 MPL, DA > 5 mg/kg) and is simply not practical. Taking a 'worst case' biotoxin year into consideration only makes matters worse, with the upper limit of the confidence interval even higher (Figure 6b). The magnitude of the uncertainty is similar for both the smooth and simple models (data not shown), and reflects binomial variation and is a consequence of having only a relatively small number of samples for each month. To illustrate, to ensure that the upper limit of the 95% confidence interval is less than 1.33%, i.e. monthly monitoring would be safe, at least 225 samples per month would be needed, all having a negative test result.

Should Group be considered as a random effect or as a fixed effect?

This is a question that is important from a statistical point of view. Generally, if the focus of interest is on one or more individual groups, it would be appropriate to regard group as a fixed effect, whereas if were interested in predicting biotoxin patterns for a future but hitherto unknown group, then treating group as a random effect would be more appropriate. The practical implications of group being treated as fixed or random effect are most noticeable for those groups for which all (or nearly all) test results were below the limit of interest (such as 0.5 MPL). When group is regarded as a fixed effect the predicted prevalence will be zero, whereas when it is regarded as a random effect the predicted prevalence will be close to zero. The average biotoxin prevalence across all groups will somewhat influence the predicted prevalence for the group of interest. When the first risk assessment was conducted in 2004, biotoxin data were limited and little was known about toxin patterns. It was felt that when all samples tested negative for a given group of pods, not too much importance should be given to this finding (as it was based on relatively small number of test results) and therefore group was incorporated as a random effect.

The smooth models employed here regard group a fixed effect, for various reasons:

- When group was incorporated as a random effect, excessive 'shrinkage towards the overall mean' occurred for the LT LCMS data, resulting in predicted biotoxin patterns that were well above the prevalence observed in the data when the observed prevalence was low, and that were well below the prevalence observed in the data when the observed prevalence was high. This seemed unrealistic as the data test results were consistently below the limit of interest during the first four months of the year (see Figure C4, observed prevalence of 0 during Jan-Apr for the majority of groups), but the predicted prevalence was 5% or higher.
- In some cases the software ran into memory problems.
- Each individual group is of explicit interest.

What are the practical implications? Based on the results from the simple model (which regards group as random effect), it can be seen that for groups that have zero prevalence data throughout the predicted prevalence is slightly higher than zero (but less than 1.33%, the cut-off for which more

frequent sampling would be required) during the summer months. See e.g. Figure 3e (but also plots 2, 7, 8, 9, 10, and 27 in Figure C2, and plots 1, 2, 3, 6, 10, 20, 24, 25, 34, 35 and 37 in Figure C3), where the simple model shows a predicted prevalence that is not quite zero (despite all test results being below the limit of interest). The practical consequence is the same however, irrespective of whether the prevalence is estimated to be zero or to be slightly higher than zero; in both cases monthly sampling would be regarded sufficiently safe.

Summary and conclusion

In summary:

- Smooth models provide a more realistic description of progression of biotoxin prevalence than simple models used previously.
- A major drawback of fitting smooth models is that it is time-consuming process.
- In broad terms there was good agreement between simple models and smooth models, and good agreement of both the simple and smooth models with prevalence patterns observed in the data.
- Simple models appear more robust, i.e. are less prone to giving unrealistic predictions for 'difficult' data.
- The simple model failed to fit a model to the LT LCMS data from 2011-15, whereas the smooth model succeeded.
- Both smooth models and simple models occasionally fail to capture toxin patterns when very few samples exceeded the limit of interest.
- Relatively small numbers of samples mean that, although the model predictions give an indication of biotoxin prevalence patterns, their predicted values are variable with high upper limits, caused by uncertainty due to small numbers of samples.
- There is no single best model, and models should only ever be regarded as a tool to aid the development of monitoring schemes.
- There are various limitations to data collection (such as small sample representative of entire harvesting field), assumptions in risk assessment and monitoring scheme development (such as the test result being valid for a week), and uncertainty in the predicted prevalences (unless huge numbers of samples are obtained).

The implications are that

- Approaches based on the simple models employed previously can continued to be used in future risk assessments.
- For a select number of biotoxin levels and shellfish species it is worthwhile exploring smooth approaches.
- Consideration of smooth models is also worthwhile when simple models can not be fitted successfully.
- Predicted prevalences, risk assessments and suggested monitoring frequencies should always be regarded cautiously, due to the aforementioned limitations.

- Where available, information from other sources should be incorporated to adapt monitoring frequencies throughout the year, based on as many sources of information as possible, such as phytoplankton and biotoxin test results from current and preceding weeks from the same and neighbouring pods.

Acknowledgements

FSS is thanked for funding this research. Rob Fryer (Marine Scotland) is thanked for helpful comments on implementation of the GAMM methodology in R. Graham Horgan (BioSS) is thanked for critical reading of the draft report.

References

Holtrop, G., Swan, S., Duff, B., Wilding, T, Naryanaswamy, B. & Davidson, K. (2016) Risk assessment of the Scottish monitoring programme for marine biotoxins in shellfish harvested from classified production areas: review of the current sampling scheme to develop an improved programme based on evidence of risk. Report to Food Standards Scotland, Project code FSS/2015/021. September 2016.

Holtrop, G. (2008) Risk assessment of the FSA Scotland inshore shellfish monitoring programme based on historical toxin data from 2004-2006. Report to Food Standards Agency Scotland, Project code S14036. February 2008.

Holtrop, G., & Horgan, G.W. (2004) Risk assessment of the FSA Scotland monitoring programme for biotoxins in shellfish harvested from classified inshore areas in Scotland: evaluation of the current scheme and development of improved alternatives based on historical data. Report to Food Standards Agency Scotland, Project code S01026. December 2004.

Lee, Y., & Nelder, J.A. (1996) Hierarchical generalized linear models. Journal of the Royal Statistical Society, series B 58: 619-678.

Lee, Y., & Nelder, J.A. (2001) Hierarchical generalized linear models: a synthesis of generalised linear models, random-effect models and structured dispersions. Biometrika 88: 987-1006.

Wood, S.N. (2006) Generalized Additive Models: an Introduction with R, Chapman & Hall, London.

Appendix A: Model details

Each test result was summarised as 0 or 1, corresponding to the absence or presence of a toxic event (i.e. exceedance of limit of interest). A cubic spline was fitted to these data using the gamm4() routine in R, as follows

```
y~s(Day,by=Group,bs="cc",k=25)+Group-1,random=~(1|Year),family=binomial(link="logit")
```
(A1)

where Day refers to day of year (1 through to 365, extra day in leap years ignored), Group is a factor identifying the (group of) pod(s) the sample came from. Each group of pods was allowed its own specific smooth spline, where each spline was assumed cyclic to ensure continuity between 31 December and 1 January. To allow for random variation from year to year, Year was included as a random effect. Predictions for each group for an average year extracted with the predict() function applied to the gam component of the output.

The basis dimension for each smoothed curve was set to k=25, corresponding to 23 evenly spaced interior knots, i.e. two knots for each month. The effective degrees of freedom (EDF) of the final models were al 5.9 or less, and as this is well below the number of knots this indicates that the model space was not limiting in any way.

Akaike's Information Criterion (AIC), which is based on maximum log likelihood and the number of parameters to be estimated, was used to compare goodness of fit. In order to compare like with like, the simple model used in Holtrop et al. 2016 was refitted with gamm4(). The maximum likelihood (including constant terms) is reported below:

Toxin level	# samples	Max loglik simple	Max loglik smooth	# pars simple[1]	EDF smooth[2]	# pars smooth[3]	AIC[4] simple	AIC smooth
PST>800 µg/kg	19018	-588.7	-718.5	445	39.85	77.85	2067	1593
PST>400 µg/kg	19018	-1135.9	-1309.3	445	64.41	102.41	3162	2823
PST>0 µg/kg	19018	-1913.2	-2103.7	445	92.44	130.44	4716	4468
DA>5 mg/kg	14557	-449.2	-550.3	445	23.49	61.49	1788	1224
DA>0 mg/kg	14557	-1956.6	-2178.6	445	84.89	122.89	4803	4603
LT>MPL	21592	-4118.9	-4394.1	445	110.06	148.06	9128	9084
LT LCMS>0.5MPL	8857	NA[5]	-2669.3	445	115.51	153.51	NA	5646

[1]Number of parameters in the simple model: 37 groups with 12 months per group, and a variance component for year.

[2]EDF: estimated degrees of freedom (EDF) for the smooth terms, presented here as the sum of the EDF for each of the 37 groups.

[3]Number of parameters in the smooth model: given as EDF plus one constant per group plus a variance component for year.

[4]AIC: Akaike's Information Criterion. Models with lower AIC are preferred.

[5]Simple model could not be fitted to these data.

The simple model, despite its name, uses more parameters than the smooth model, and although it gives a better maximum likelihood, the extra cost due to more parameters does not outweigh the benefits, as is shown by higher Akaike's Information Criterion (AIC) for the simple models.

Before settling on the final model as shown in equation (A1), several model structures were explored, among which were:

- Models that regard group as a random effect. This gave good predictions for the PST> 0.5 MPL and DA> 5 mg/kg data, albeit that the between group variation component was estimated to be zero. For the LT > MPL data the program ran into memory problems (on a standard PC). For the more recent LT LCMS > 0.5 MPL data the predictions showed a large amount of shrinkage, with large prevalence in the data strongly suppressed and largely elevated predictions when the data showed zero prevalence. This resulted in suggested weekly monitoring all year round and seemed unrealistic, as the fast majority of the groups showed zero prevalence for the first four months of the year. Furthermore, as individual groups are specifically of interest to us it is counterintuitive to regard group as random effect.
- In an attempt to emulate some shrinkage towards the overall mean for locations that consistently show negative test results, a Scotland-wide smooth curve was incorporated as well as group-specific curves. This resulted in somewhat unrealistic predicted biotoxin patterns. For example, for DA > 5 mg/kg, the predictions for each of the individual groups showed three distinctive peaks, which seemed unrealistic. This was a consequence of biotoxin prevalence peaking at different times of the year, depending on the location. It is known that prevalence rises early in summer along the west coast of Scotland, it peaks in the middle of summer in the north-west, while in Shetland biotoxin prevalence only starts to increase later in summer. As a consequence, the overall mean pattern shows three distinctive peaks. We know, however, that individual groups generally do not show all these peaks, so this approach was abandoned.

Ultimately model A1 was chosen as the final model as it could be fitted to all toxin levels of interest without crashing, memory problems or convergence issues, and which gave reasonable predictions.

The estimated between year variances (on the logit scale) for the simple and smooth models are similar:

Toxin level	Simple model variance(Year)	Smooth model variance(Year)
PST>800 µg/kg	2.30	2.16
PST>400 µg/kg	1.37	1.36
PST>0 µg/kg	1.79	1.82
DA>5 mg/kg	0.76	0.70
DA>0 mg/kg	0.12	0.11
LT>MPL	0.45	0.45
LT LCMS>0.5MPL	NA[1]	0.19

[1]Simple model could not be fitted to these data.

Appendix B: Data summaries by group

Table B1: Summary of biotoxin test results in mussels from April 2001 to September 2015, summarised by group.

		DA (mg/kg)			PST (µg/kg)				LT		LT LCMS			
Group	Groupname	n	DA>0	DA>5	n	PST>0	PST>400	PST>800	n	LT>MPL	n	LT>0	LT>0.5MPL	LT> MPL
G80	Eastcoast	173	3	0	173	9	4	1	201	23	58	34	24	15
G26	Dumfries	286	1	0	319	0	0	0	411	2	186	6	0	0
G8	Ayr-LochStriven	331	4	0	412	35	25	20	514	113	191	132	89	59
P16	Ayr-LochFyneArdkinglas	277	1	1	355	10	5	4	444	51	172	98	39	18
G18	Ayr-other	643	9	2	906	57	43	29	1126	225	416	264	149	113
G123	WC-Gigha	218	10	0	310	14	7	4	386	4	199	64	10	1
P6	WC-LochMelfort	243	33	5	322	6	0	0	379	26	189	99	47	22
G10	WC-LochEtive	516	10	1	607	0	0	0	680	2	217	47	2	0
G9	WC-LochCreranLynnhe	499	15	2	620	0	0	0	722	8	204	58	0	0
G31	WC-LochLevenEil	524	1	0	612	0	0	0	747	4	299	37	7	1
G28	WC-Lochaber	623	23	6	812	43	21	9	896	99	352	178	98	59
P5	Mull-LochSpelve	320	24	5	333	0	0	0	437	3	171	32	1	0
P7	Mull-LochScridain	308	15	3	421	44	29	7	453	57	172	103	68	41
G1	Mull-other	384	9	1	483	4	2	0	478	9	74	7	5	3
P41	Skye-LochEishort	332	4	1	454	43	27	13	489	52	184	109	66	32
G42	Skye-other	587	53	10	797	31	18	5	802	37	241	75	26	8
G21	Lewis-LochLeurbostErisort	449	22	3	529	3	1	0	713	23	379	122	47	13
G23	Lewis-LochRoag	864	65	22	1038	33	11	7	1267	111	564	234	133	73
G22	HarrisUist	686	32	5	788	12	2	2	947	12	564	146	33	6
G35	NWC-LochTorridon	502	23	0	684	37	24	16	730	45	251	95	50	31
G39	NWC-LochEweBroom	460	24	3	620	5	1	0	660	40	236	93	44	26
G48	NWC-LochLaxfordInchard	433	21	6	628	61	34	19	723	124	354	201	132	87
G49	NWC-other	416	8	1	566	14	7	4	566	32	179	71	38	21
P38	Tain	258	13	0	367	21	13	6	372	11	142	55	20	7
G54	Orkney	99	1	0	126	12	9	6	133	13	68	10	1	0
G67	Shetland-SE-CliftSound	425	32	3	581	24	5	2	656	55	295	99	42	22
G56	Shetland-SE-DalesVoe	309	22	8	367	3	0	0	439	31	205	72	26	16
G57	Shetland-SE-SandsoundWeisdale	540	45	7	716	22	10	4	809	51	363	175	75	31
P61	Shetland-SW-GrutingVoe	331	17	6	399	12	4	0	472	42	173	78	30	11
P68	Shetland-SW-VailaVoe	347	15	4	455	39	21	5	484	51	180	69	23	12
P72	Shetland-W-AithVoe	241	9	1	376	25	9	2	412	40	186	75	25	13
P64	Shetland-W-BustaVoe	296	15	3	457	34	16	7	466	36	191	80	24	12
P70	Shetland-W-OlnaFirth	268	9	2	412	13	5	2	420	23	191	58	21	11
G58	Shetland-W-VementryVoe	332	5	0	518	36	12	4	545	38	266	119	50	27
G71	Shetland-W-RonasVoe	213	5	0	288	20	7	3	301	52	122	70	38	27
P65	Shetland-N-Basta	244	9	1	382	11	4	0	412	31	186	57	20	14
G81	Shetland-N-Uyea	580	8	0	785	49	20	6	900	58	437	163	74	44
	total	14557	615	112	19018	782	396	187	21592	1634	8857	3485	1577	876

Appendix C: Smooth model fits for main biotoxin levels of interest (LT > MPL, PST > 0.5 MPL, DA > 5 mg /kg, LT LCMS > 0.5 MPL)

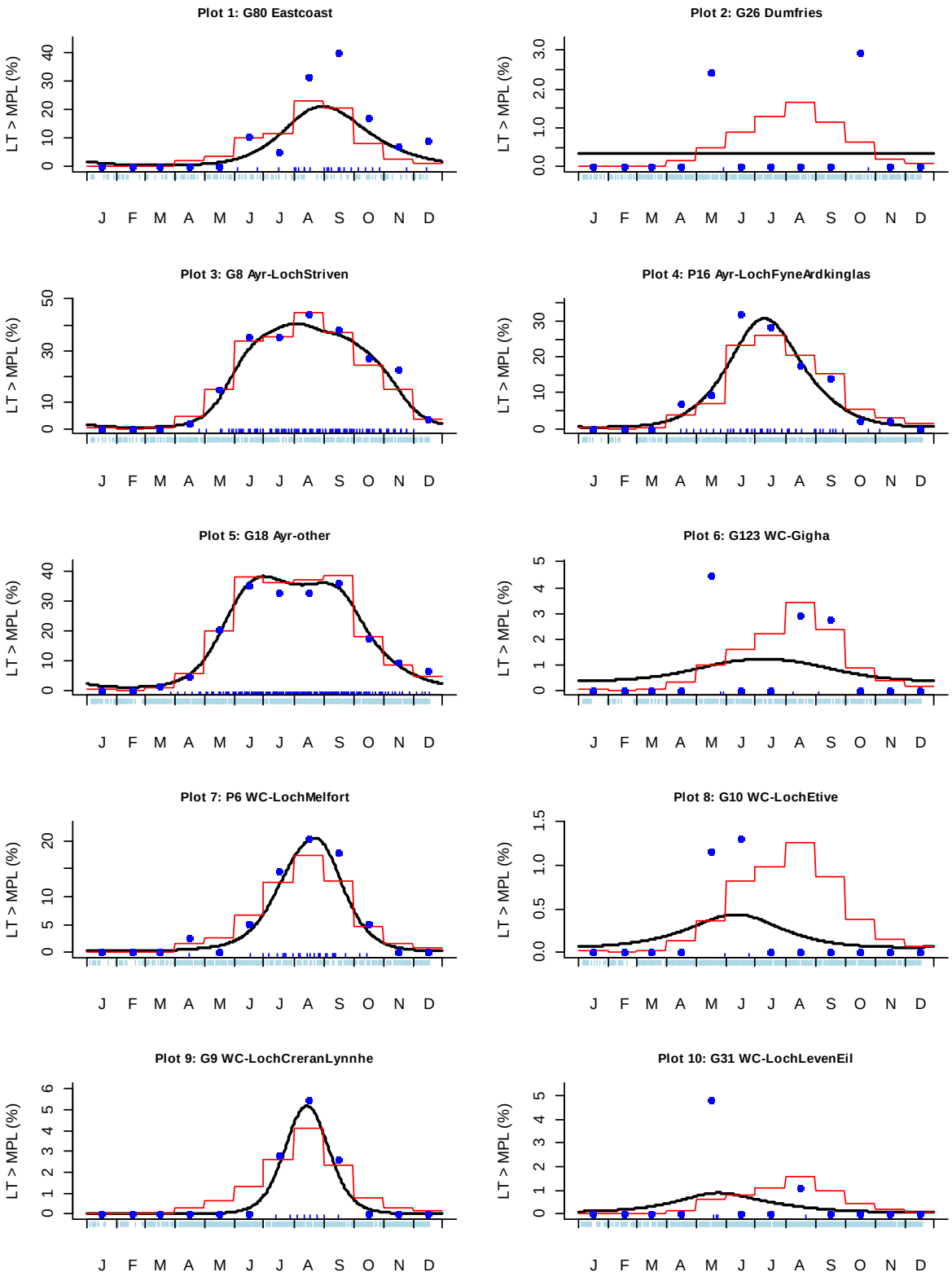


Figure C1: For each (group of) pod(s) the prevalence of LT > MPL is shown, based on data (blue circles) from 2001-15, from simple models where predicted prevalence for an average toxin year is constant for each month (red lines, based on Holtrop et al. 2016), and predicted prevalence for an average toxin year from smooth models (black curve). The x-axis shows the days of the year with ticks indicating when samples were obtained and whether their test results was below (light blue) or above (blue) MPL.

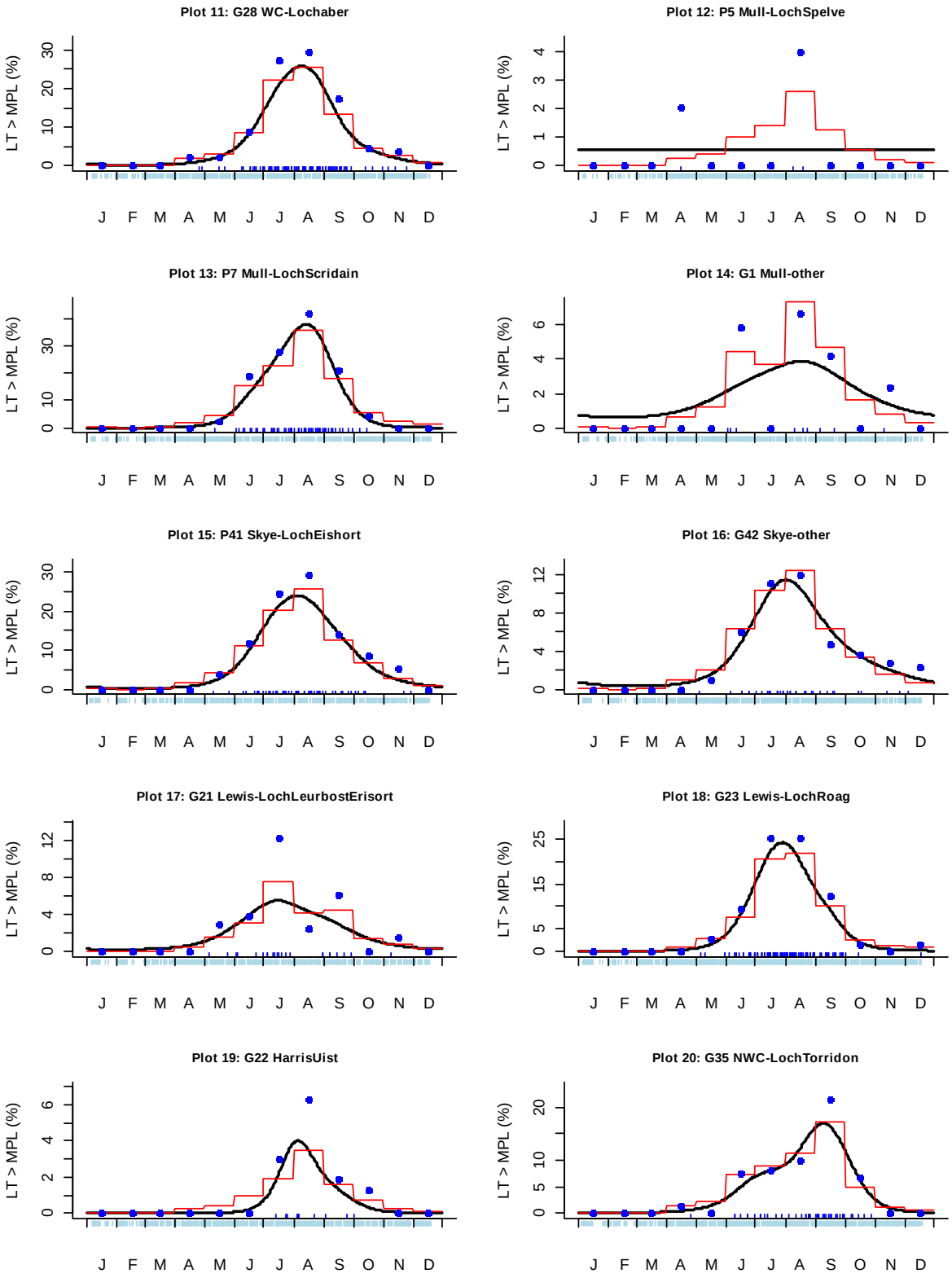


Figure C1 continued

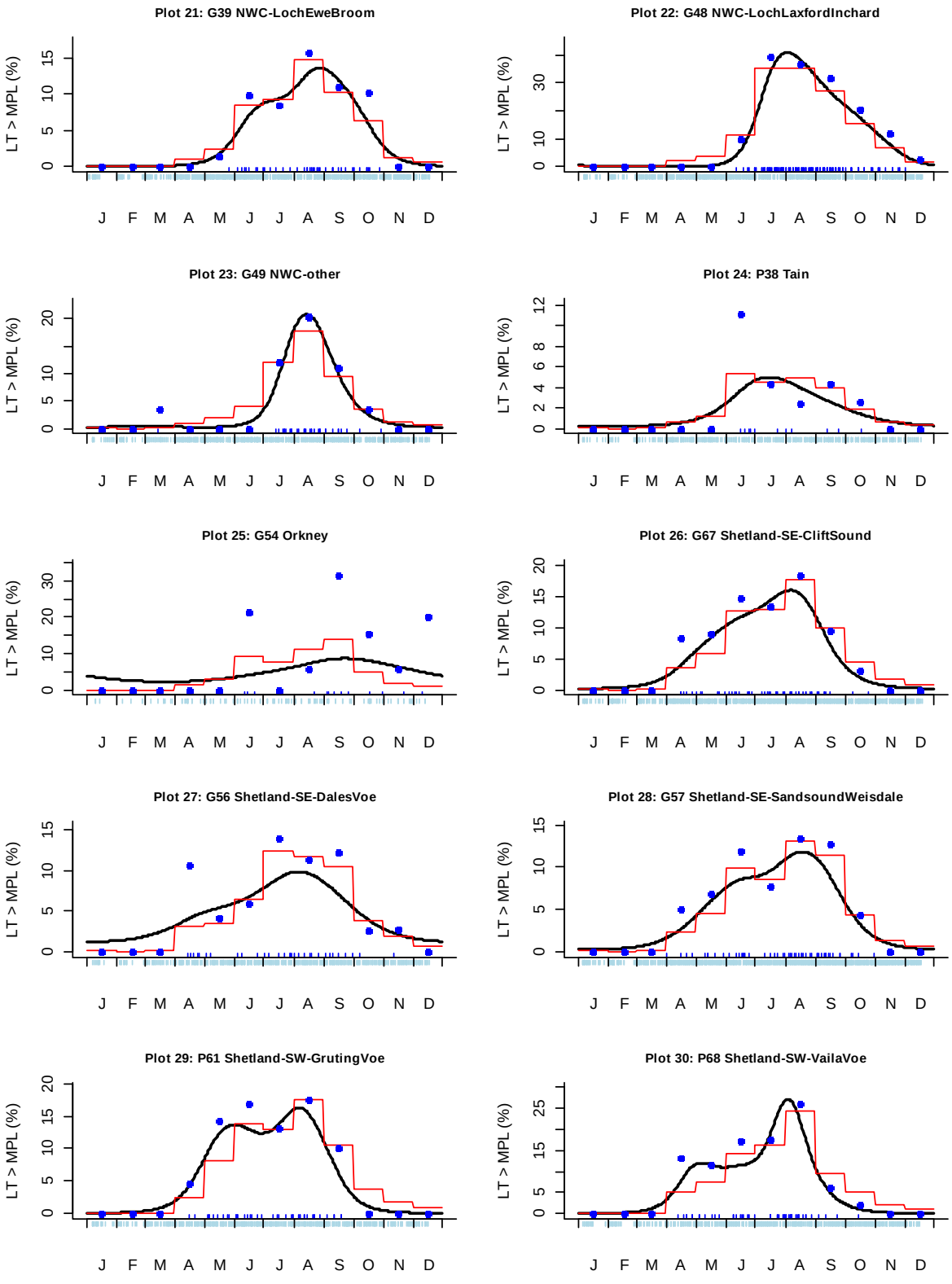


Figure C1 continued

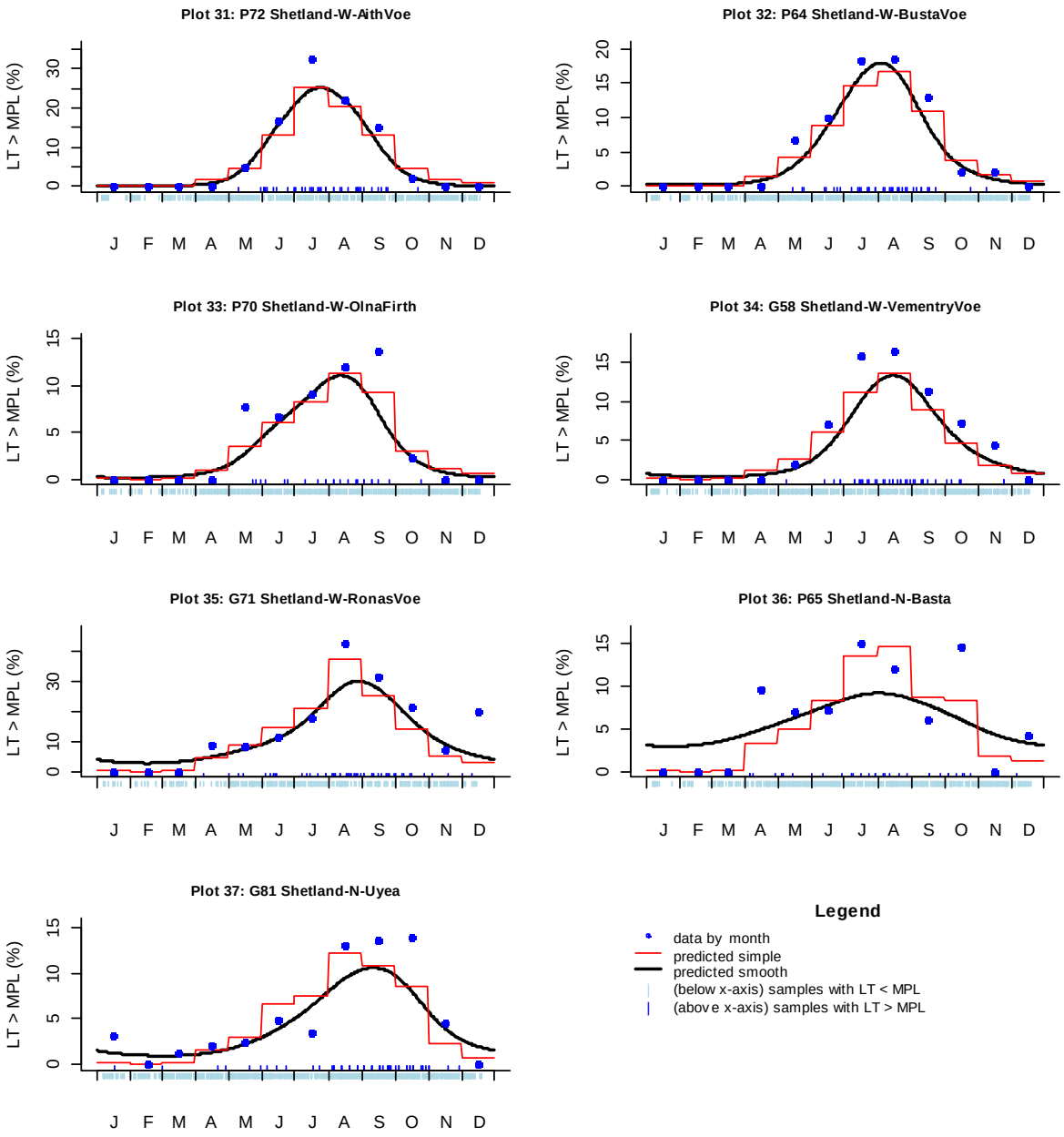


Figure C1 continued

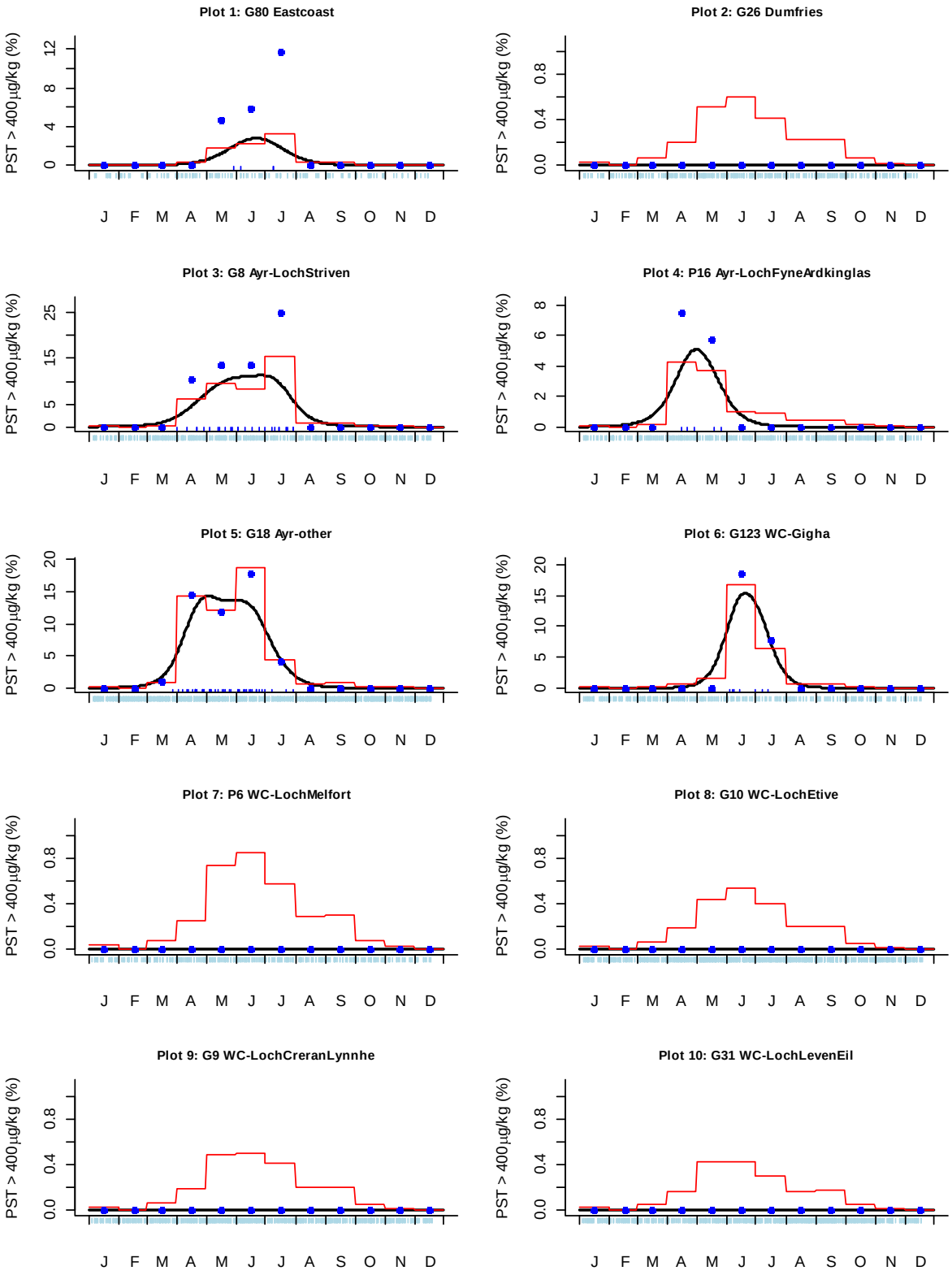


Figure C2: For each (group of) pod(s) the prevalence of PST > 400 µg/kg is shown, based on data (blue circles) from 2001-15, from simple models where predicted prevalence for an average toxin year is constant for each month (red lines, based on Holtrop et al. 2016), and predicted prevalence for an average toxin year from smooth models (black curve). The x-axis shows the days of the year

with ticks indicating when samples were obtained and whether their test results was below (light blue) or above (blue) 400 µg/kg.

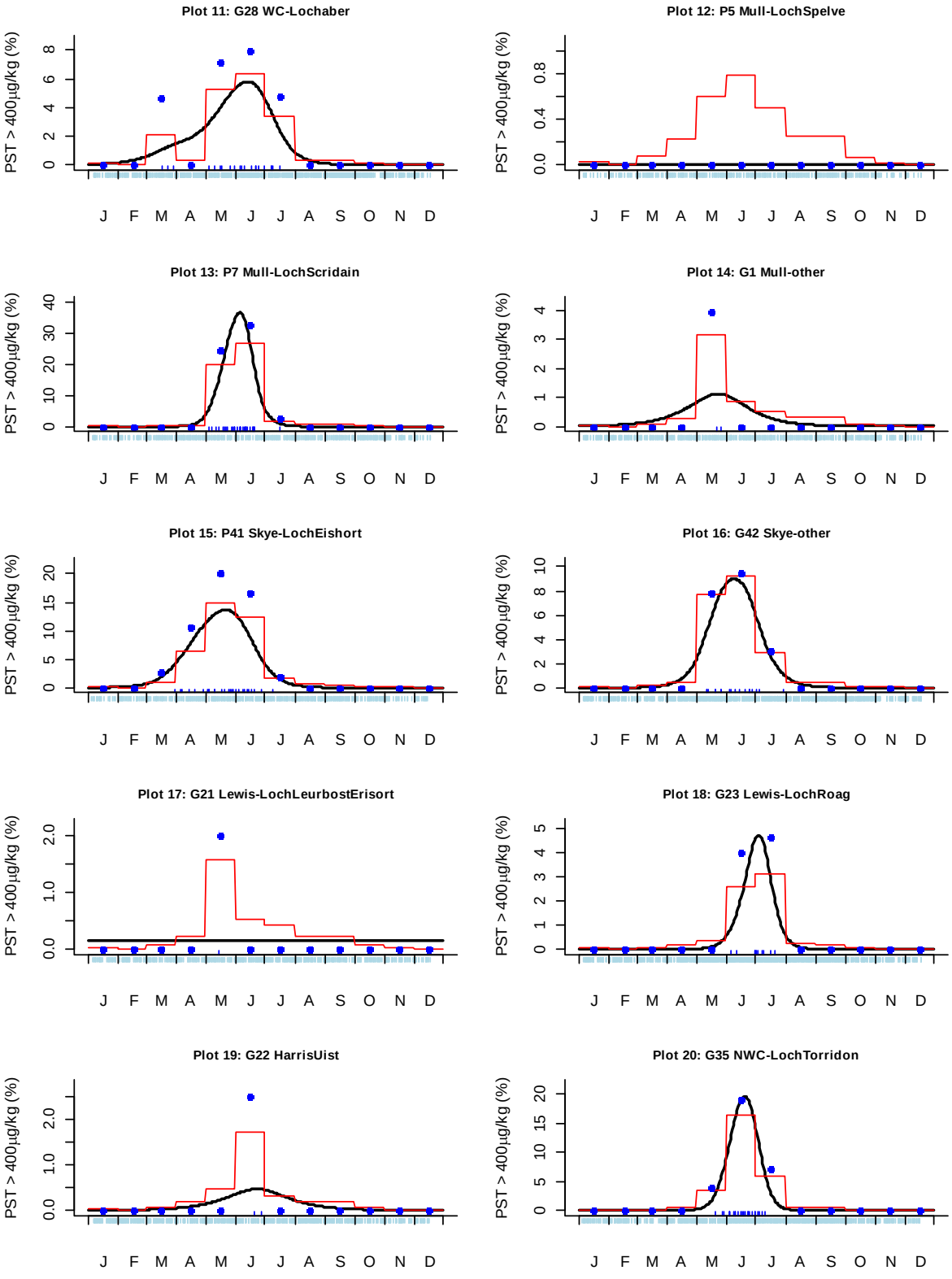


Figure C2 continued

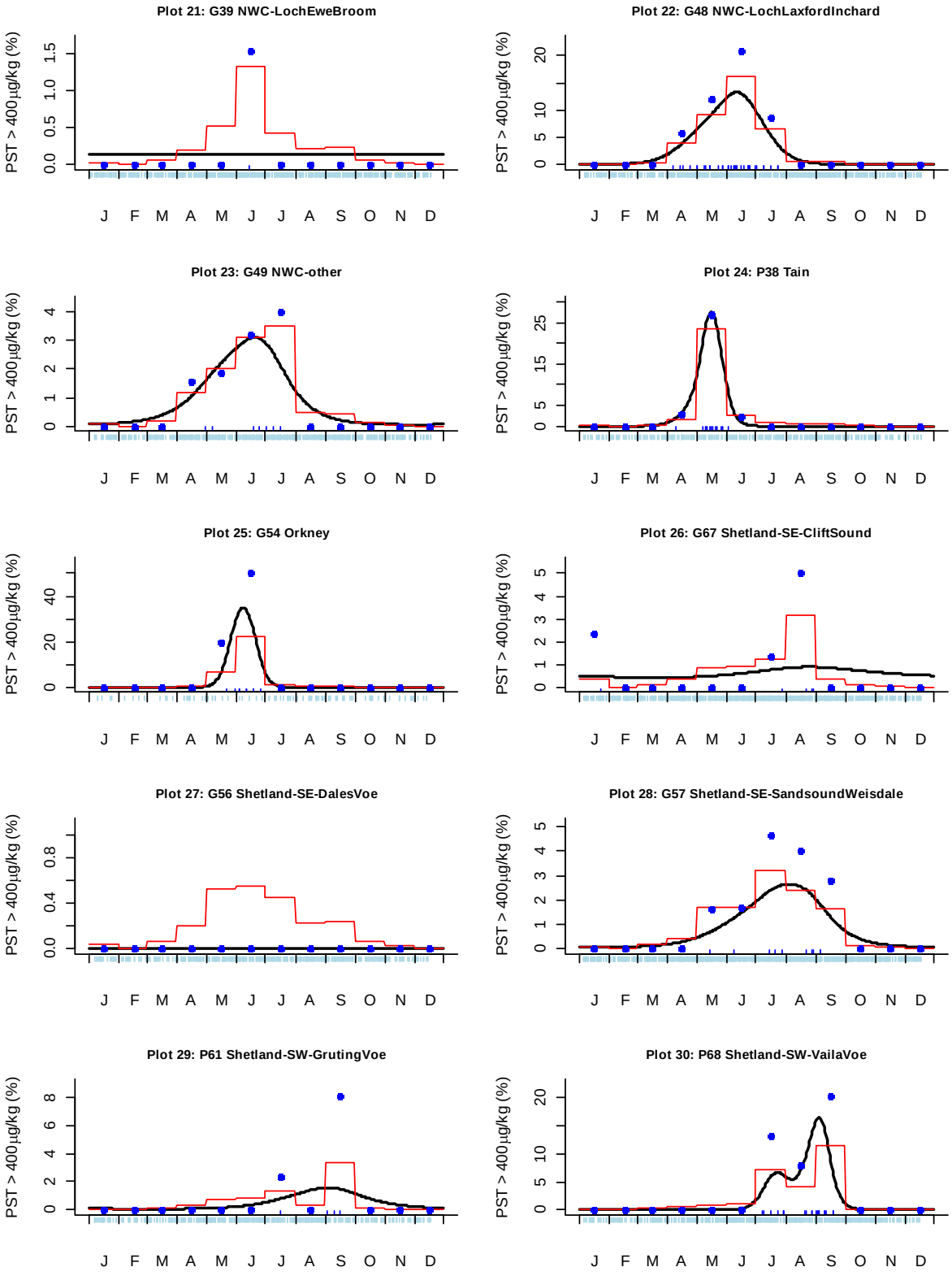


Figure C2 continued

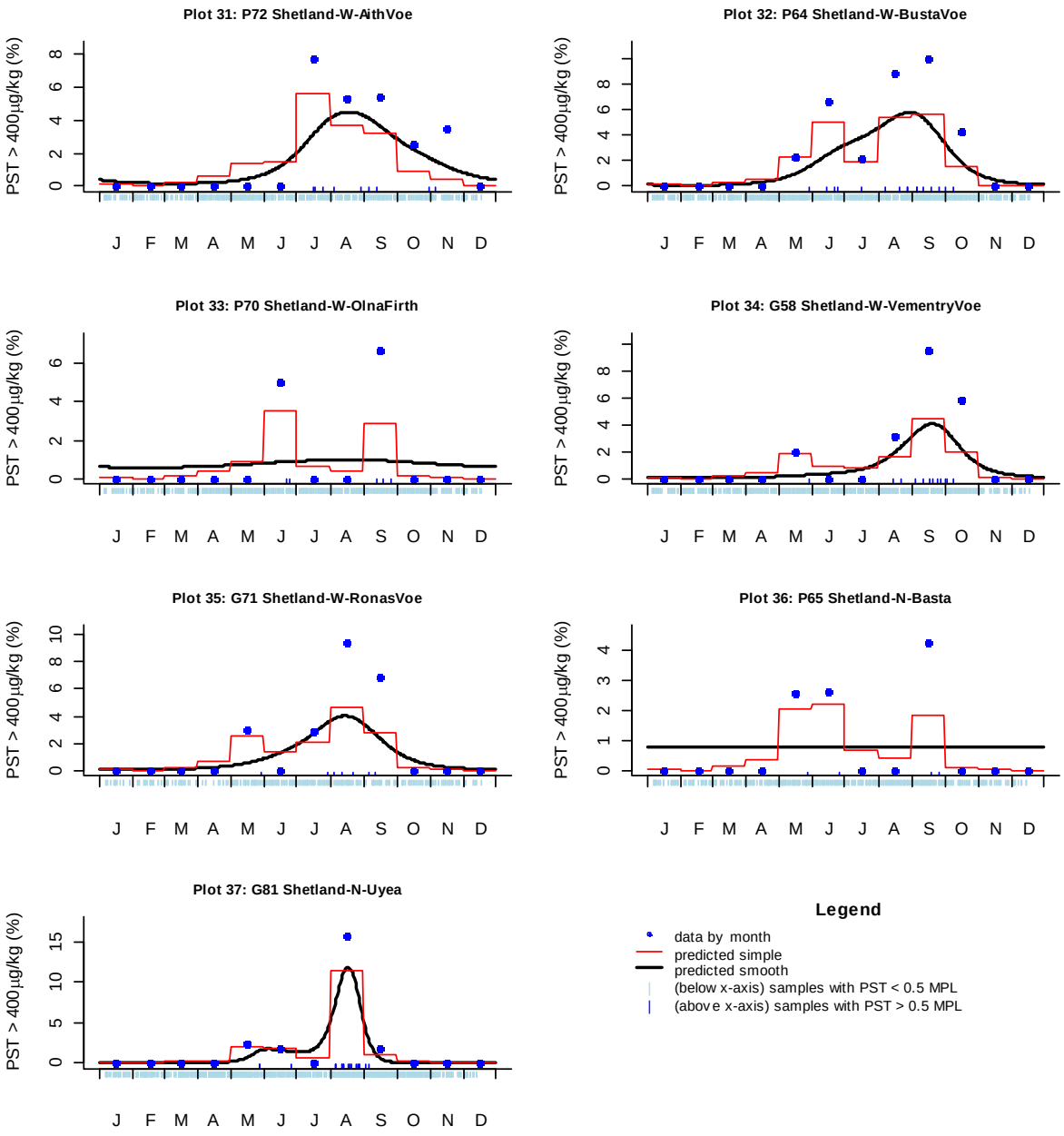


Figure C2 continued

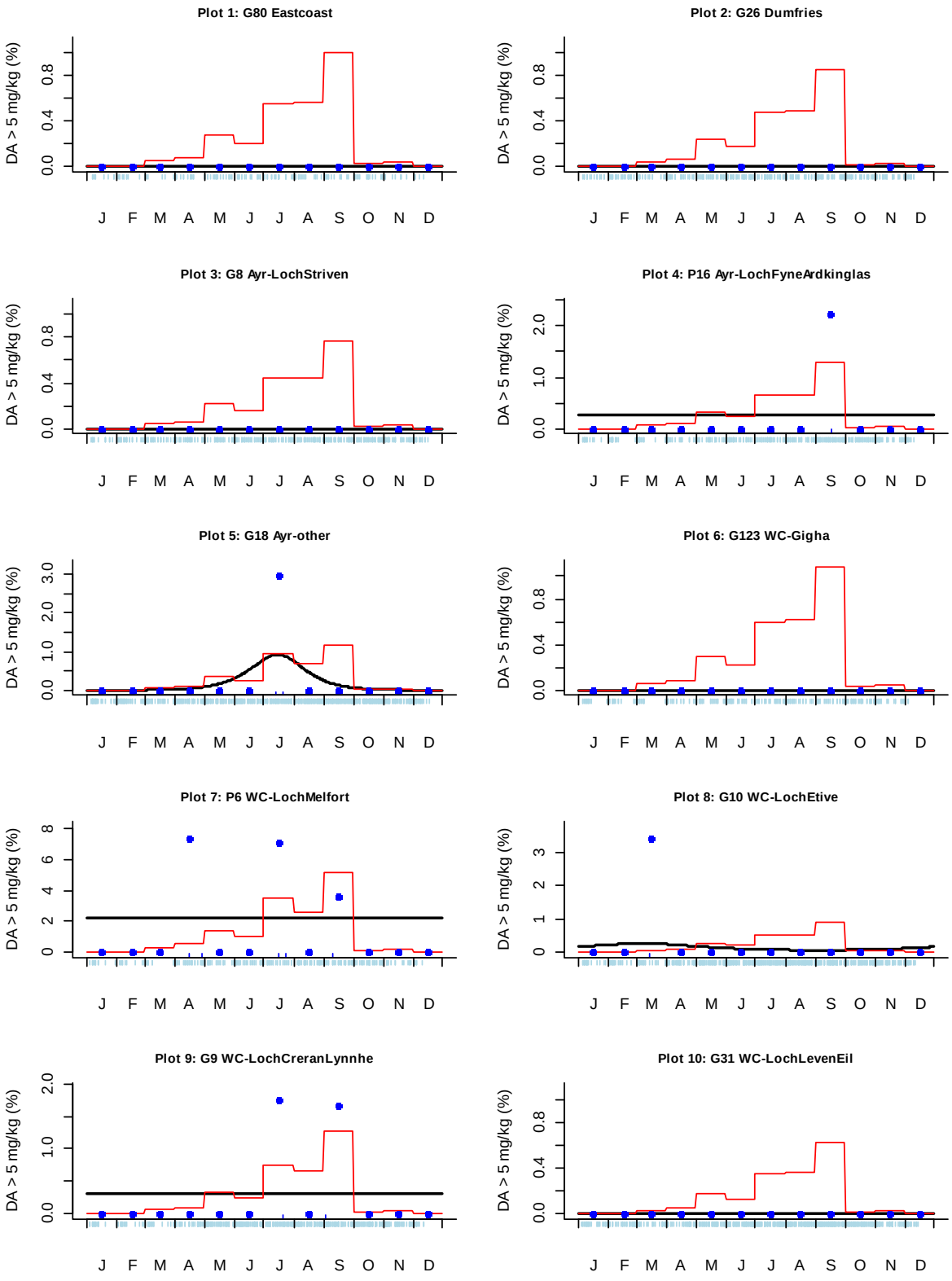


Figure C3: For each (group of) pod(s) the prevalence of DA > 5 mg/kg is shown, based on data (blue circles) from 2001-15, from simple models where predicted prevalence for an average toxin year is constant for each month (red lines, based on Holtrop et al. 2016), and predicted prevalence for an average toxin year from smooth models (black curve). The x-axis shows the days of the year with ticks indicating when samples were obtained and whether their test results was below (light blue) or above (blue) 5 mg/kg.

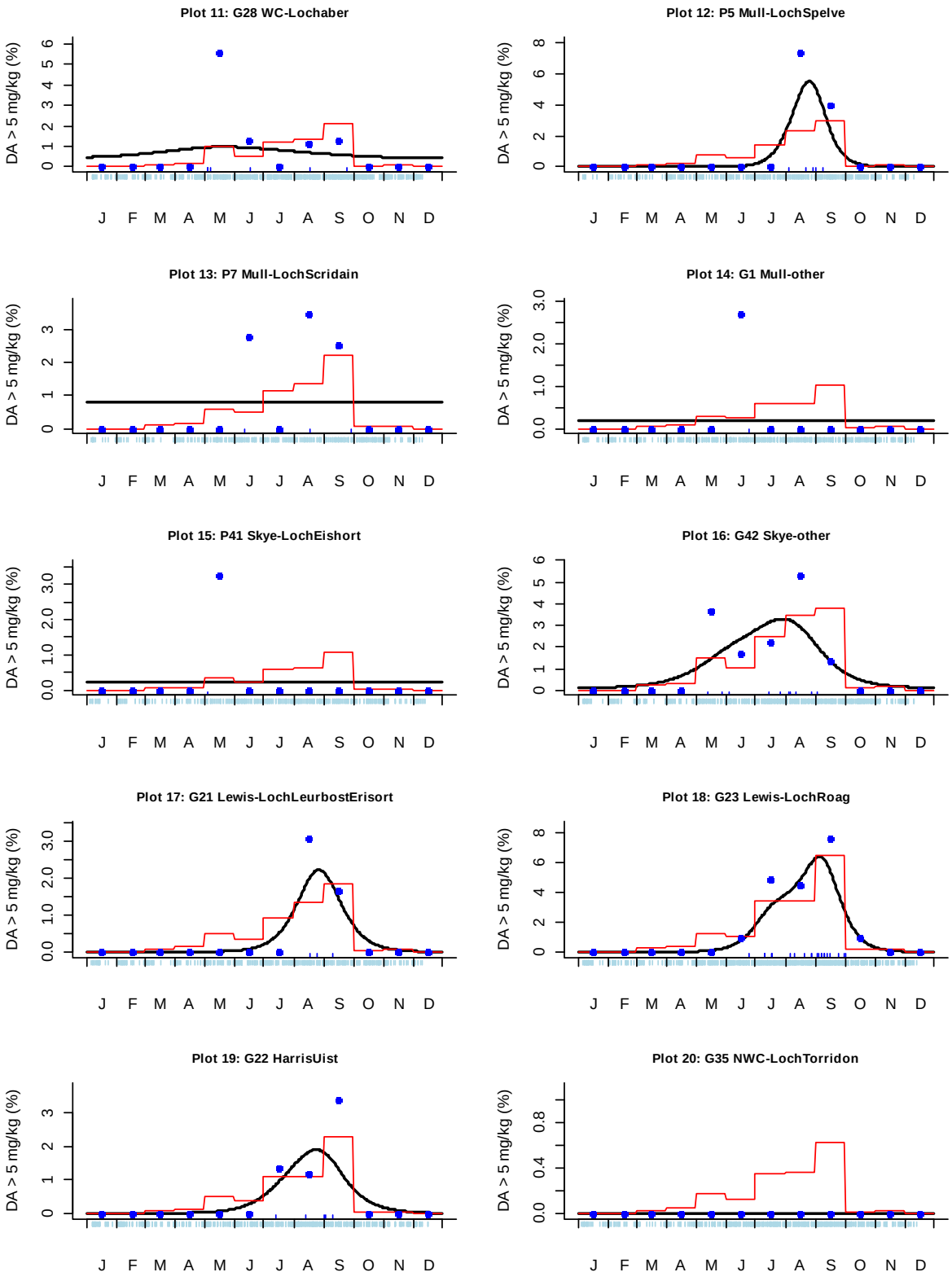


Figure C3 continued

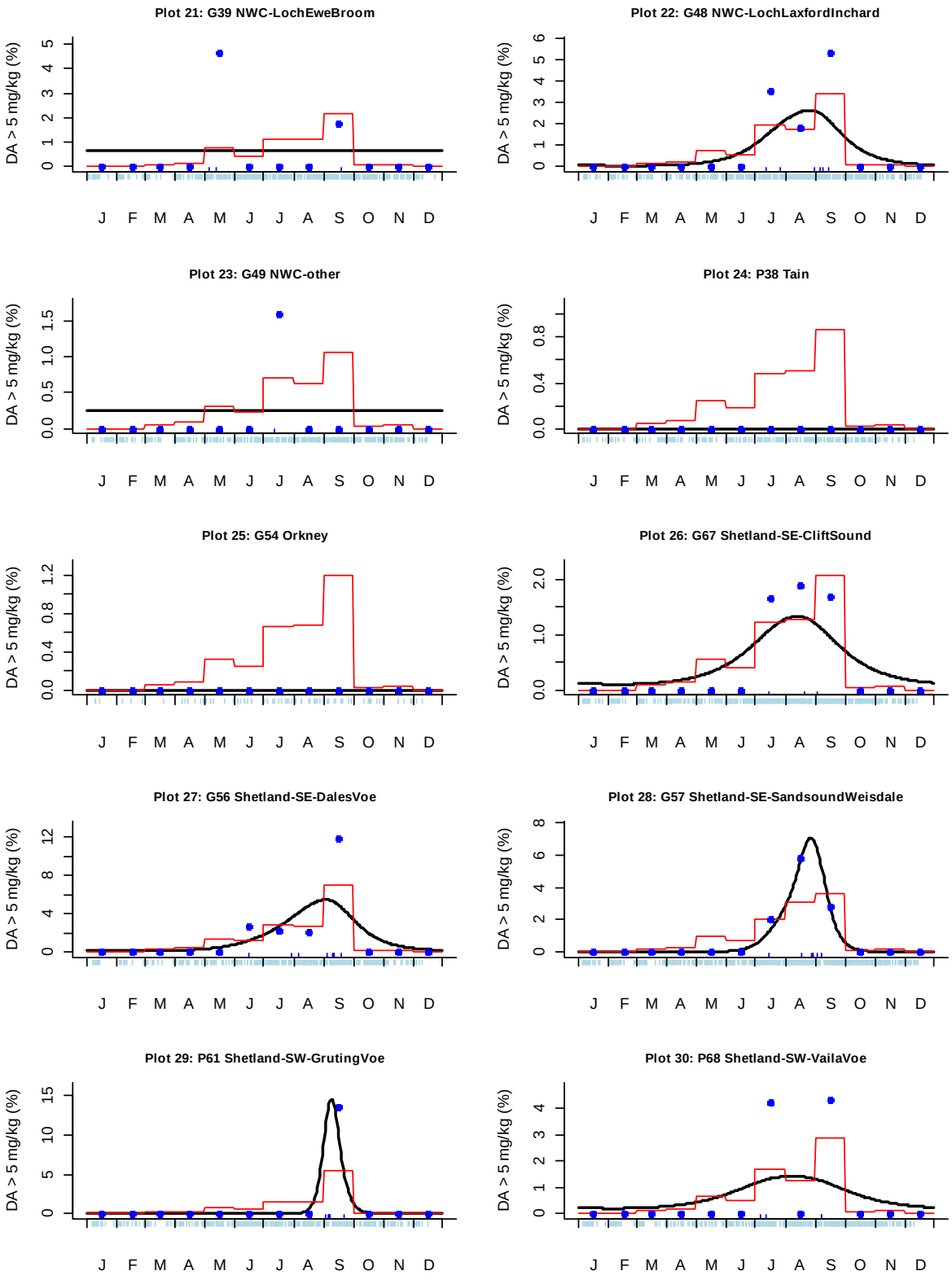


Figure C3 continued

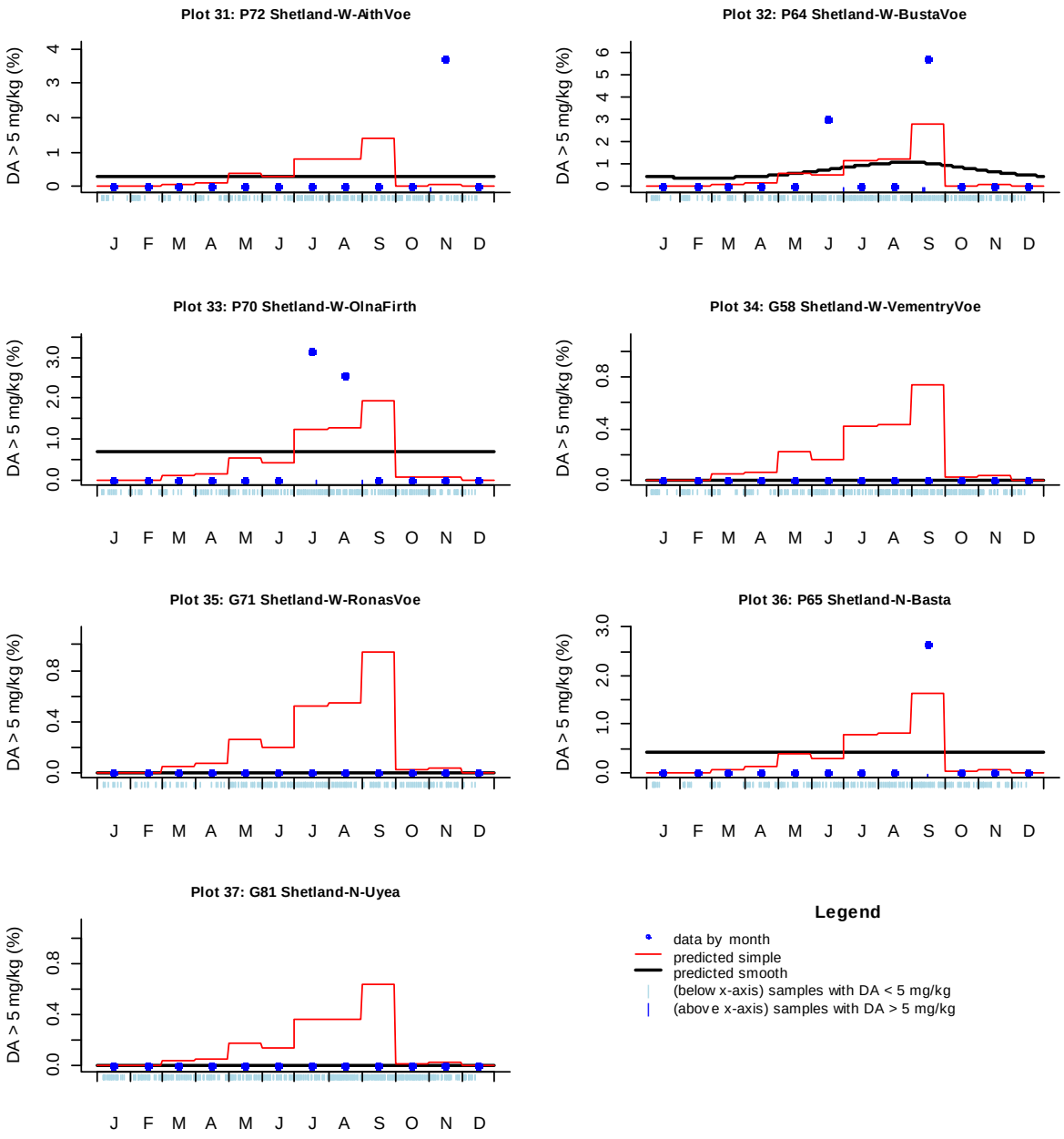


Figure C3 continued

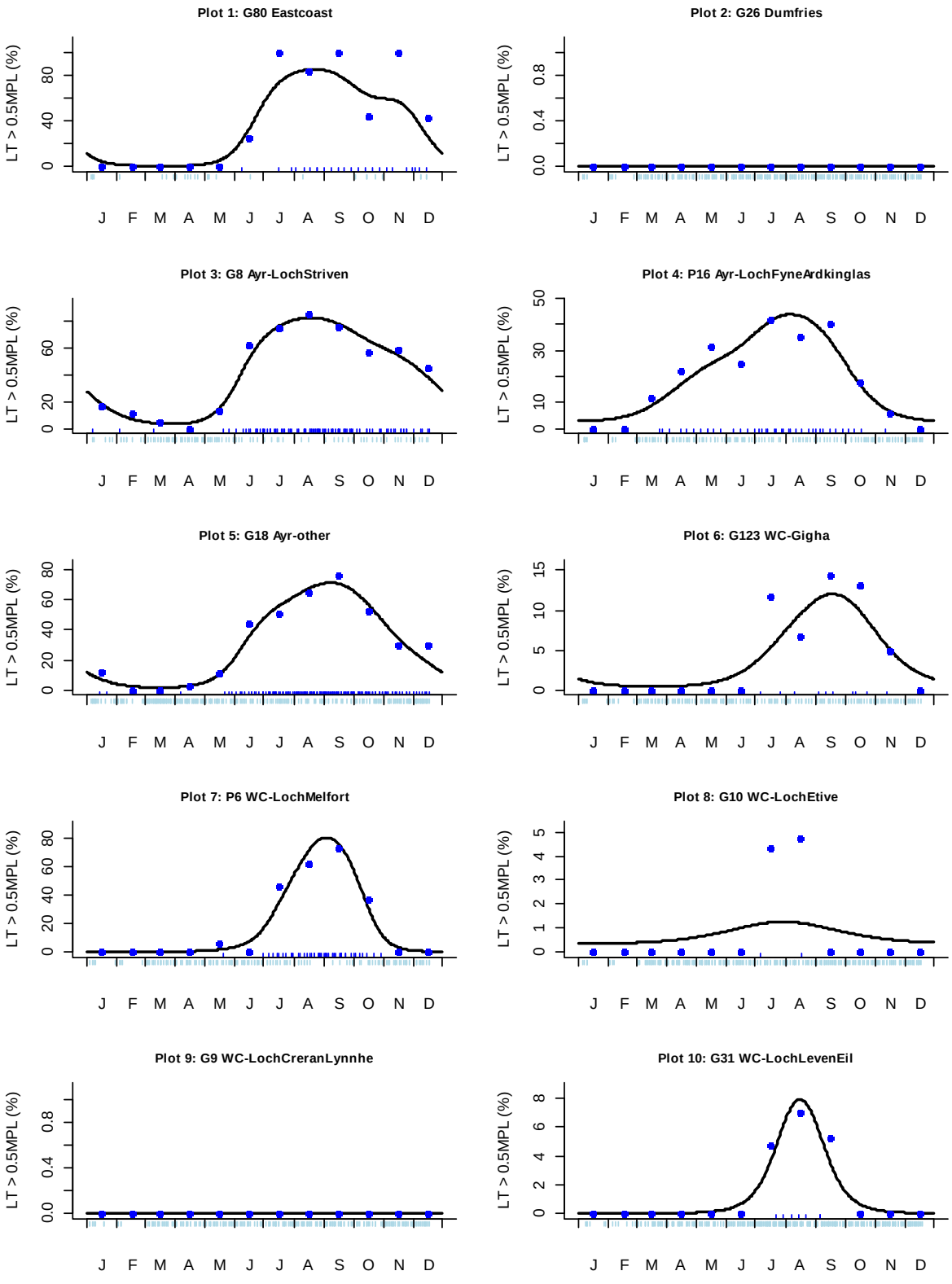


Figure C4: For each (group of) pod(s) the prevalence of LT-LCMS > 0.5 MPL is shown, based on data (blue circles) from 2011-15, and predicted prevalence for an average toxin year from smooth models (black curve). The x-axis shows the days of the year with ticks indicating when samples were obtained and whether their test results was below (light blue) or above (blue) 0.5 MPL.

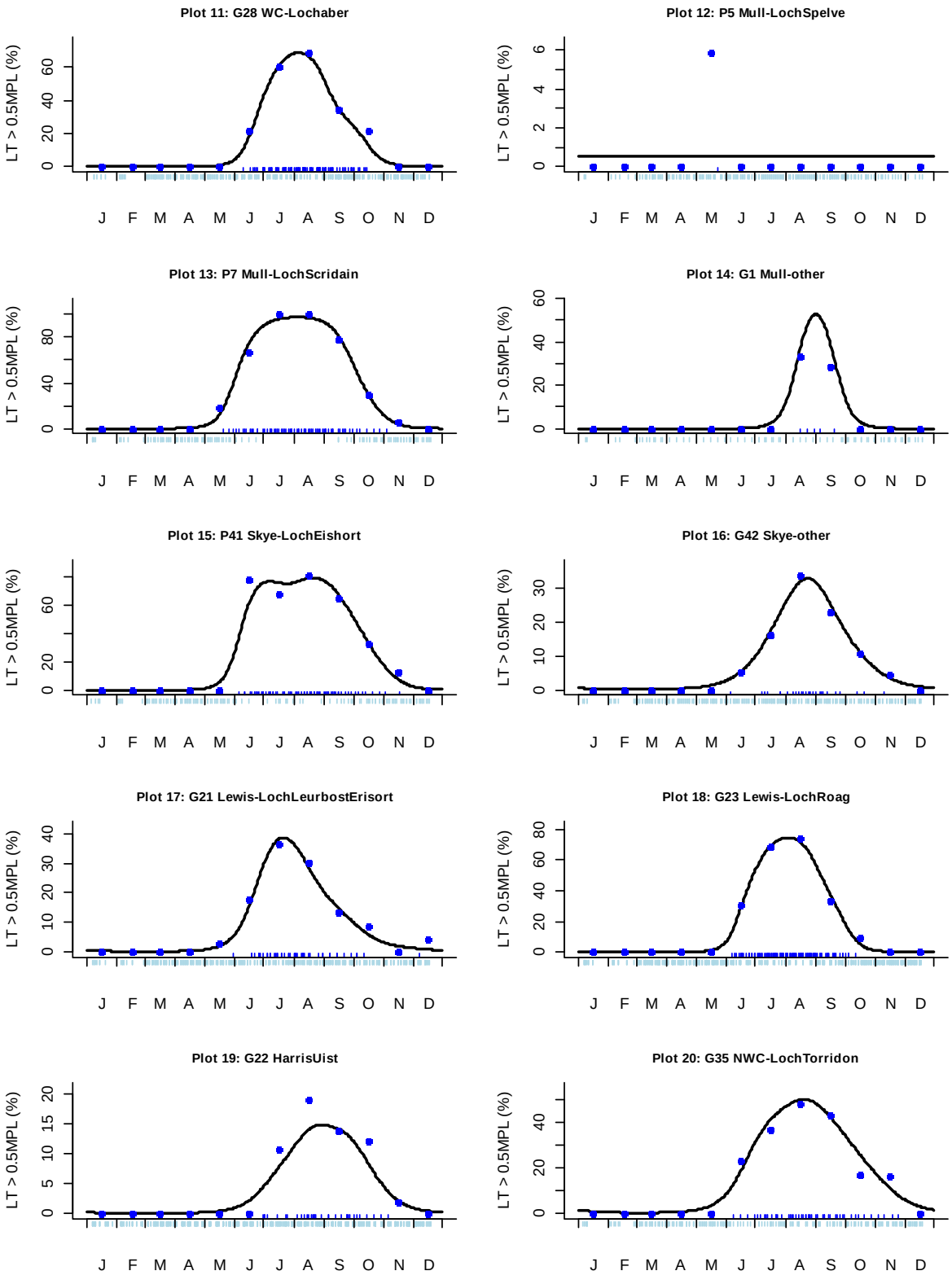


Figure C4 continued

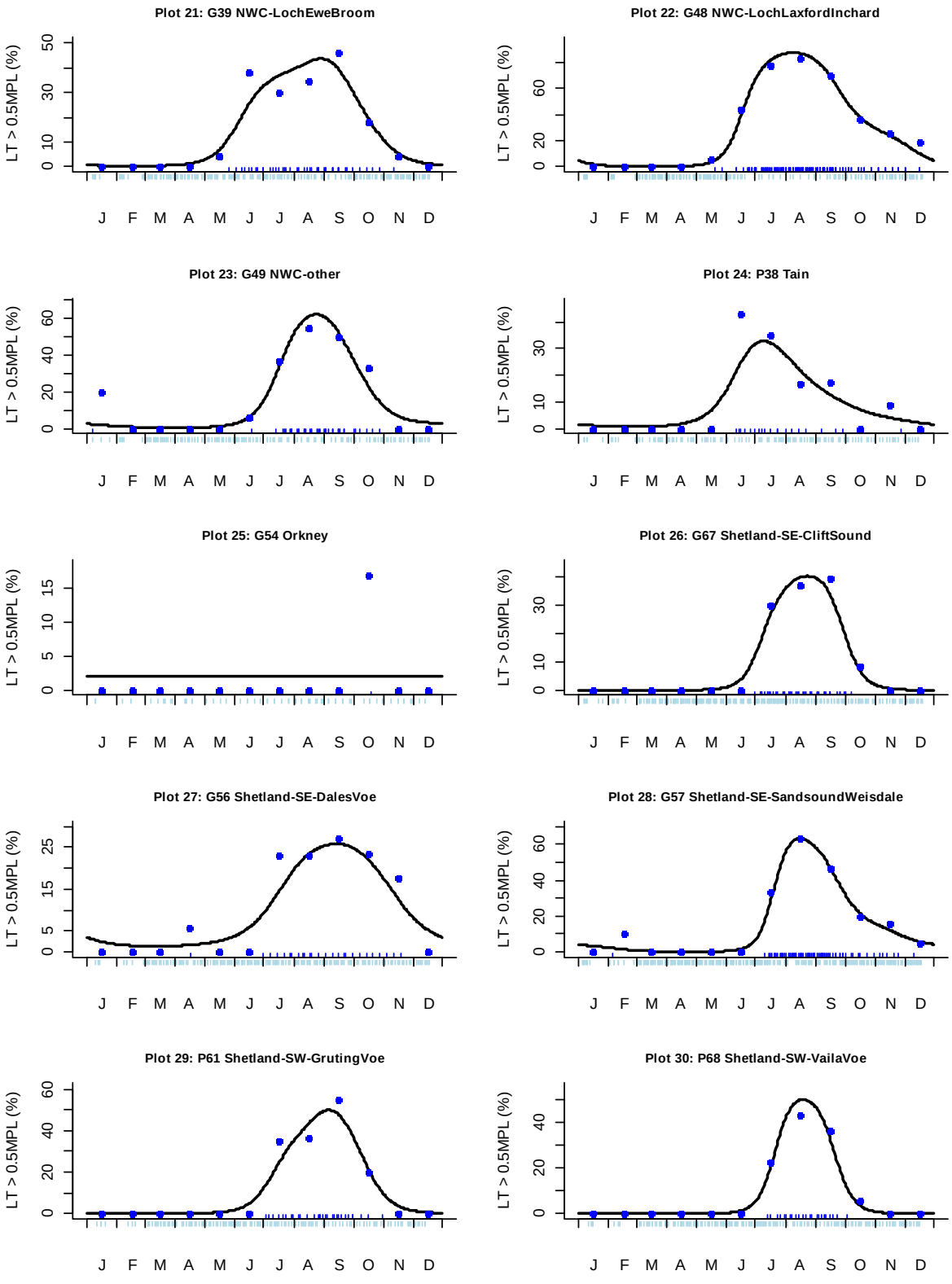


Figure C4 continued

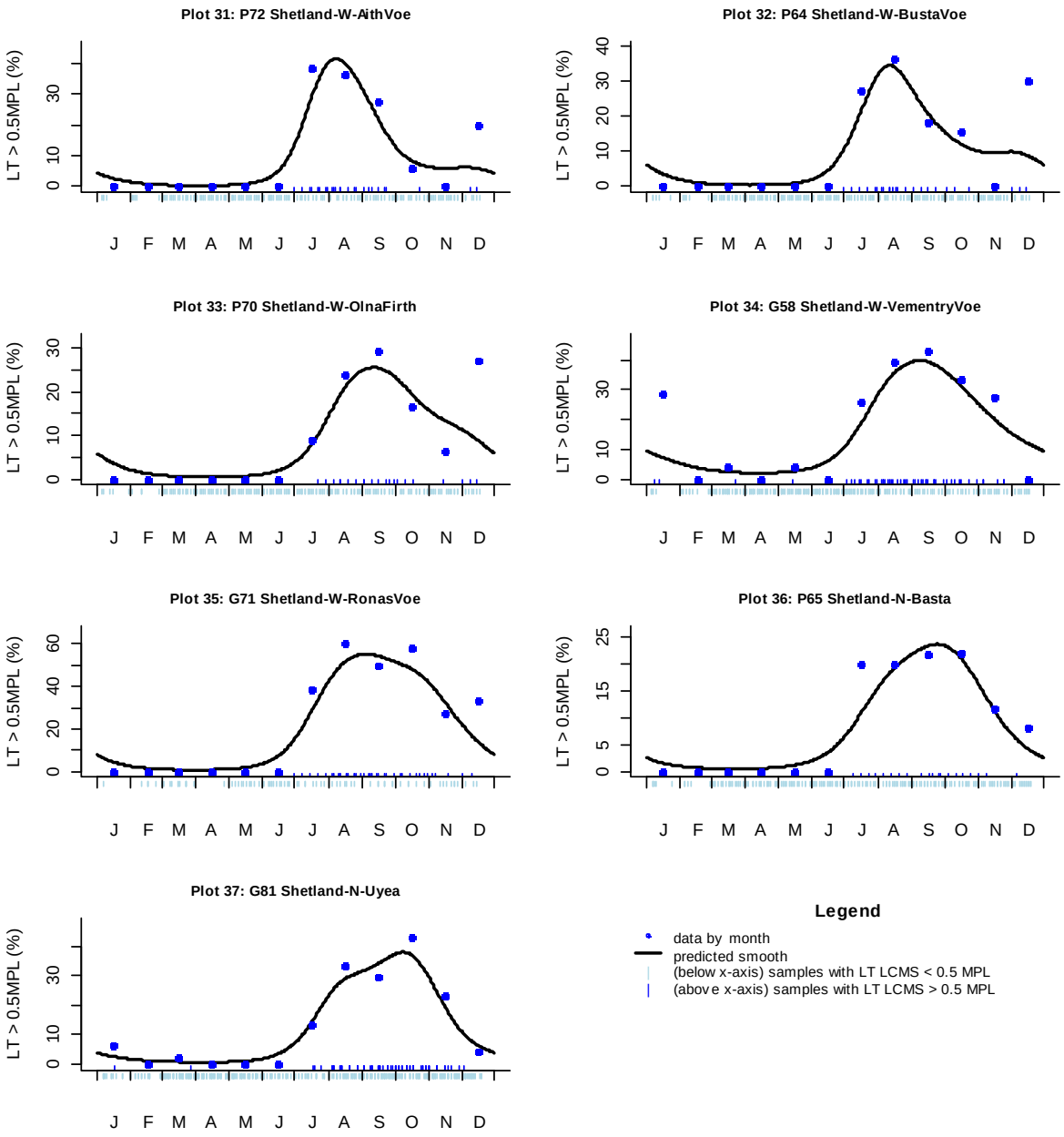


Figure C4 continued

Appendix D: Integrated display of predicted prevalence and observed biotoxin levels

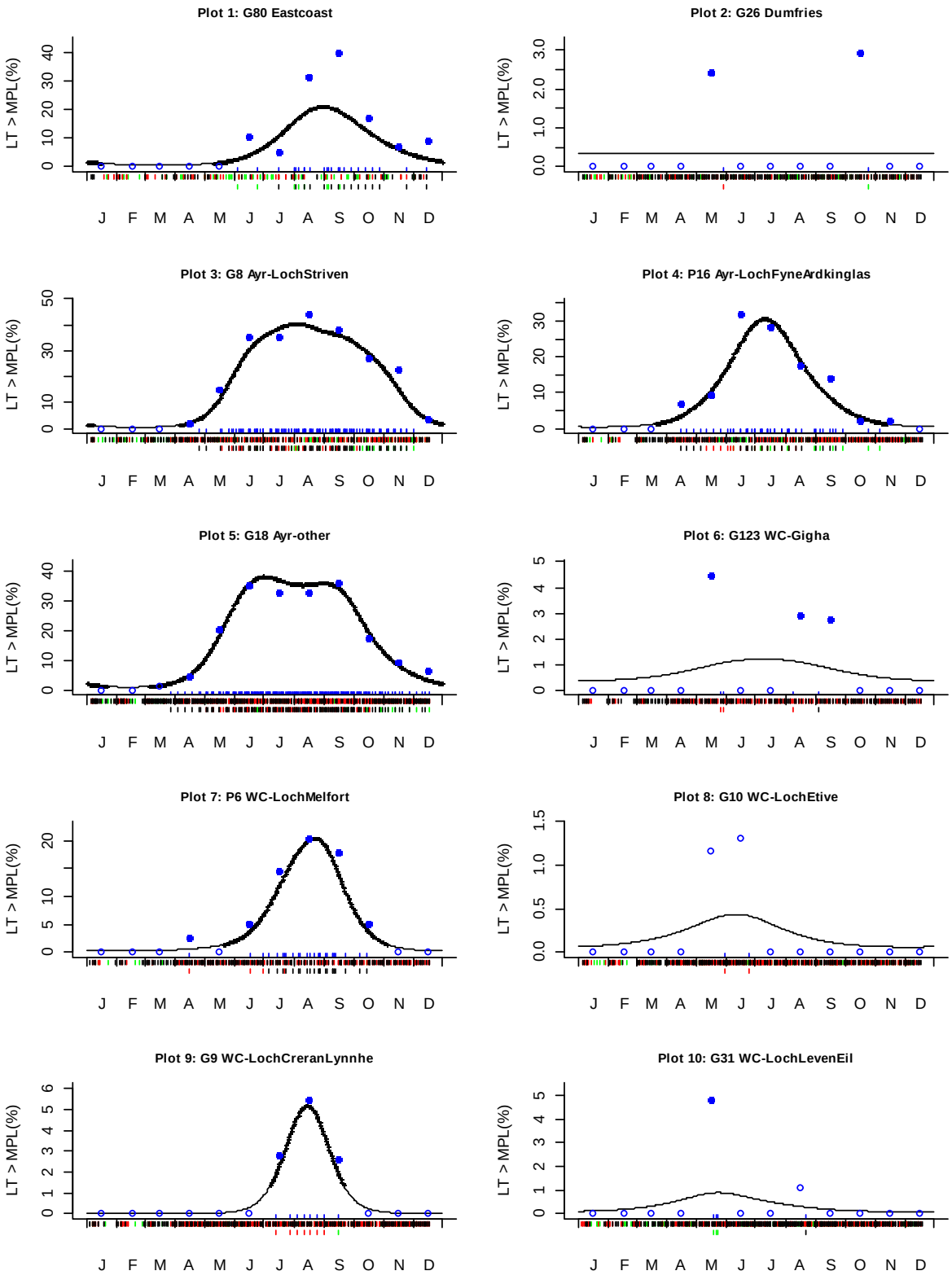


Figure D1: For each (group of) pod(s) the prevalence of LT > MPL is shown, based on data (blue circles) from 2001-15, and predicted prevalence for an average toxin year from smooth models (black curve). Bold lines and closed symbols indicated is where monthly sampling is insufficient (prevalence > cut-off and fortnightly or weekly sampling would be required to keep the risk of non-detection below 1%). The x-axis shows the days of the year with ticks indicating when samples were obtained and whether their test results was below (light blue) or above (blue) MPL. The colouring of these ticks is as follows: green 2001-5, red 2006-10, black 2011-15.

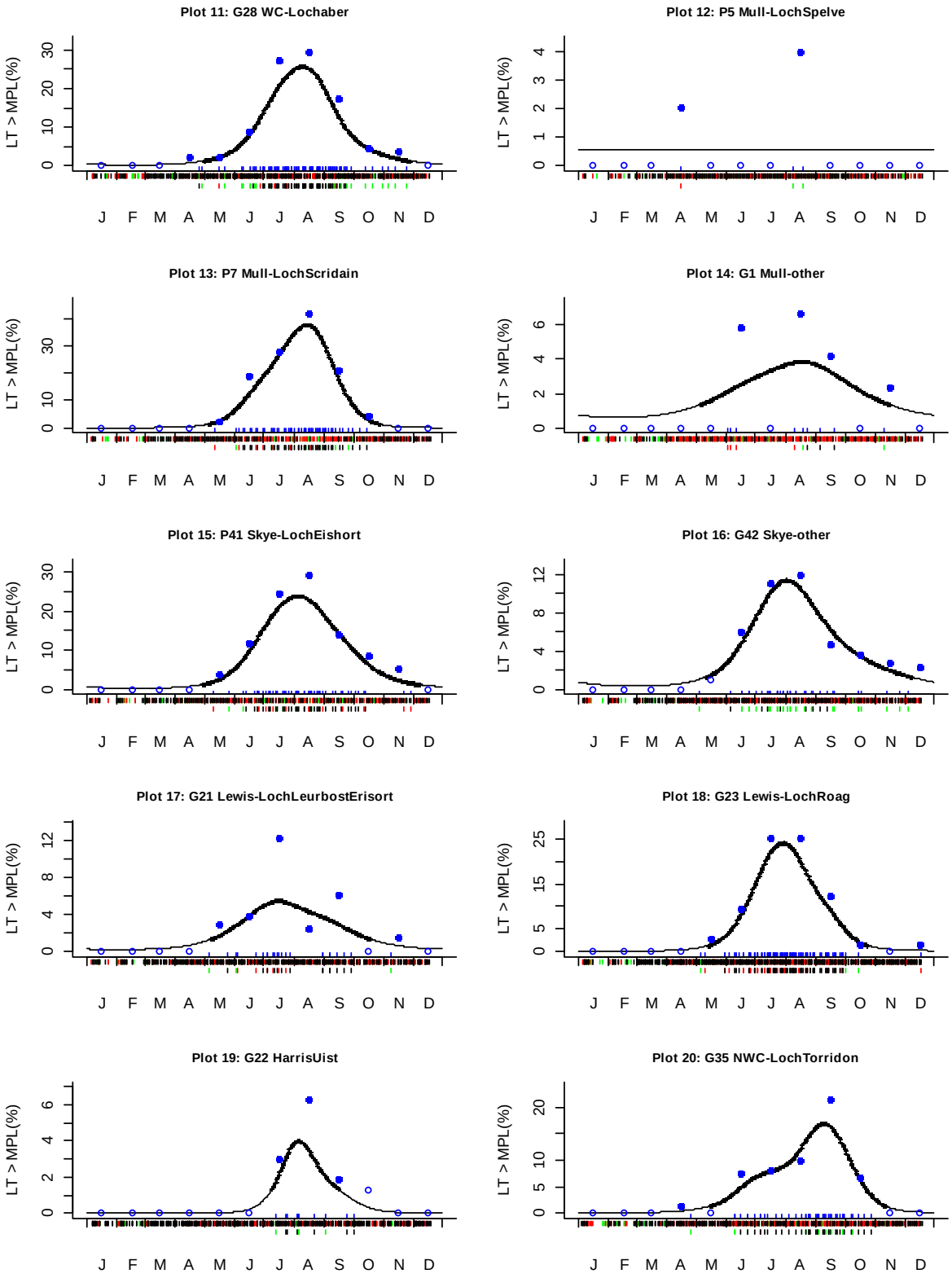


Figure D1 continued

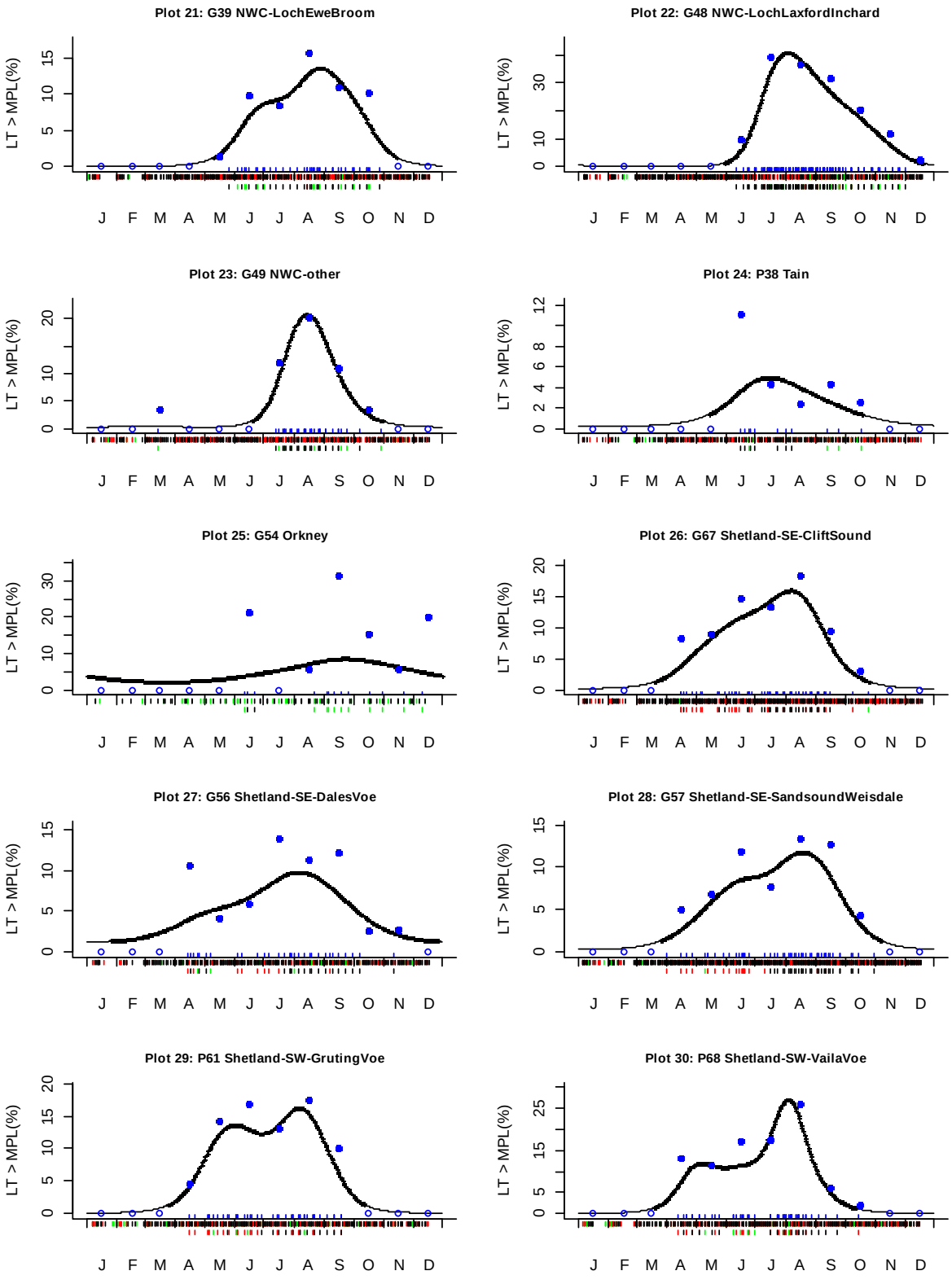


Figure D1 continued

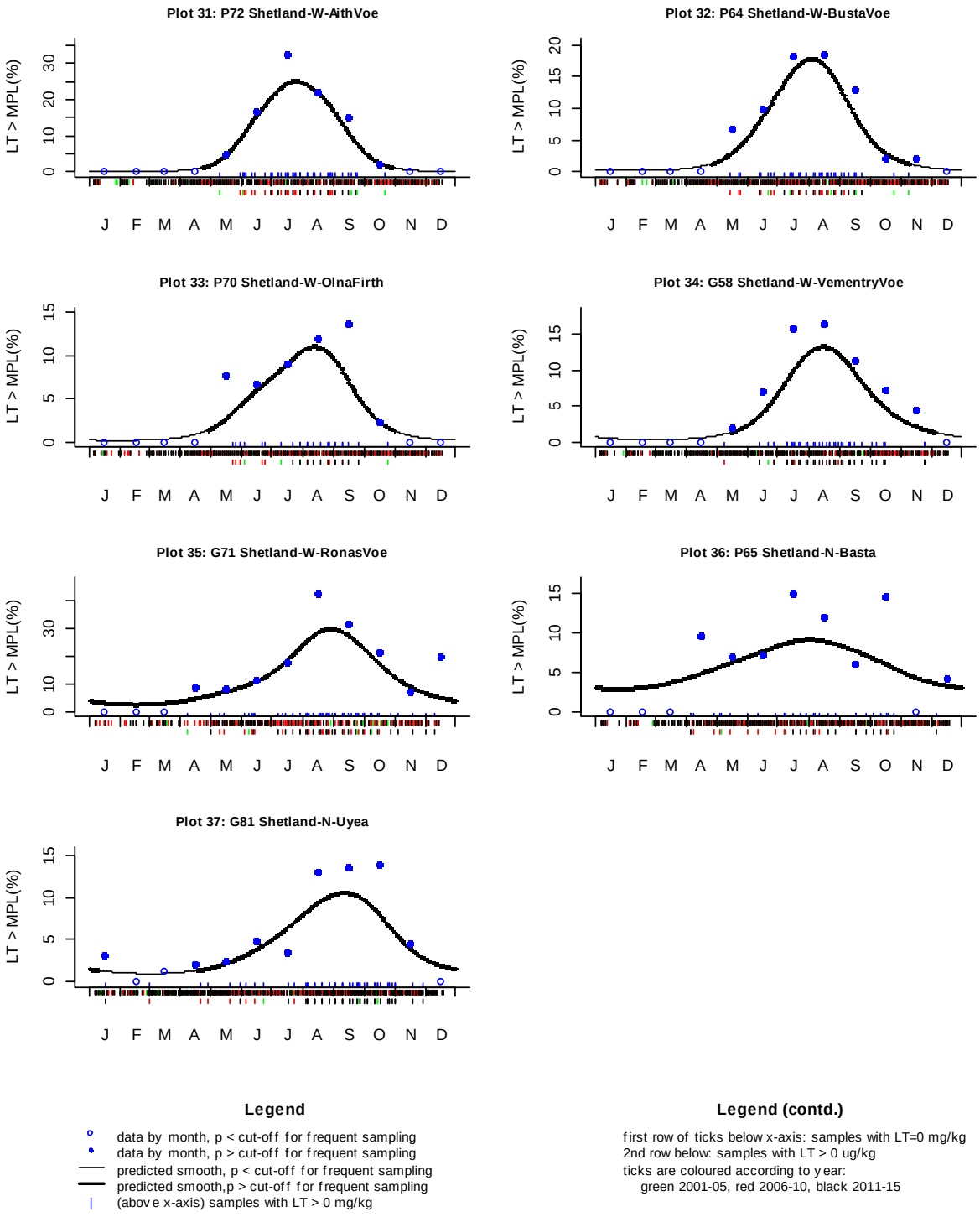


Figure D1 continued

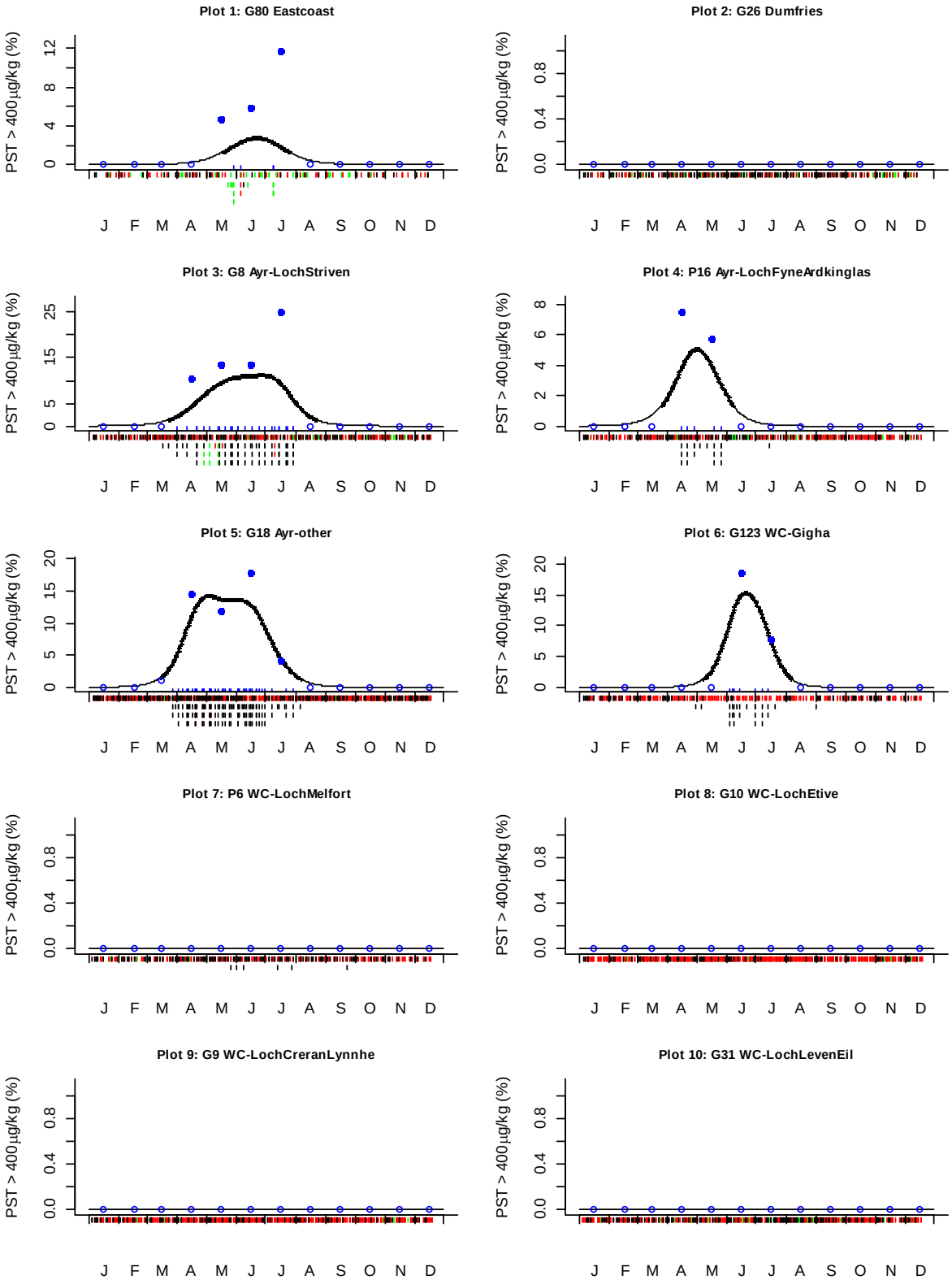


Figure D2: For each (group of) pod(s) the prevalence of PST > 400 µg/kg is shown, based on data (blue circles) from 2001-15, and predicted prevalence for an average toxin year from smooth models (black curve). Bold lines and closed symbols indicated is where monthly sampling is insufficient (prevalence > cut-off and fortnightly or weekly sampling would be required to keep the risk of non-detection below 1%). The x-axis shows the days of the year. Ticks above the x-axis are samples for which PST > 400 µg/kg. Ticks below the x-axis indicate the following: first row of ticks: samples for which PST = 0mg/kg, second row of ticks: 0 < PST < 400 µg/kg, the third row of ticks: 400 ≤ PST < 800 mµ/kg, and the fourth row of ticks: PST ≥ 800 µg/kg. The colouring of these ticks is as follows: green 2001-5, red 2006-10, black 2011-15.

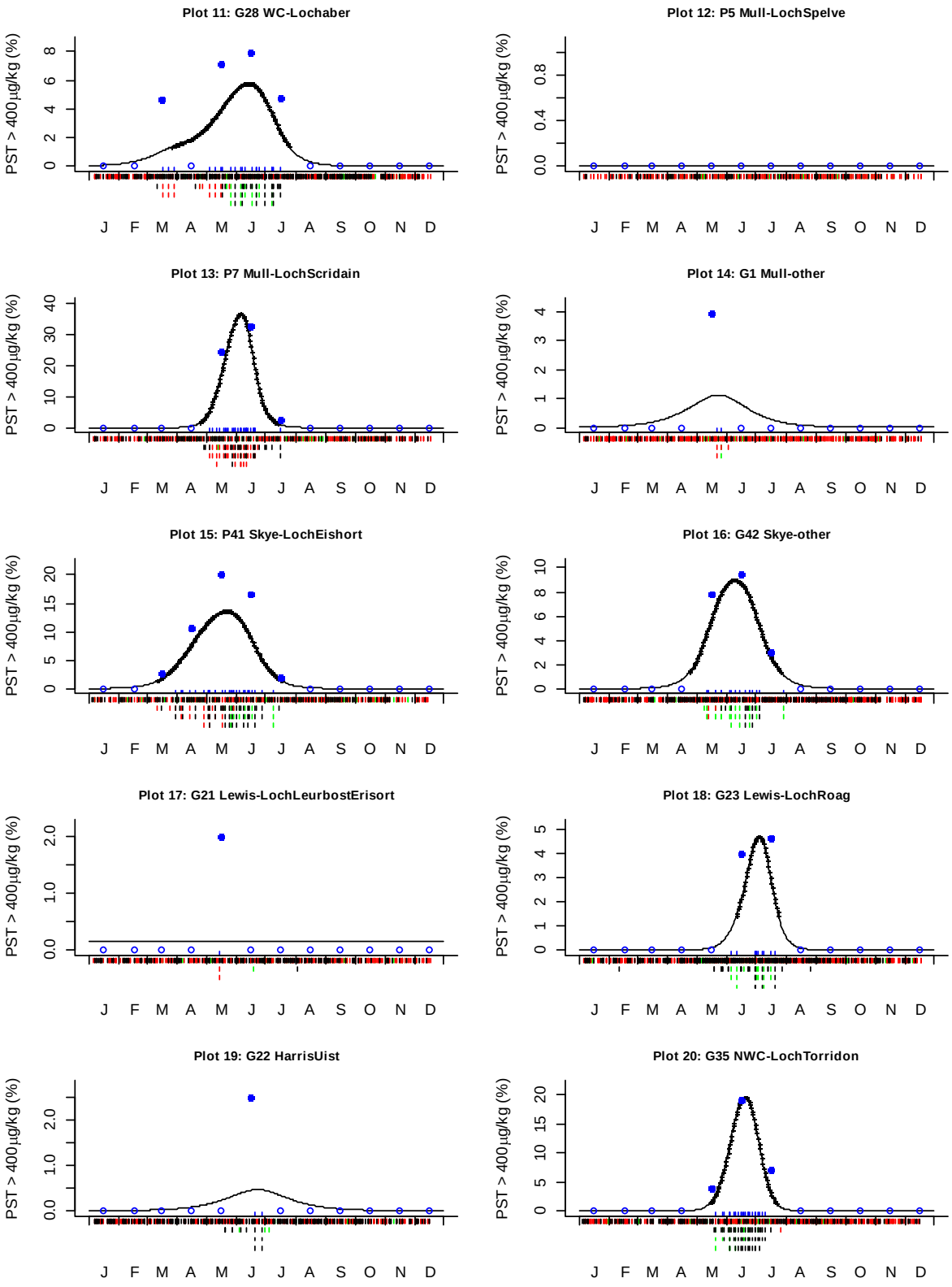


Figure D2 continued

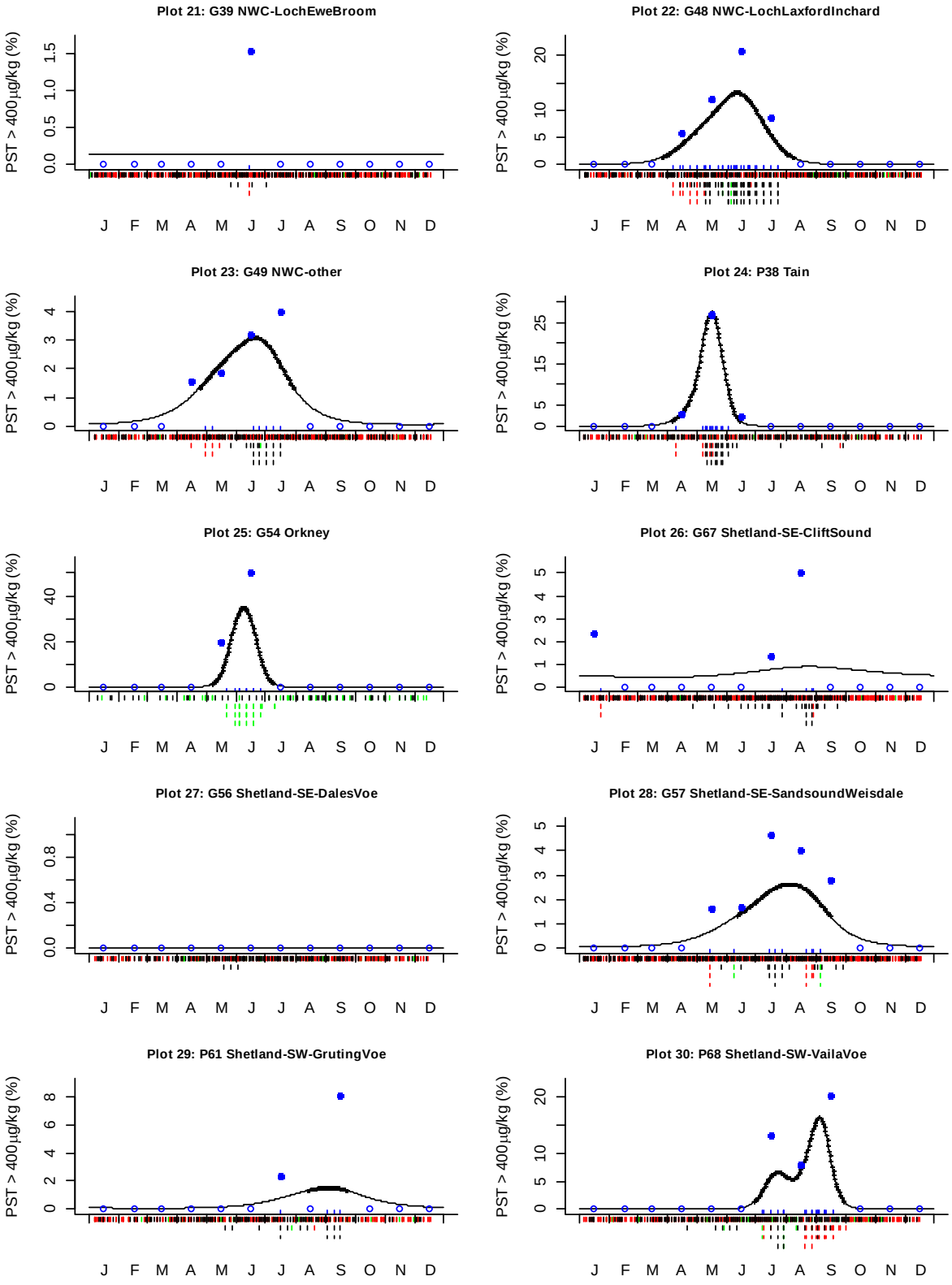


Figure D2 continued

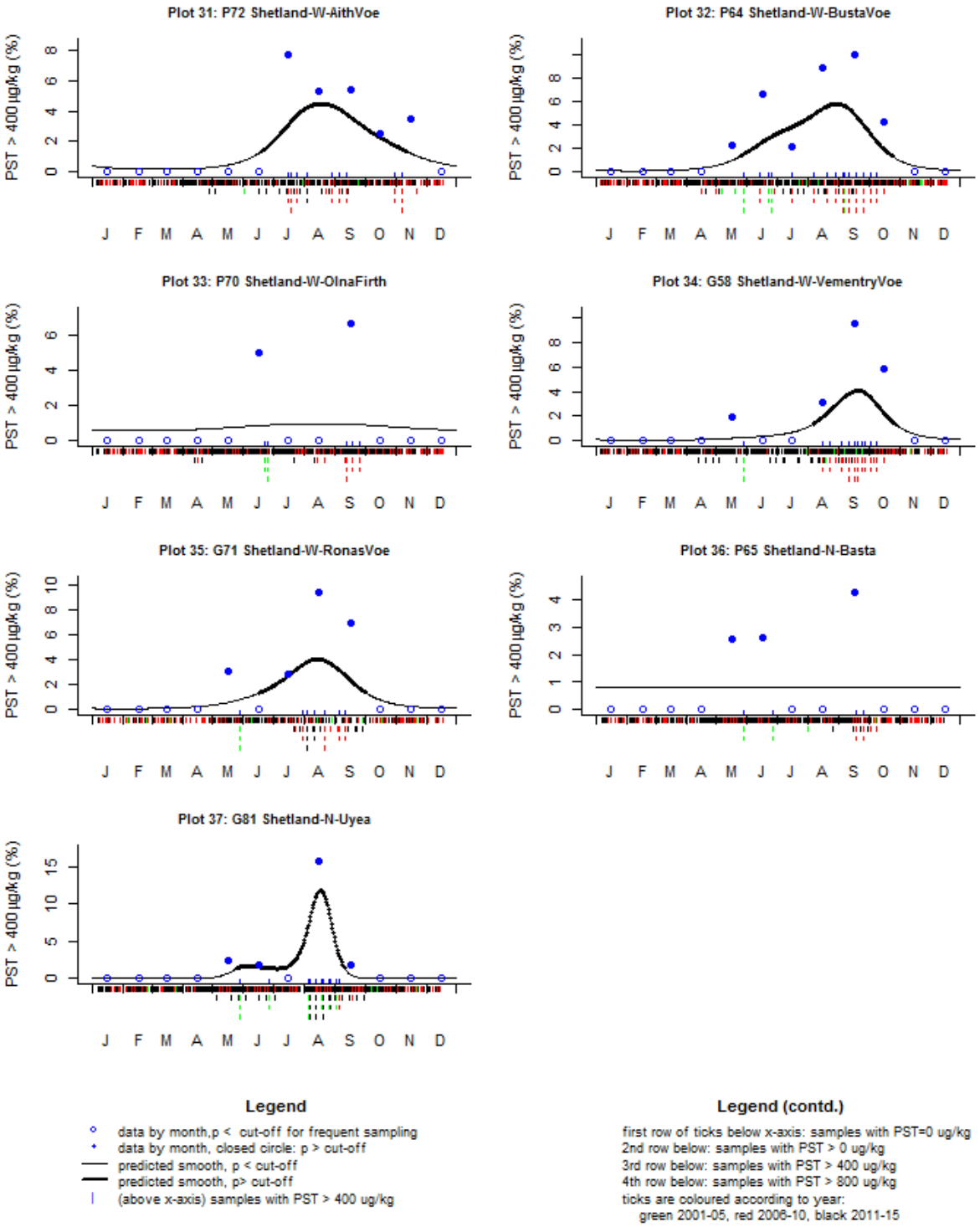


Figure D2 continued.

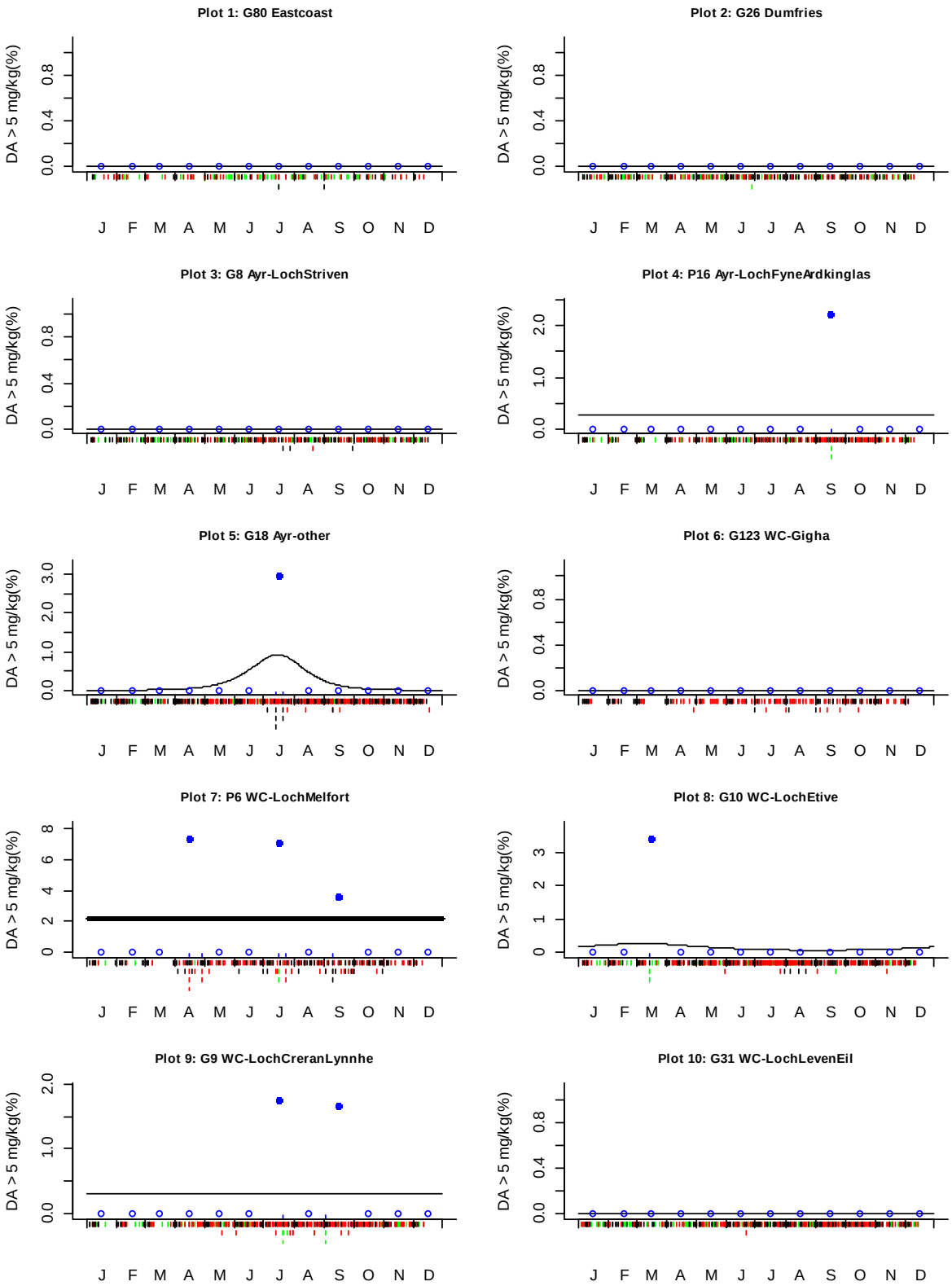


Figure D3: For each (group of) pod(s) the prevalence of DA > 5 mg/kg is shown, based on data (blue circles) from 2001-15, and predicted prevalence for an average toxin year from smooth models (black curve). Bold lines and closed symbols indicated is where monthly sampling is insufficient (prevalence > cut-off and fortnightly or weekly sampling would be required to keep the risk of non-detection below 1%). The x-axis shows the days of the year. Ticks above the x-axis are samples for which DA > 5 mg/kg. Ticks below the x-axis indicate the following: first row of ticks: samples for which DA = 0 mg/kg, second row of ticks: 0 < DA < 5 mg/kg, the third row of ticks: 5 ≤ DA < 10mg/kg, and the fourth row of ticks: DA ≥ 10 mg/kg. The colouring of these ticks is as follows: green 2001-5, red 2006-10, black 2011-15.

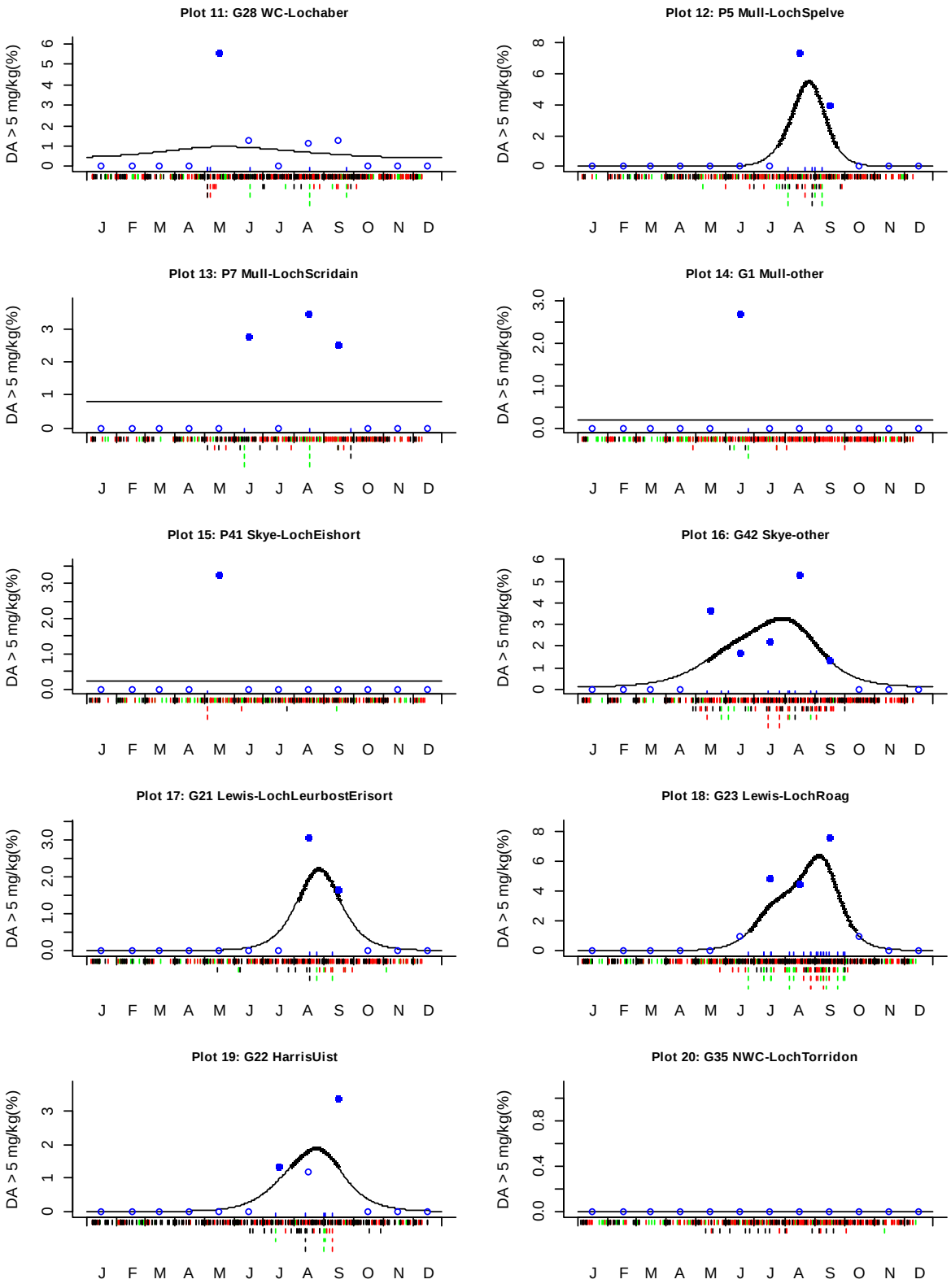


Figure D3 continued

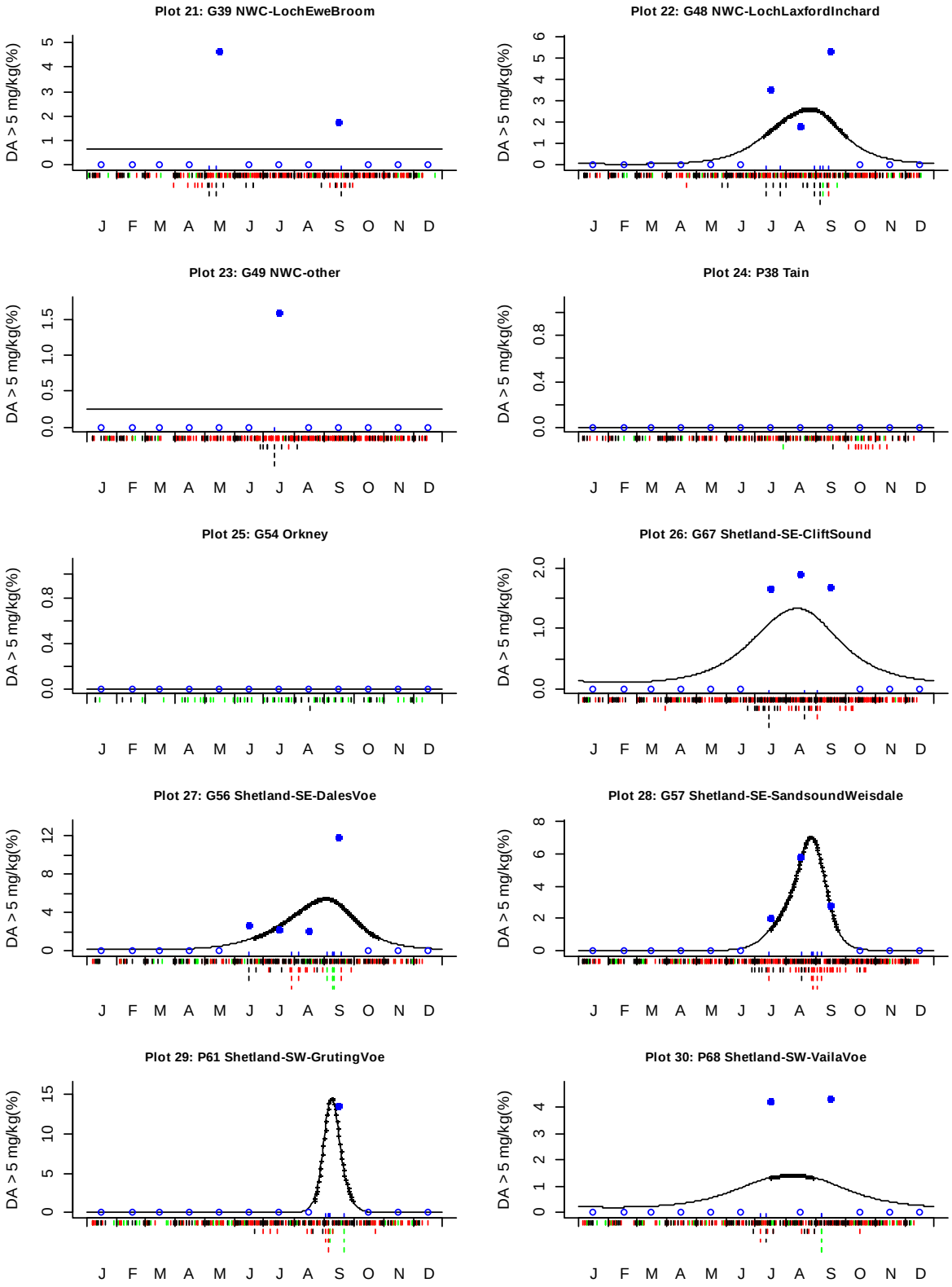


Figure D3 continued

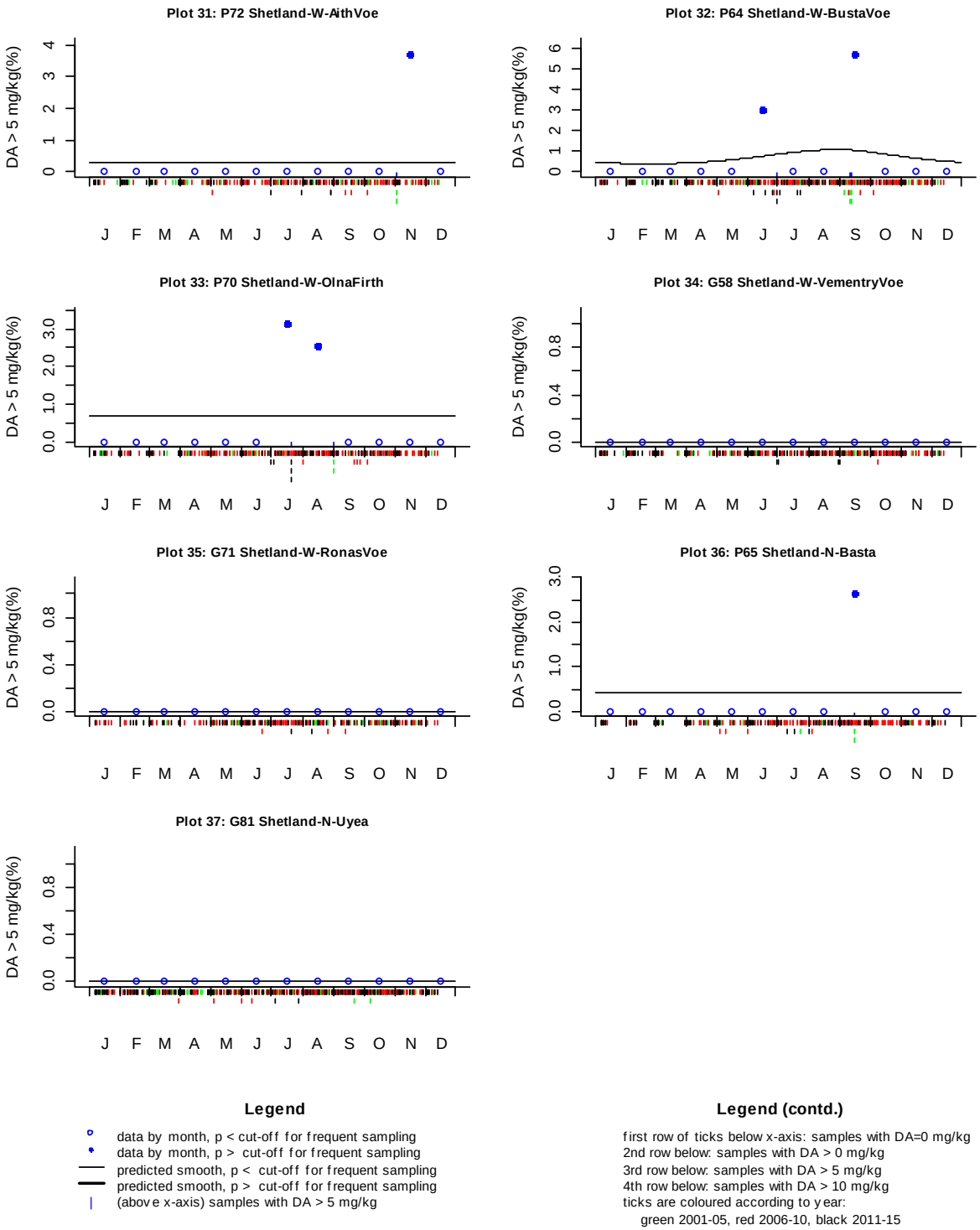


Figure D3 continued

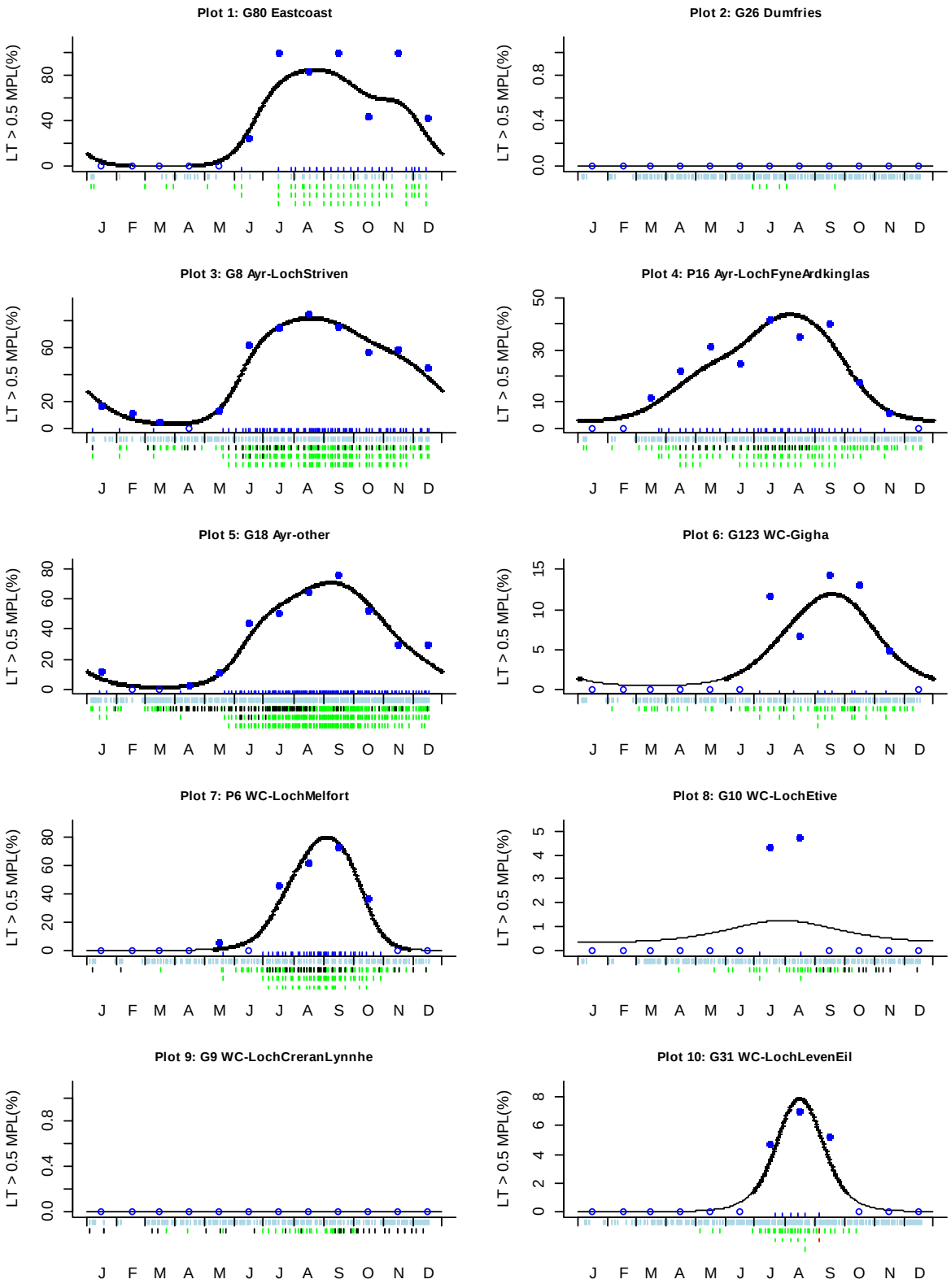


Figure D4: For each (group of) pod(s) the prevalence of LT LCMS > 0.5 MPL is shown, based on data (blue circles) from 2011-15, and predicted prevalence for an average toxin year from smooth models (black curve). Bold lines and closed symbols indicated is where monthly sampling is insufficient (prevalence > cut-off and fortnightly or weekly sampling would be required to keep the risk of non-detection below 1%). The x-axis shows the days of the year. Ticks above the x-axis are samples for which LT > 0.5 MPL. Ticks below the x-axis indicate the following: first row of ticks: samples for which LT = 0 mg/kg, second row of ticks: 0 < LT < 0.5 MPL, the third row of ticks: 0.5 ≤ LT < MPL, and the fourth row of ticks: LT ≥ MPL. The colouring of these ticks is according to the type of LT toxin: green OA, red AZA, black YTX.

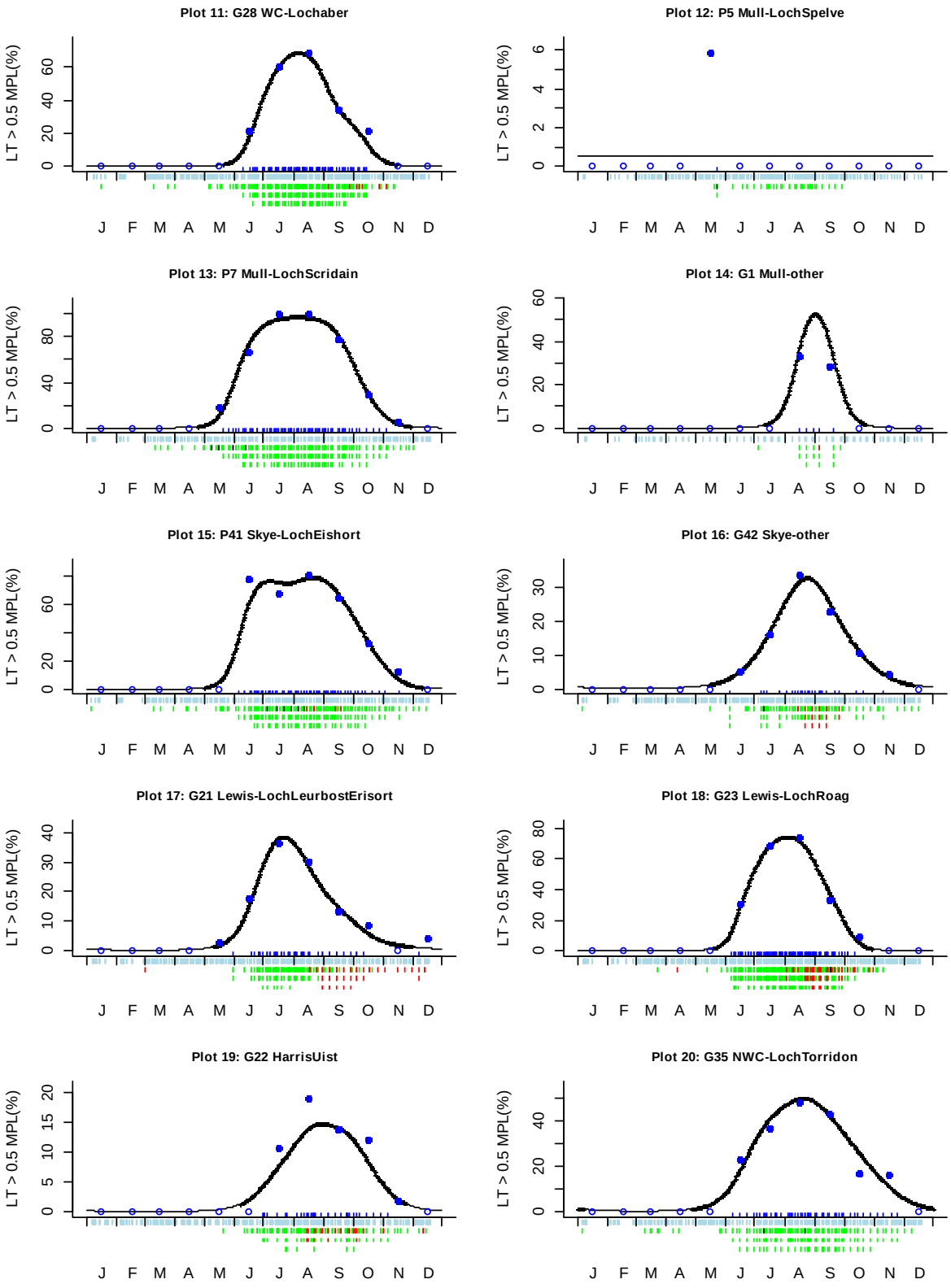
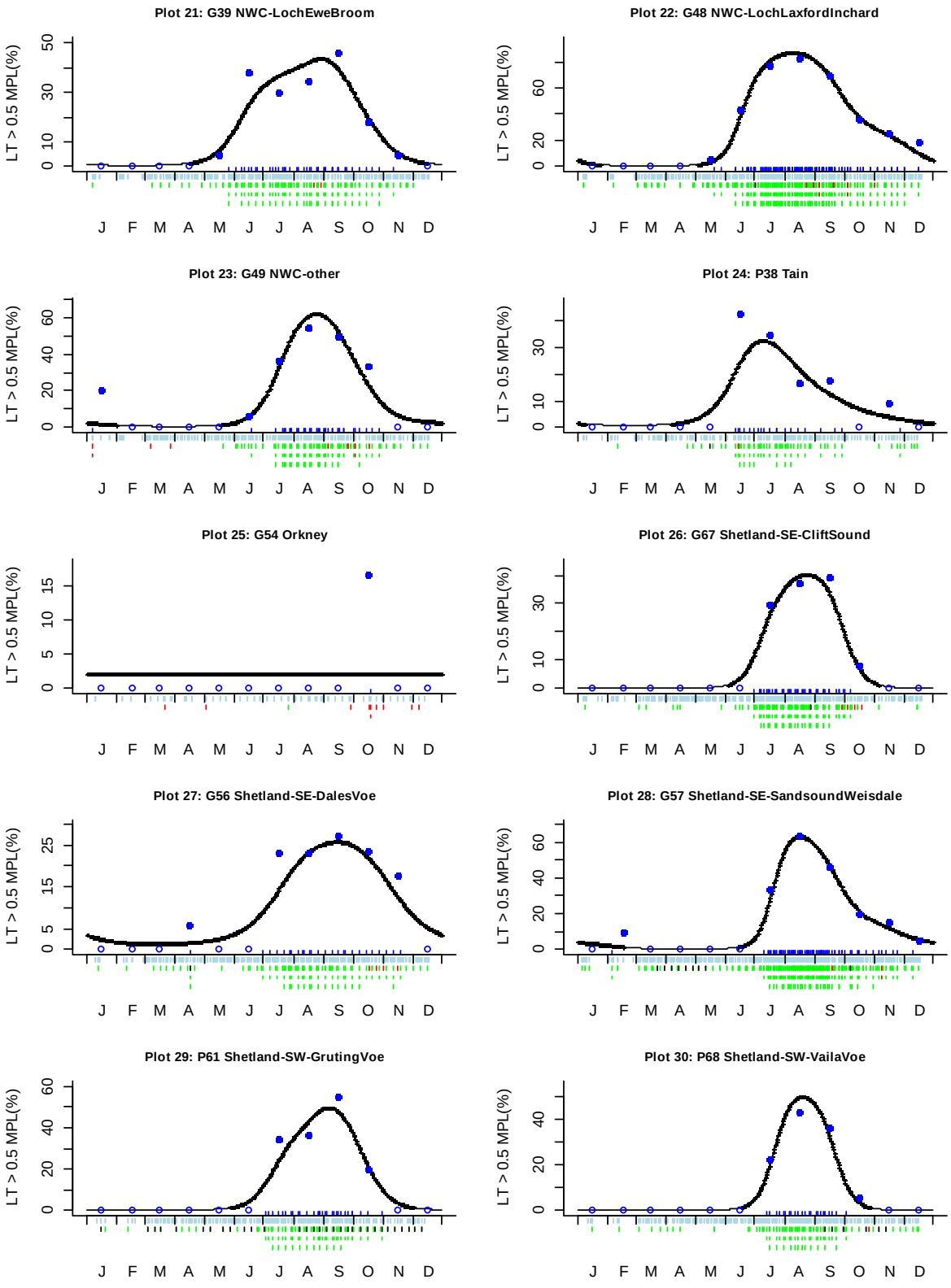


Figure D4 continued



FigureD4 continued

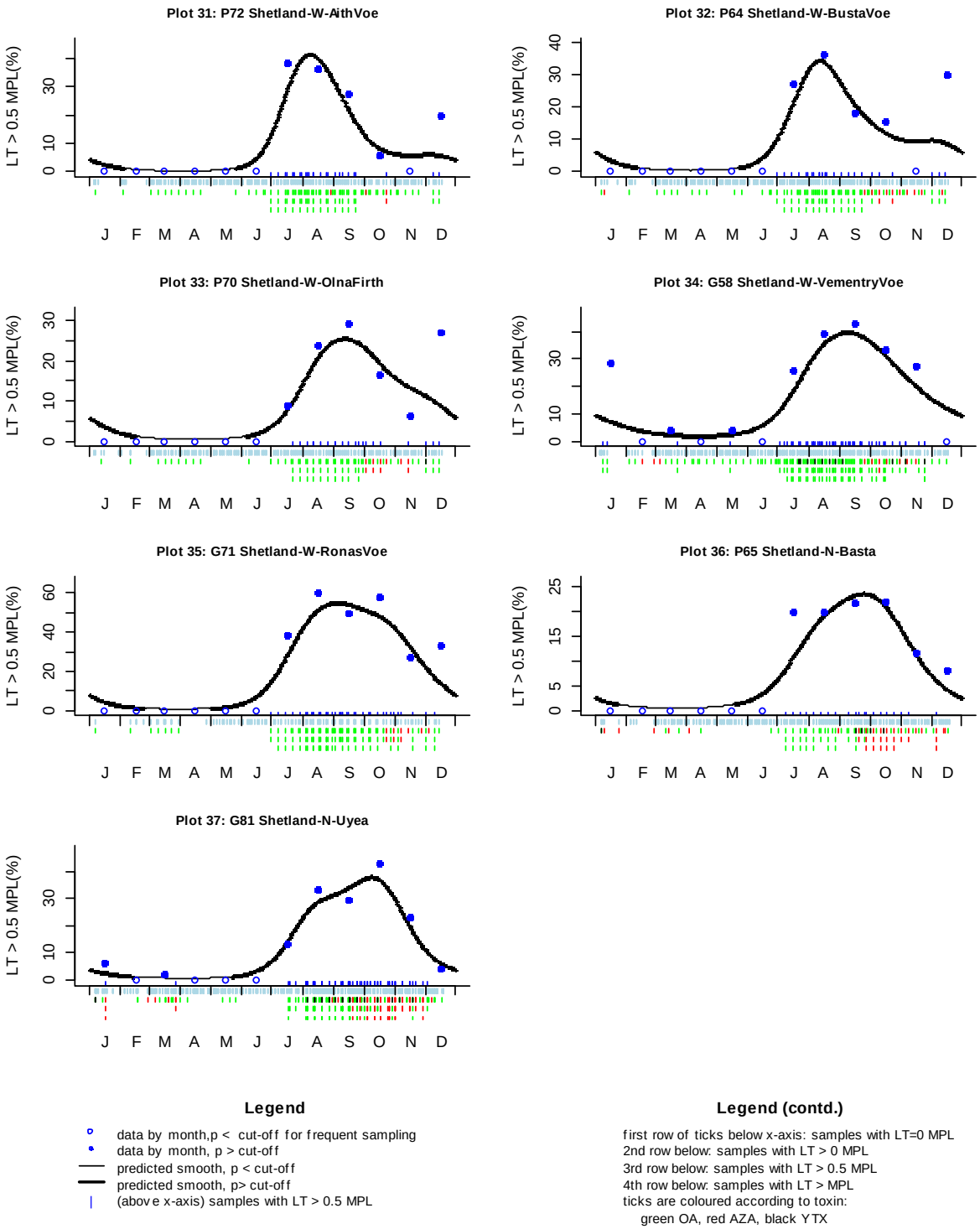


Figure D4 continued

Appendix E: Smooth model fits for biotoxin levels of secondary interest (PST > 0, PST > MPL, DA > 0, LT LCMS > MPL)

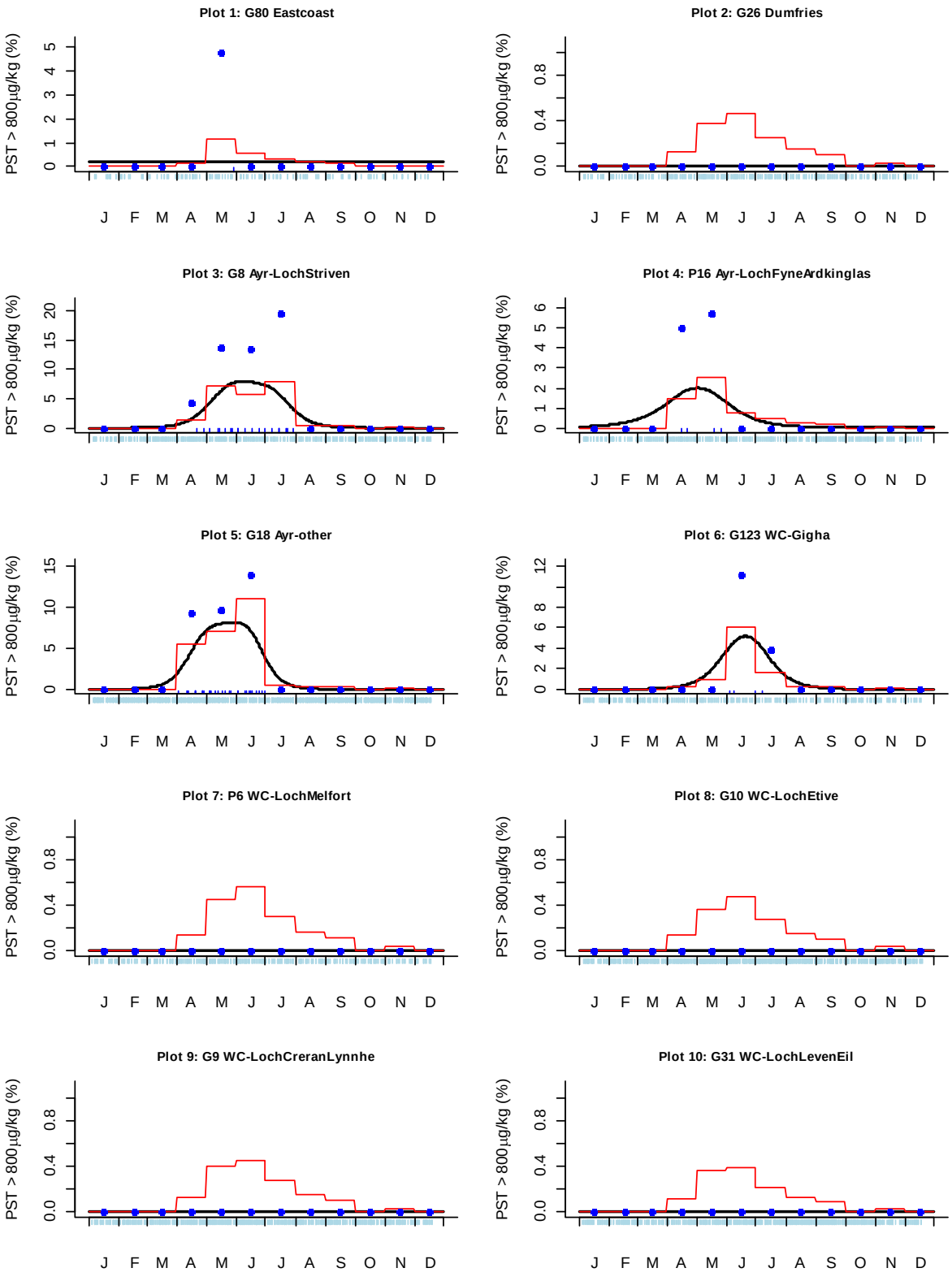


Figure E1: For each (group of) pod(s) the prevalence of PST > 800 µg/kg is shown, based on data (blue circles) from 2001-15, from simple models where predicted prevalence for an average toxin year is constant for each month (red lines, based on Holtrop et al. 2016), and predicted prevalence for an average toxin year from smooth models (black curve). The x-axis shows the days of the year with ticks indicating when samples were obtained and whether their test results was below (light blue) or above (blue) 800 µg/kg.

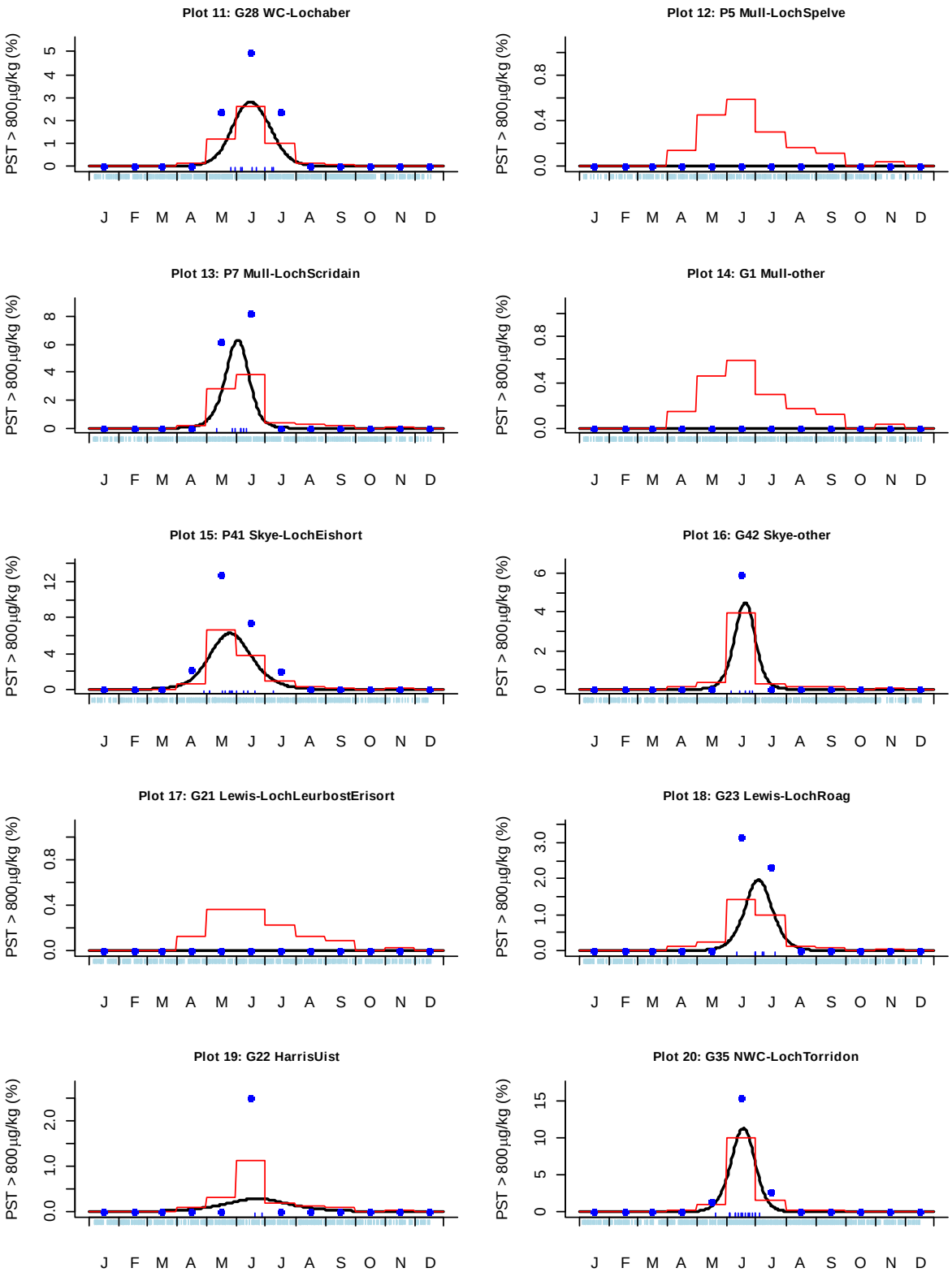


Figure E1 continued

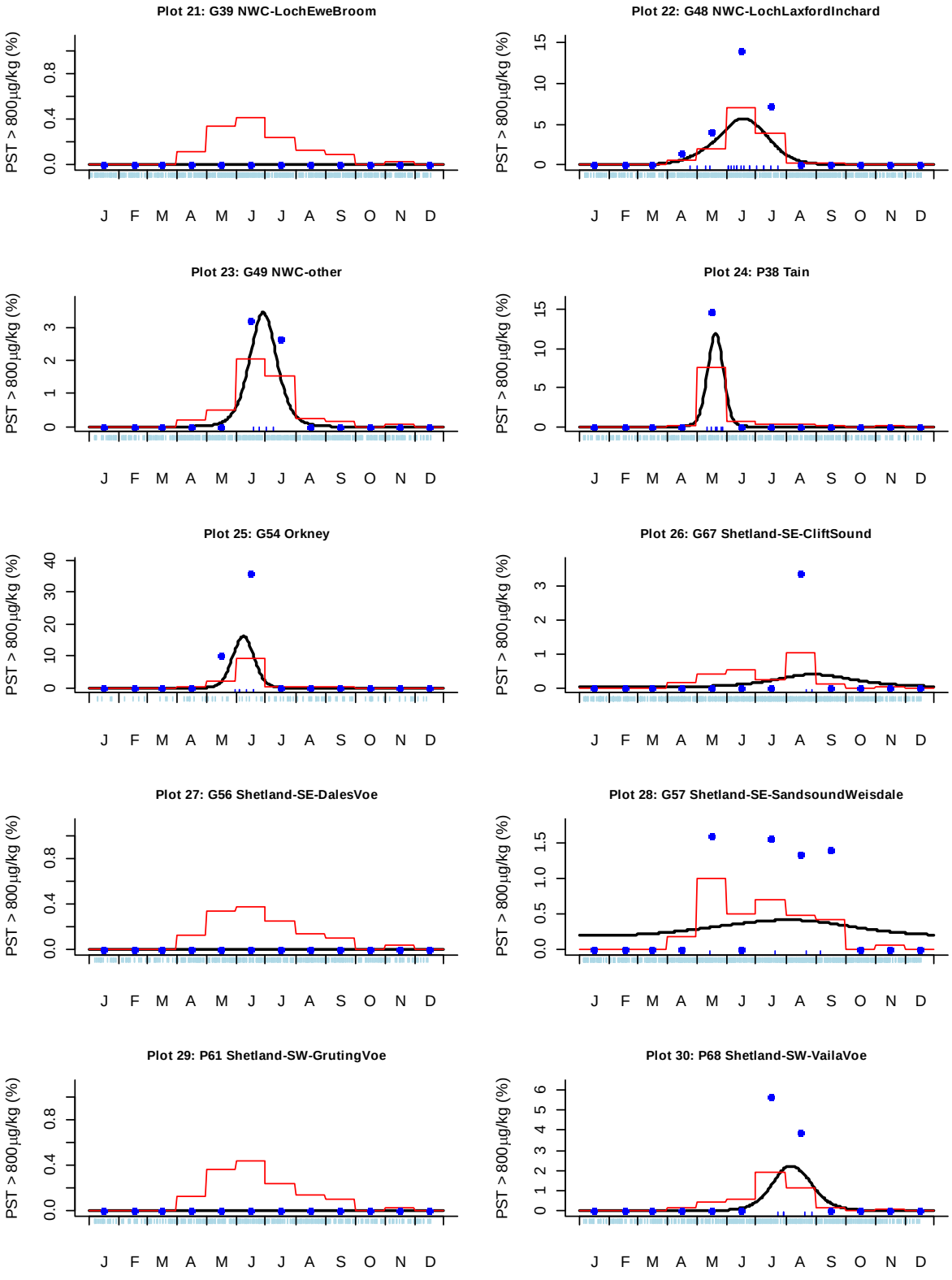


Figure E1 continued

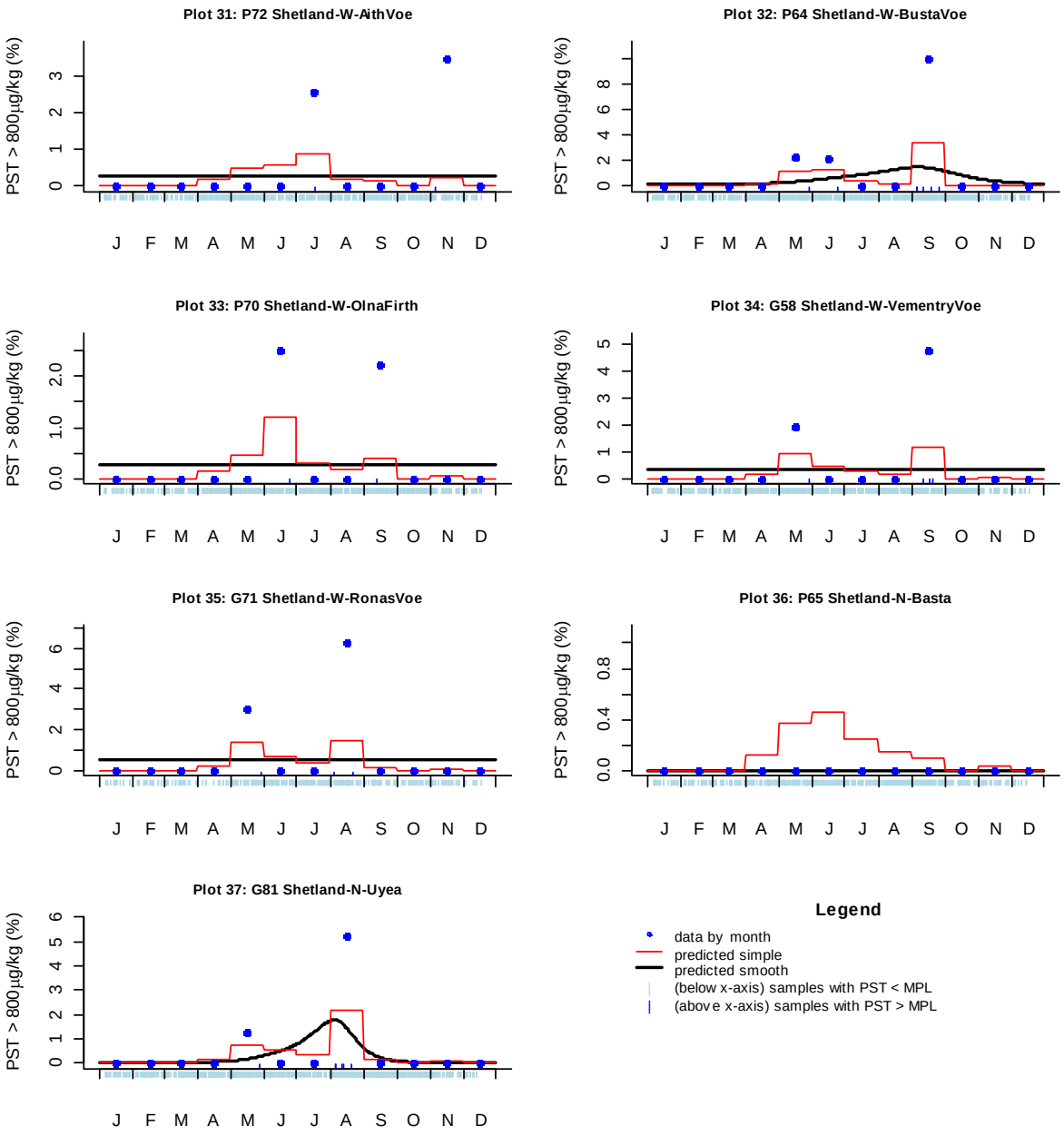


Figure E1 continued

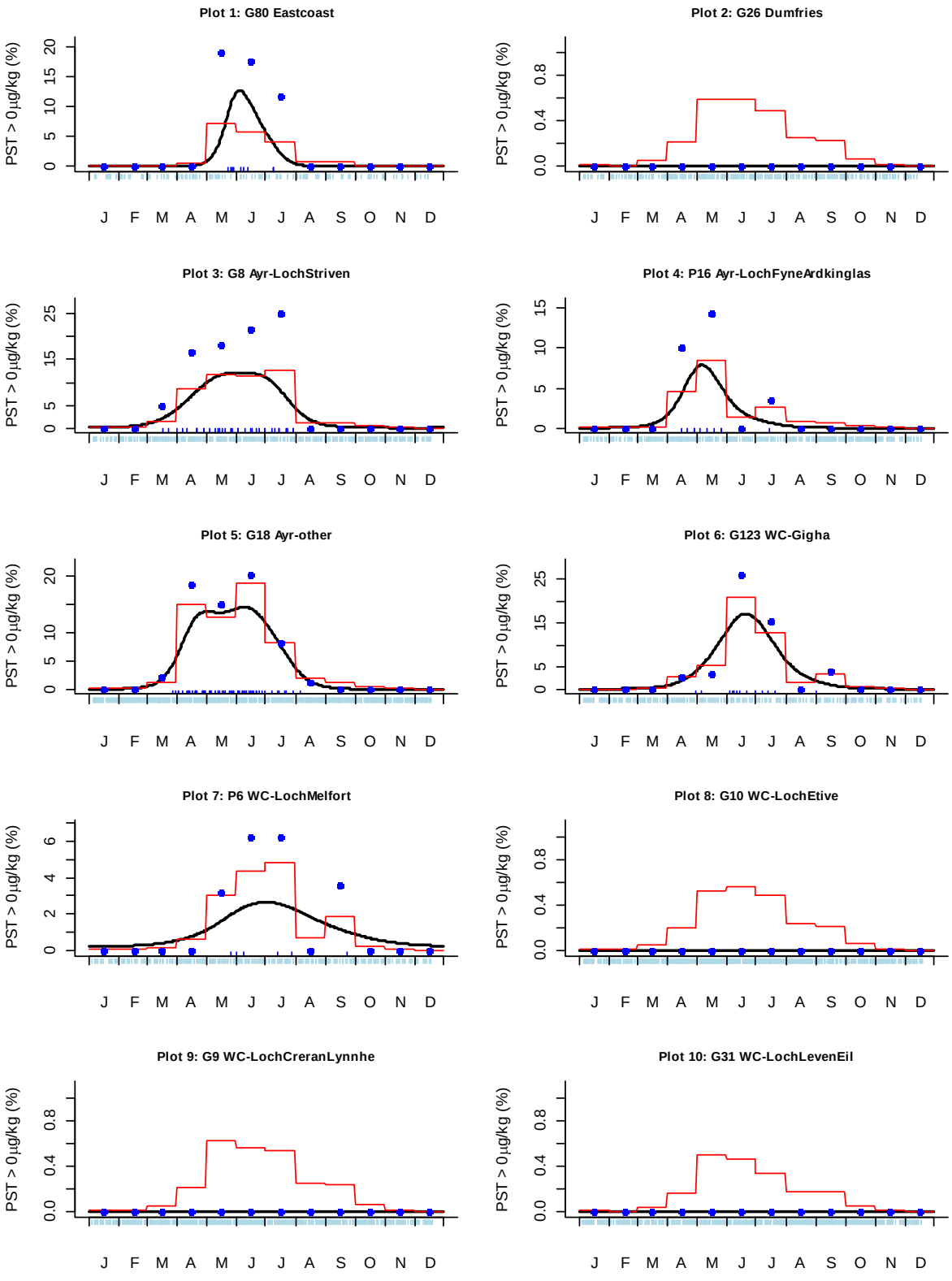


Figure E2: For each (group of) pod(s) the prevalence of PST > 0 µg/kg is shown, based on data (blue circles) from 2001-15, from simple models where predicted prevalence for an average toxin year is constant for each month (red lines, based on Holtrop et al. 2016), and predicted prevalence for an average toxin year from smooth models (black curve). The x-axis shows the days of the year with ticks indicating when samples were obtained and whether their test results was equal to (light blue) or above (blue) 0 µg/kg.

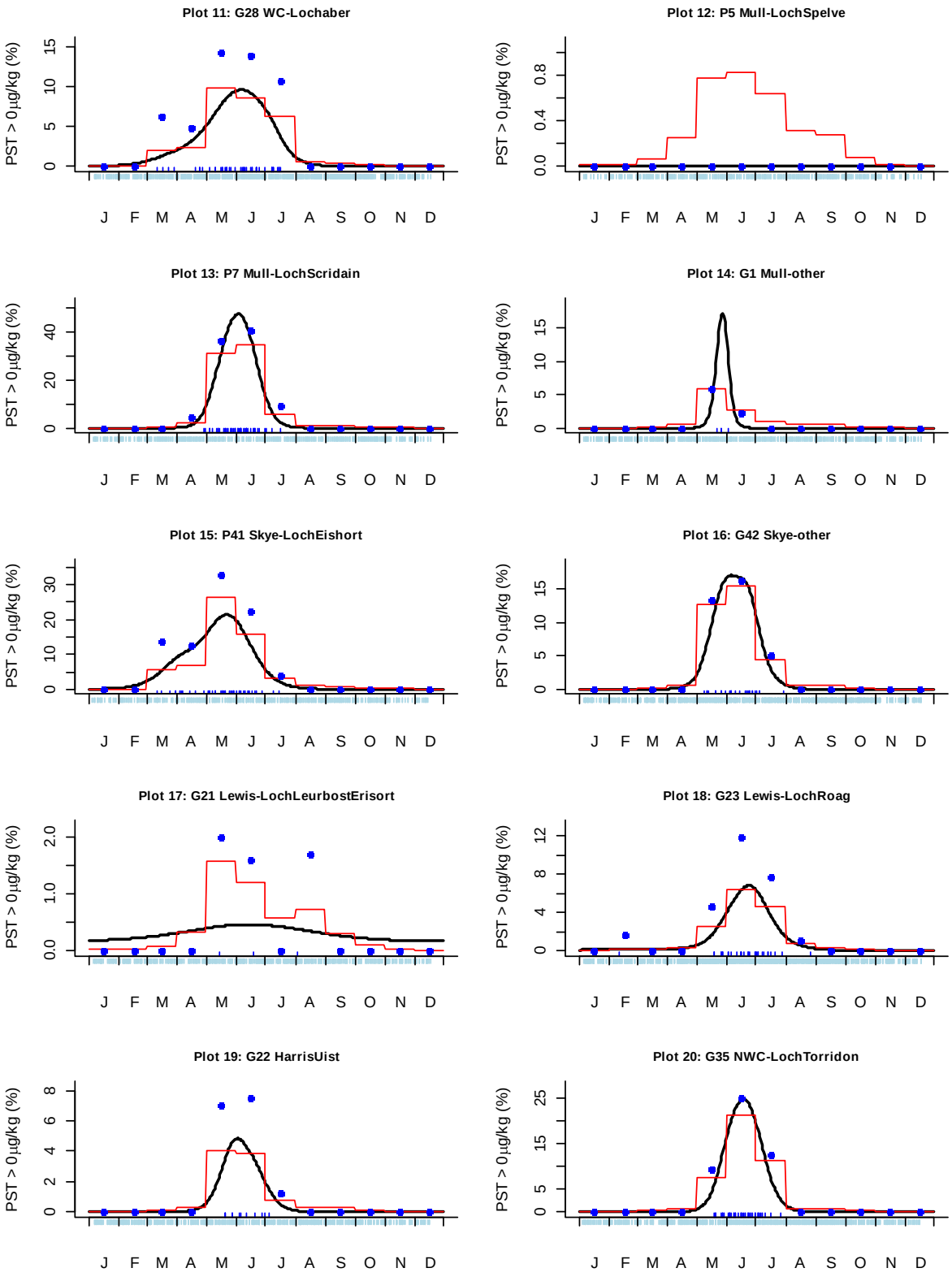


Figure E2 continued

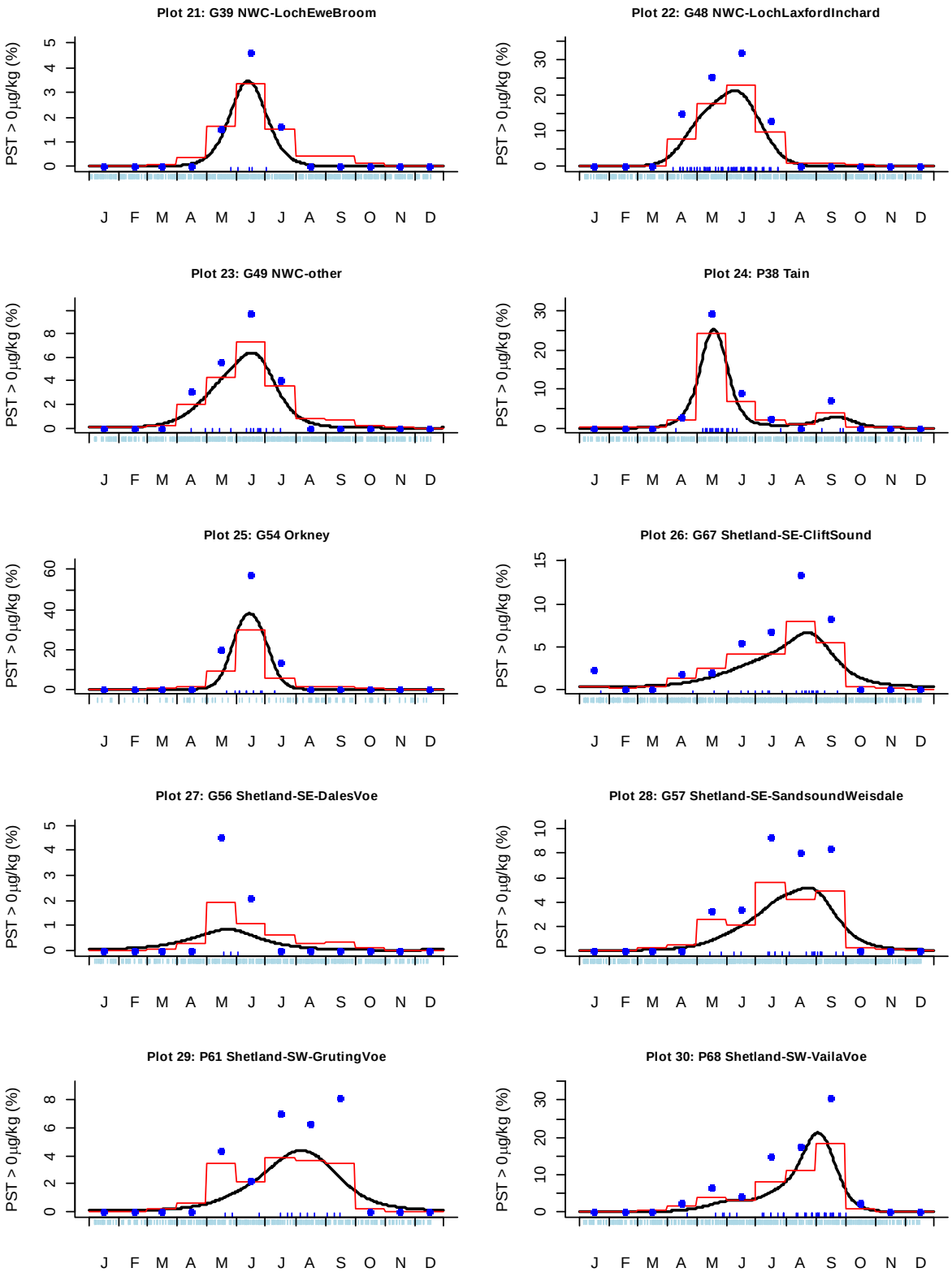


Figure E2 continued

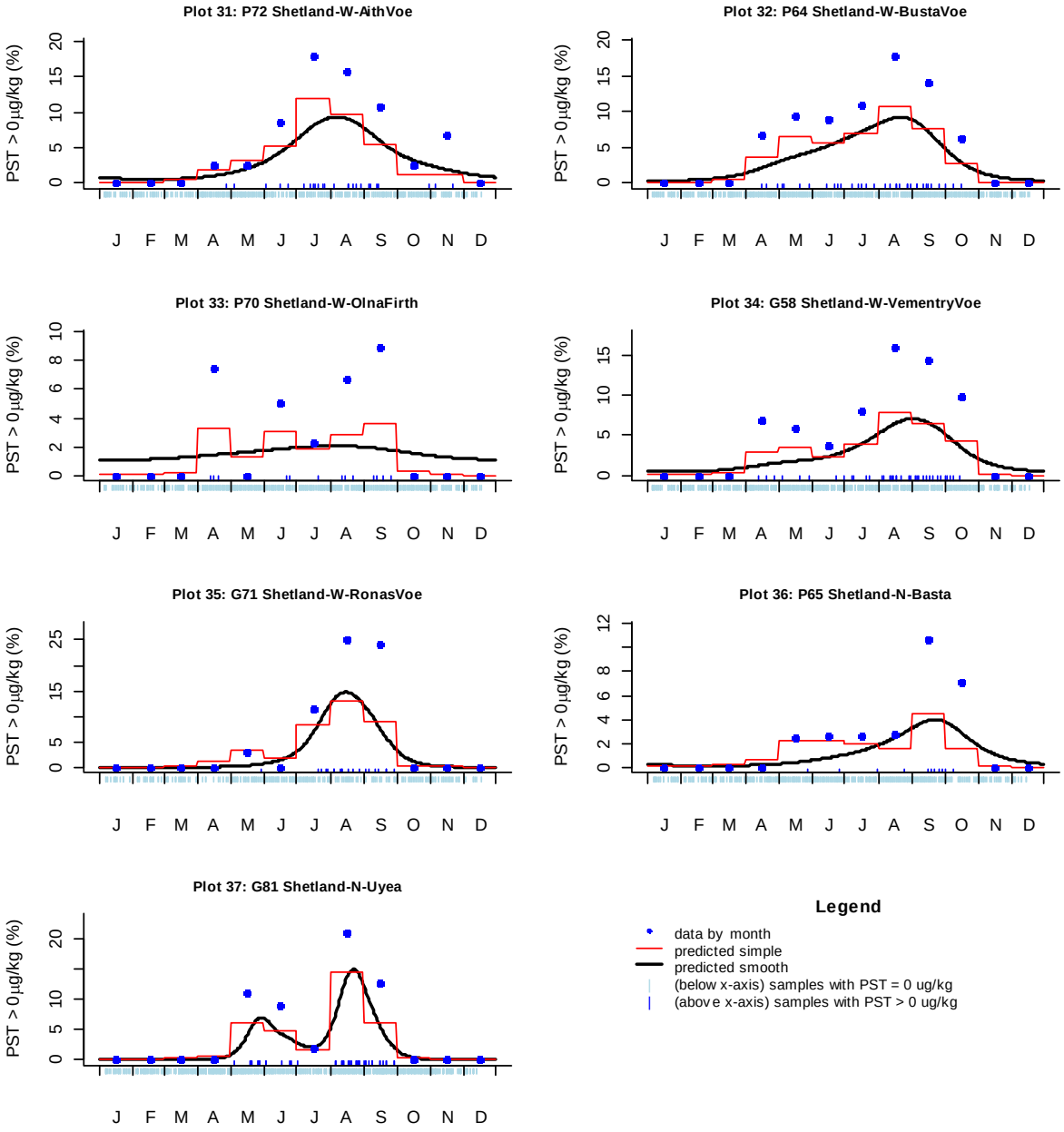


Figure E2 continued

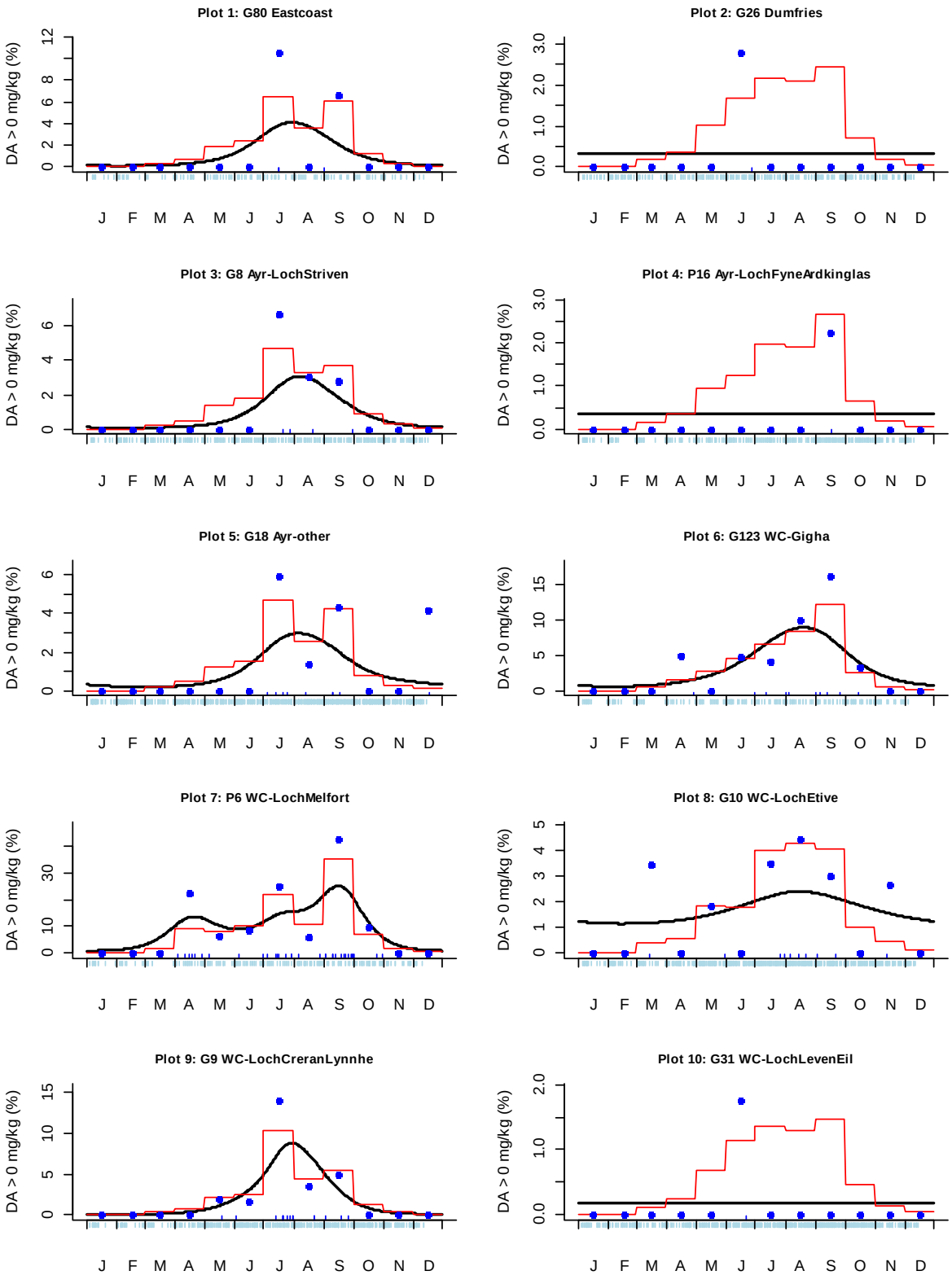


Figure E3: For each (group of) pod(s) the prevalence of DA > 0 mg/kg is shown, based on data (blue circles) from 2001-15, from simple models where predicted prevalence for an average toxin year is constant for each month (red lines, based on Holtrop et al. 2016), and predicted prevalence for an average toxin year from smooth models (black curve). The x-axis shows the days of the year with ticks indicating when samples were obtained and whether their test results are equal to (light blue) or above (blue) 0 mg/kg.

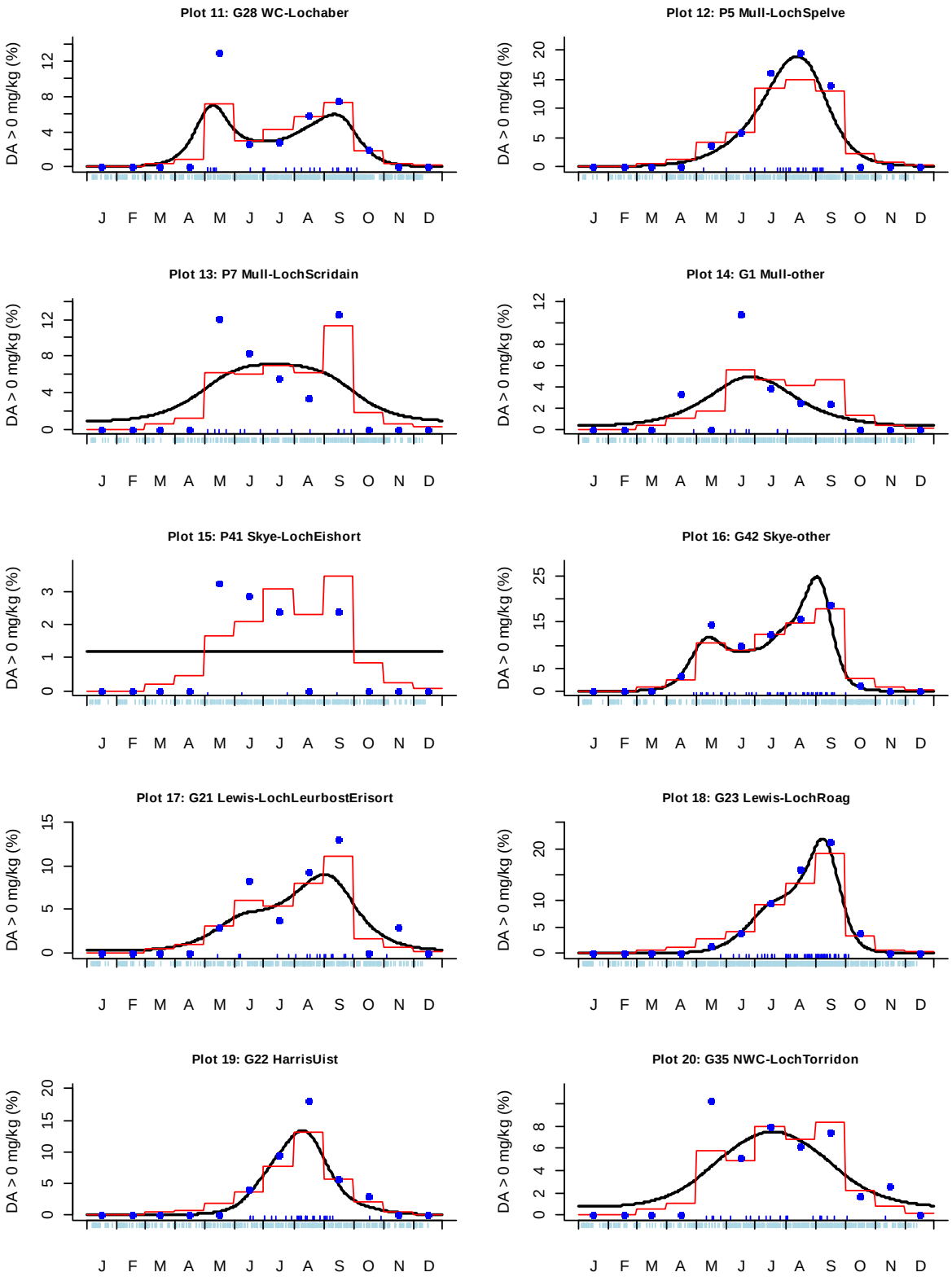


Figure E3 continued

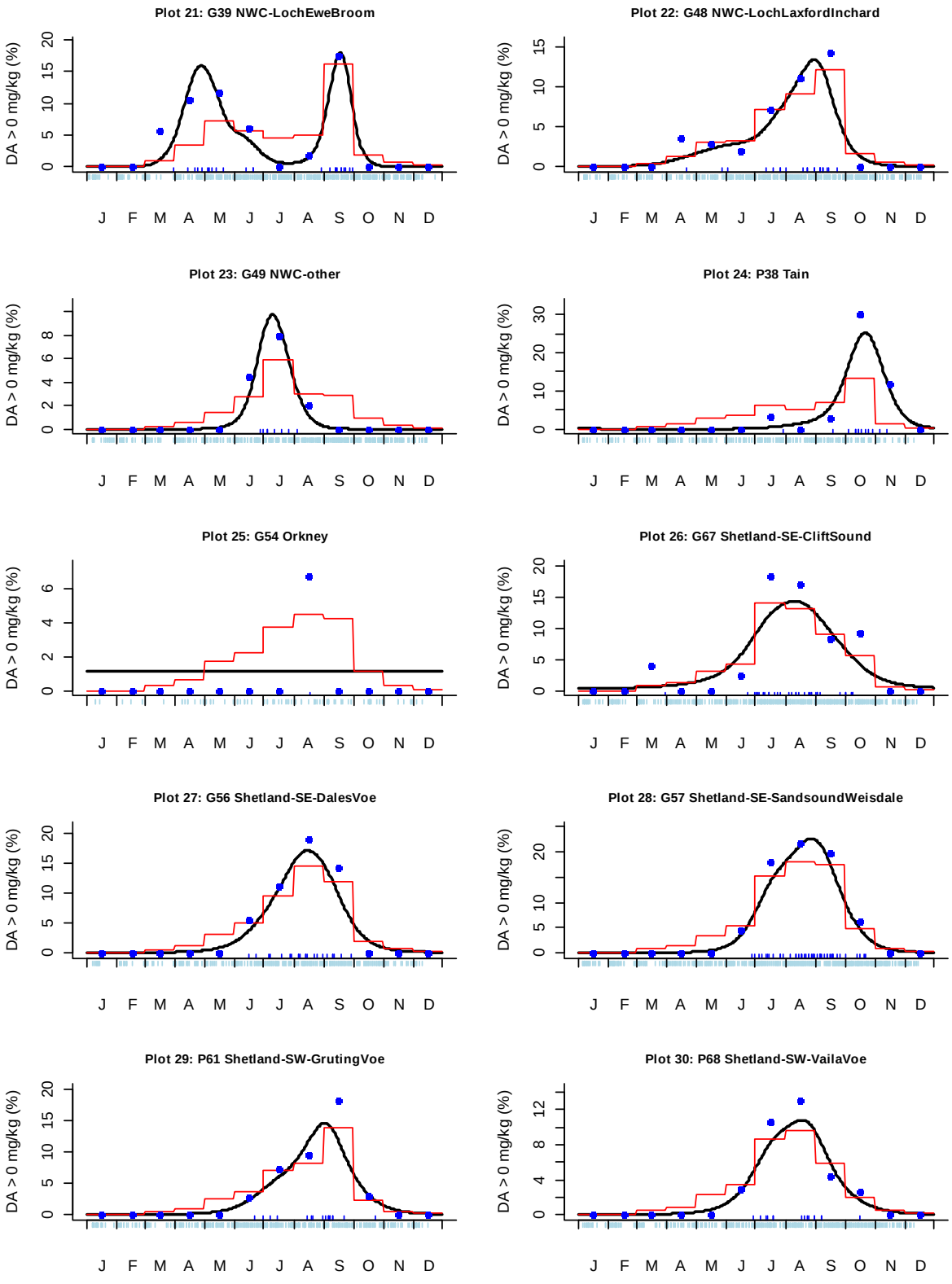


Figure E3 continued

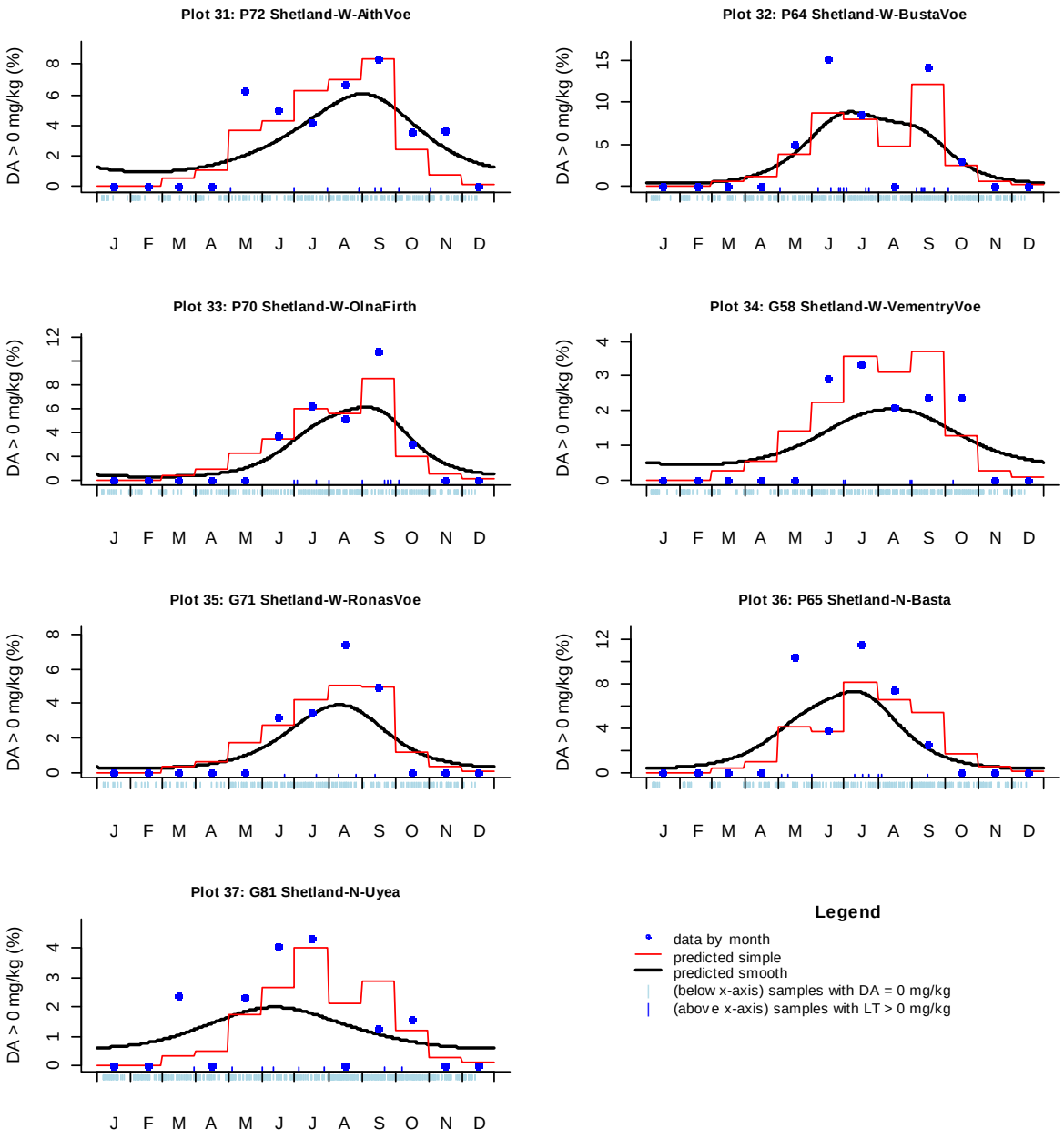


Figure E3 continued

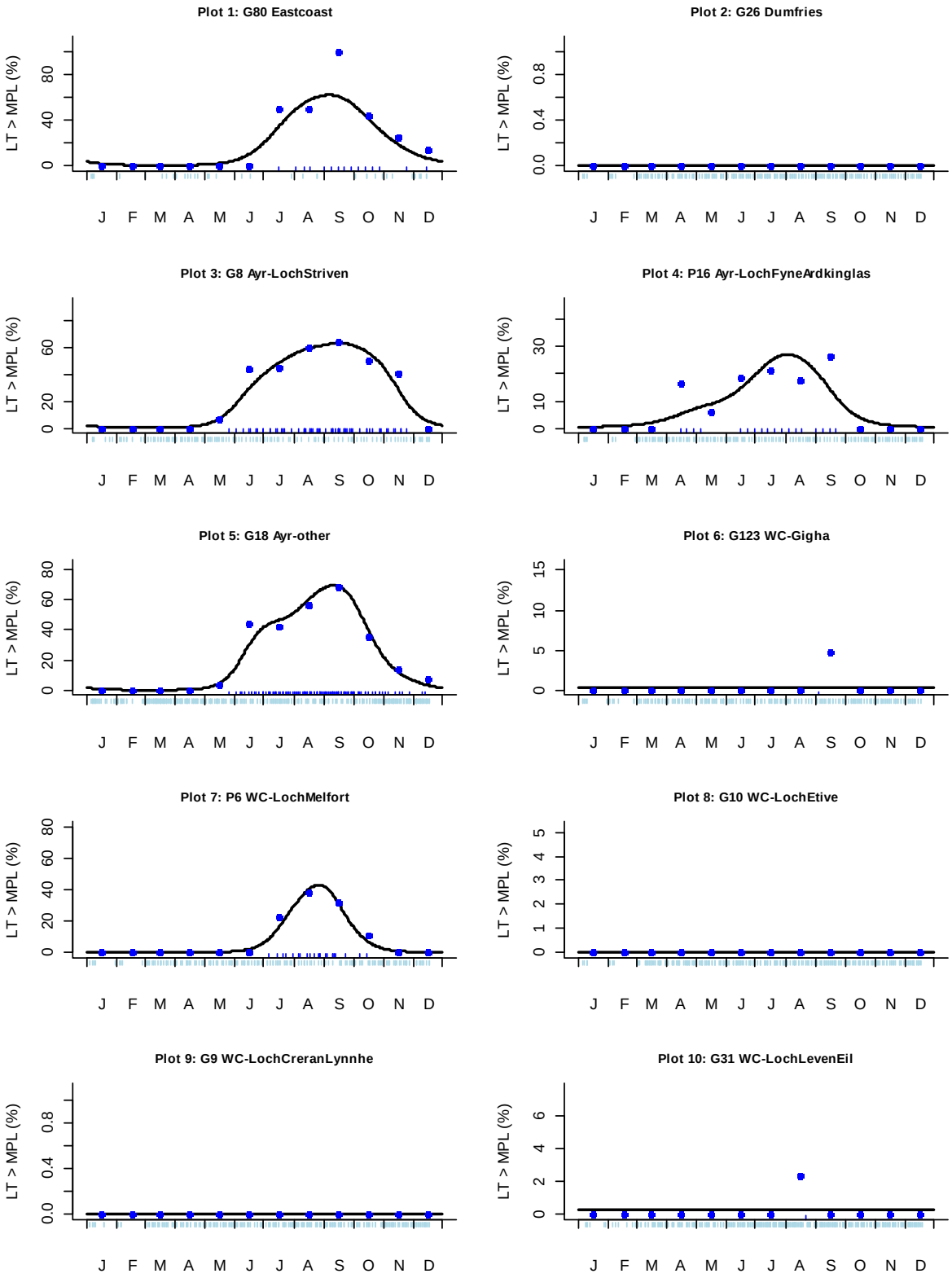


Figure E4: For each (group of) pod(s) the prevalence of LT-LCMS > MPL is shown, based on data (blue circles) from 2011-15, and predicted prevalence for an average toxin year from smooth models (black curve). The x-axis shows the days of the year with ticks indicating when samples were obtained and whether their test results was below (light blue) or above (blue) MPL.

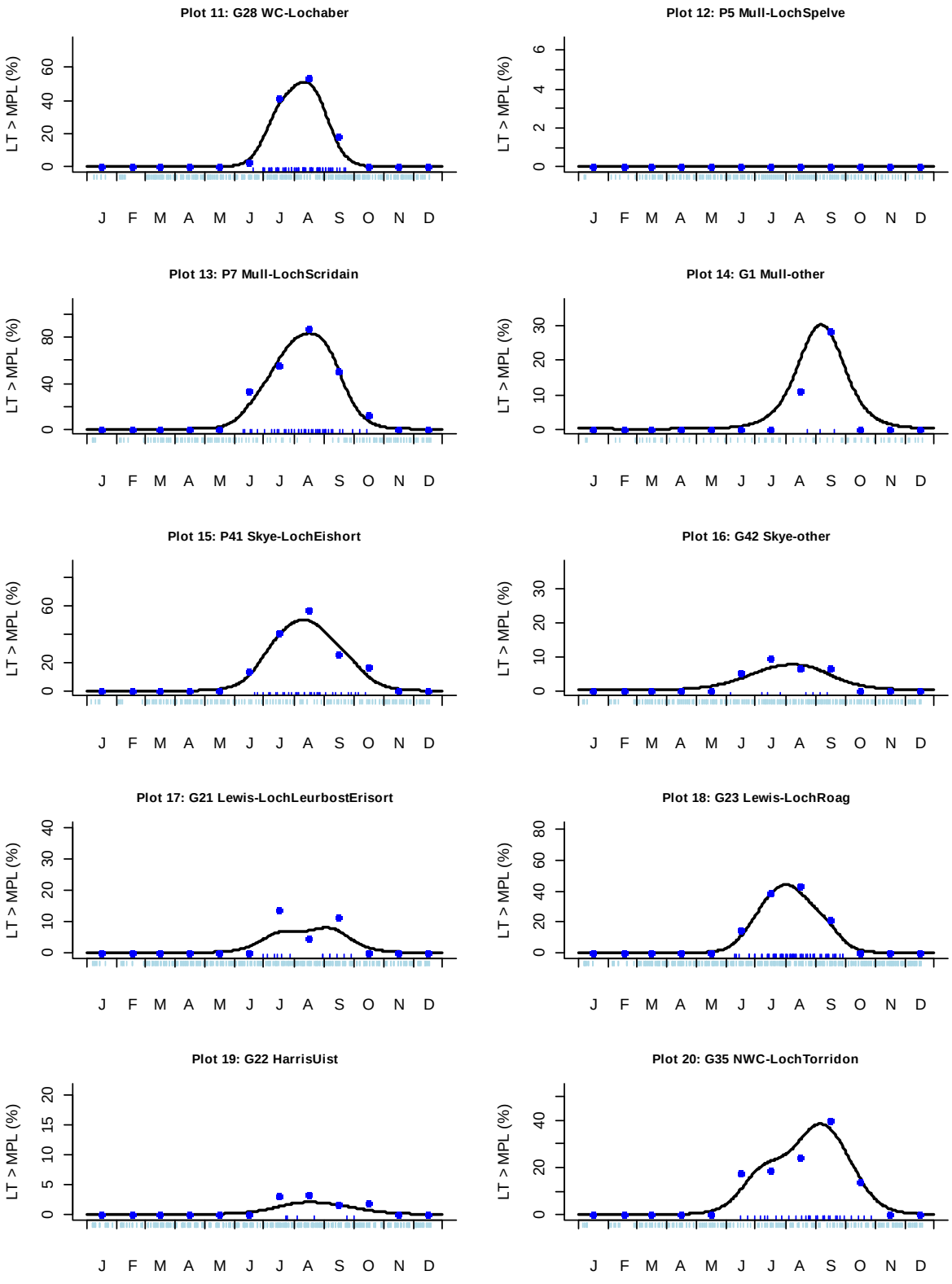


Figure E4 continued

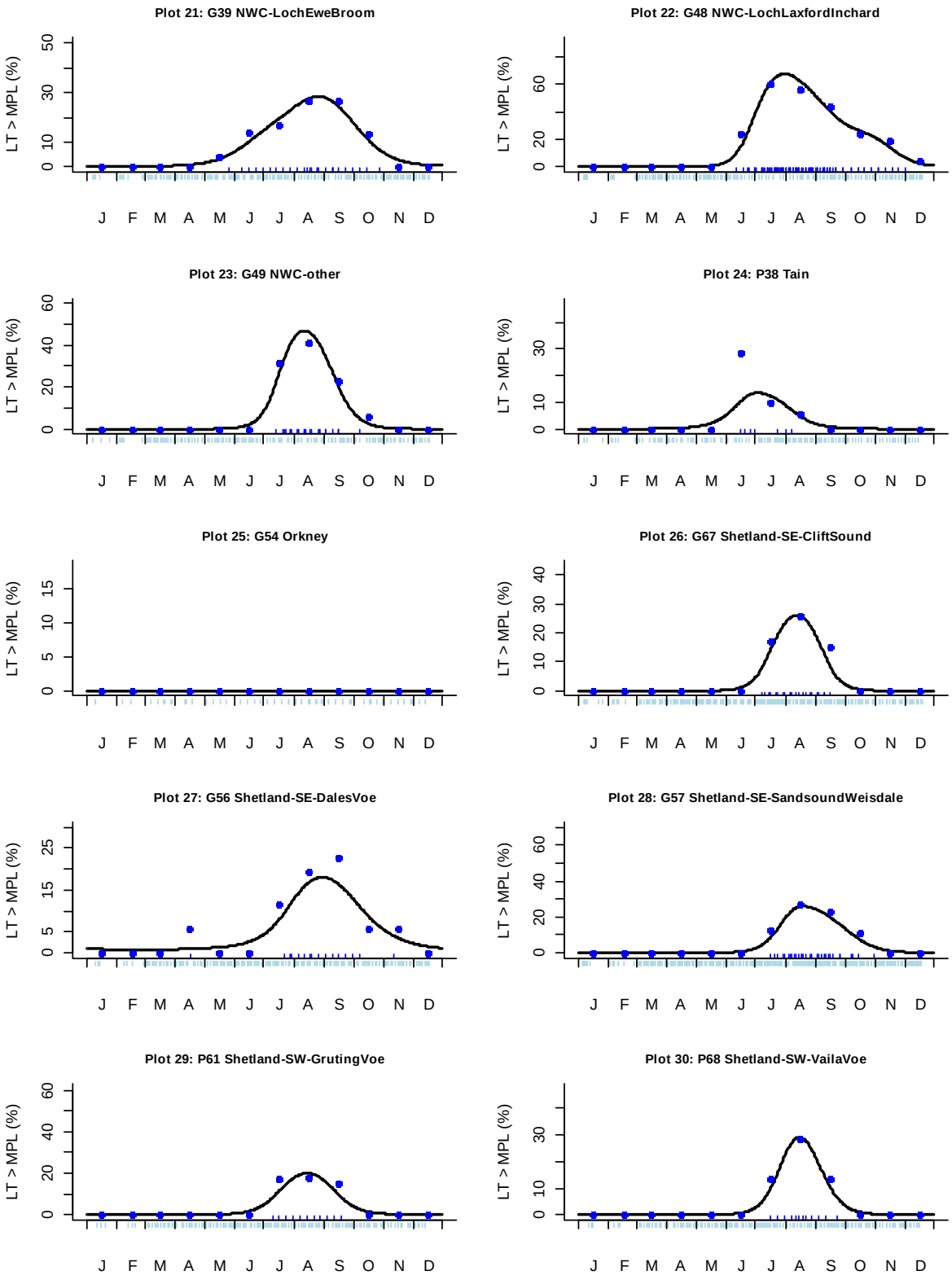


Figure E4 continued

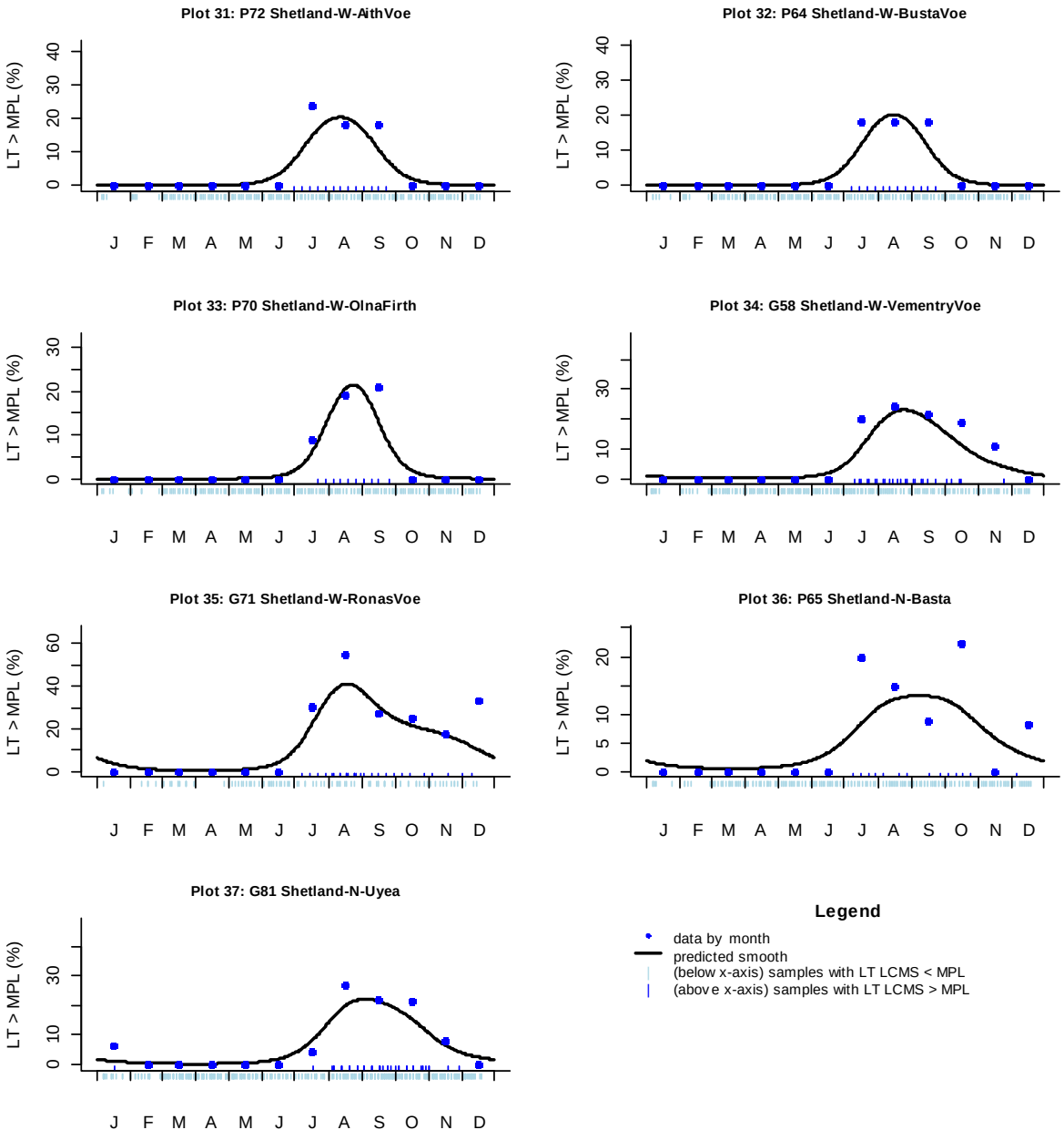


Figure E4 continued