Modelling the Growth Domain of *Clostridium botulinum* via Kernel Survival Analysis

Robert J. Foxall
School of Information Systems,
University of East Anglia,
Norwich NR4 7TJ, U.K.
rjf@sys.uea.ac.uk

Gavin C. Cawley
School of Information Systems,
University of East Anglia,
Norwich NR4 7TJ, U.K.
gcc@sys.uea.ac.uk

Michael W. Peck
Institute for Food Research,
Colney Research Park,
Norwich NR4 7UA, U.K.
Mike.Peck@bbsrc.ac.uk

***Abstract*— *Clostridium botulinum* is a bacterium present in the raw ingredients of many foods. It produces a powerful neurotoxin as part of its growth process, that can prove fatal when doses as small as 30ng are consumed. It is therefore vital to be able to accurately determine the food processing and storage conditions where toxin production is possible, known as the "growth domain". This paper describes a new approach to modelling the growth domain of microbial pathogens, by constructing a regularised kernel model relating heat treatment and subsequent incubation conditions to the parameters of a statistical distribution modelling the probability of growth as a function of incubation time. We demonstrate that the use of the "kernel trick" permits the extension of methods from classical survival analysis to account for non-linear dependencies in a principled manner.**

I. Introduction

Clostridium botulinum is an anaerobic bacterium that produces one of the most powerful toxins known to science as a by-product of its growth processes. Ingestion of only 30ng of the toxin can result in severe illness and even death [1]. It is therefore vital that steps should be taken to ensure that the toxin is not present in food. As *C. botulinum* spores are ubiquitous in raw ingredients, food must be processed to ensure that all of these spores are destroyed, or so that the spores are prevented from germinating, leading to cell division and subsequent toxin production. Growth of *C. botulinum* is, in most cases, principally dependent on environmental factors such as temperature, pH, NaCl concentration and gas atmosphere. It is important then to be able to define the conditions under which the spores are prevented from germinating, and giving rise to toxin production. This is especially true in the case of minimally processed chilled foods, as non-proteolytic *C. botulinum* is capable of growth and toxin production at chill temperatures. The safety of these foods with respect to non-proteolytic *C. botulinum* is likely to rely on a combination of heat treatment and subsequent incubation at refrigeration temperatures (Lund and Notermans [1], Peck [2]).

This paper describes a neural model of toxin production by *C. botulinum* in a meat-based medium containing lysozyme, following a range of heat treatments. We demonstrate that the familiar "kernel trick" can be used to construct non-linear equivalents of traditional linear parametric survival analysis models. The use of regularisation means that the bias-variance trade-off can be effectively controlled by a small set of regularisation parameters. This allows the use of powerful non-linear kernel functions without a significant risk of over-fitting.

The remainder of this paper is structured as follows: Section II gives an overview of classical parametric survival analysis techniques. Section III describes a dataset covering the growth domain of *C. botulinum* used in the comparison of kernel and existing growth domain models. Sections IV - VI, describe the construction of growth domain models, using a growth domain model representative of current practice, conventional linear survival analysis and kernel survival analysis methods. Finally the work is summarised in section VII.

II. Parametric Survival Analysis

Survival analysis (see e.g. [3]) is a field of classical statistics concerned with data recording the time that elapses before the occurrence of each of a set of point events, known as "failures". Many applications of survival analysis arise in medical studies, for example it might be of interest to model the length of time that patients survive following each of a range of competing treatments for a given illness. Survival analysis requires there to be a well-defined origin at which time $t = 0$, an appropriate scale for measuring time and a unambiguous definition of failure. In the case of our medical example, the origin is defined by the time of treatment, the time scale is measured in days following treatment and failure defined by the patient's death. Parametric survival analysis therefore aims to determine the optimal parameters of a fixed distribution describing time to failure, T,

$$F(t) = Pr(T \geq t). \tag{1}$$

Note this differs from the usual statistical convention where cumulative distribution functions are defined in terms of right continuity, i.e. $F(x) = Pr(X \leq x)$, and hence the probability density function is defined as,

$$f(t) = -F'(t). \tag{2}$$

Another function of interest in survival analysis is the *hazard* function, given by

$$h(t) = \frac{f(t)}{F(t)}, \tag{3}$$

which represents the instantaneous probability of failure at time t, given survival until time t. Any parametric distribution over non-negative values of t may be used (in most applications it does not make sense to consider failure before time $t = 0$). Table I shows a number of statistical distributions commonly used in parametric survival analysis. The optimal

parameters of the survival distribution are traditionally found via a maximum likelihood approach. Given a dataset $\mathcal{D} = \{t_i\}_{i=1}^{\ell}$, recording the failure time for ℓ events, then assuming that the data represents an independent and identically distributed (i.i.d.) sample from some underlying distribution, then the *likelihood* of the data is given by the product of the density function over the observed data, i.e.

$$L = \prod_{i=1}^{n} f(t_i). \tag{4}$$

The optimal parameters are then determined by minimising the negative logarithm of the likelihood function.

TABLE I

STATISTICAL DISTRIBUTIONS COMMONLY ENCOUNTERED IN PARAMETRIC SURVIVAL ANALYSIS.

Distribution	**Density Function** $f(t)$	**Survivor Function** $F(t)$
Exponential	$\frac{1}{b}\exp\left\{-\frac{t}{b}\right\}$	$\exp\left\{-\frac{t}{b}\right\}$
Weibull	$\left(\frac{\beta t^{\beta-1}}{\eta^{\beta}}\right)\exp\left\{-\left(\frac{t}{\eta}\right)^{\beta}\right\}$	$\exp\left\{-\left(\frac{t}{\eta}\right)^{\beta}\right\}$
Log-logistic	$\frac{\exp\left[\frac{(\log t-a)}{b}\right]}{bt\left\{1+\exp\left[\frac{(\log t-a)}{b}\right]\right\}^2}$	$\frac{1}{1+\exp\left[\frac{(\log t-a)}{b}\right]}$
Log-normal	$\frac{1}{\sqrt{2\pi}\sigma t}\exp\left\{-\frac{1}{\sigma^2}(\log t-\mu)^2\right\}$	$\int_t^{\infty} f(w)dw$
Gamma	$\frac{\lambda^r}{\Gamma(r)}t^{r-1}e^{-\lambda t}$	$\int_t^{\infty} f(w)dw$

A. Censoring of Data in Parametric Survival Analysis

In many applications of survival analysis it is not practical to observe every trial until failure occurs, and instead a fixed observation period is imposed. Trials where failure is not observed are said to have been "censored". Returning to our medical example, it is only to be expected that not all of the patients will have died by the time the observation period has finished (a period of 5 years is commonly used in defining survival rates for medical procedures), other patients may have died from totally unrelated causes, for instance road traffic accidents, or simply may have moved away and are no longer in contact with the medical institution conducting the study. Clearly, even though the failure time is unknown, censored data should still be included in fitting the survival distribution, as they provide information on an interval of time where failure was not observed. It is a simple matter to incorporate censoring into the likelihood function: Uncensored data are handled as before, for censored data, however, all that is known is that the failure will occur at some time greater than the censoring time, so the likelihood for censored observations is given by $F(t)$. The likelihood function then becomes

$$L = \prod_{i\in\mathcal{U}} f(t_i) \prod_{i\in\mathcal{C}} F(t_i). \tag{5}$$

where $\mathcal{U}$ and $\mathcal{C}$ represent the index sets of uncensored and censored observations respectively. Note that the censoring time may vary for each observation, or may be constant for all trials.

B. An Illustrative Example

Consider an experiment where spores of *C. botulinum* were introduced into five vials containing nutrient media and subjected to identical heat treatment and incubation conditions. The observed time to growth for these five replicates were $15, 25, 30, 30$ and 38 days. Figure 1 shows exponential, Weibull, log-logistic and log-normal distributions fitted using a maximum likelihood approach. Subjectively, the exponential model clearly does not fit the observed data as well as the other three distributions investigated. This is perhaps unsurprising as the exponential distribution is defined by a single parameter and hence is less flexible than the Weibull, log-logistic, or log-normal distributions, each of which are defined by two parameters. The exponential model is appropriate for survival data where there is a constant hazard rate, i.e. the probability of failure is is independent the amount of time the subject has already survived. An objective measure of goodness of fit is given by the log-likelihoods for the four models (given in table II), which also demonstrates the superiority of the log-normal, log-logistic and Weibull distributions over the exponential distribution.

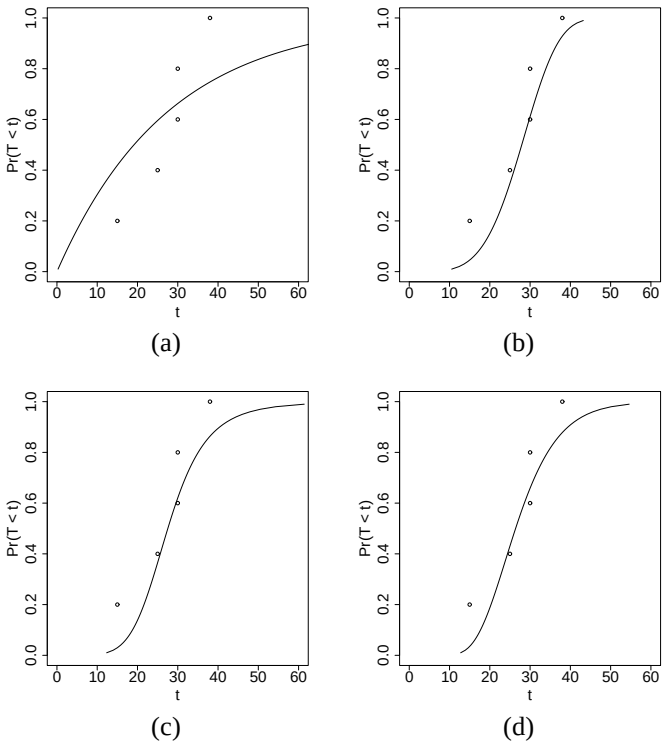


Fig. 1. Maximum likelihood survival models of example dataset using exponential (a), Weibull (b), log-logistic (c) and log-normal (d) survivor functions.

TABLE II

LOG-LIKELIHOODS FOR THE EXAMPLE DATASET.

Model	Log-likelihood
Exponential	-21.6
Weibull	-17.1
Log-logistic	-17.7
Log-normal	-17.6

C. Dealing with Explanatory Variables

In most applications of survival analysis, we seek to construct a model capturing the relationship between survival times and the values of a set of d explanatory variables, $\{\boldsymbol{x}_i \in \mathbb{R}^d\}_{i=1}^{\ell}$. In the case of linear parametric survival analysis, a linear combination of the explanatory variables is used to provide conditional estimates for the parameters of the survivor function $F(t)$, for example the Weibull distribution has shape and scale parameters β and η respectively, so the linear survival model is given by

$$\beta(\boldsymbol{x}) = \langle \boldsymbol{w}_\beta \cdot \boldsymbol{x} \rangle + b_\beta \quad \text{and} \quad \eta(\boldsymbol{x}) = \langle \boldsymbol{w}_\eta \cdot \boldsymbol{x} \rangle + b_\eta, \quad (6)$$

where the model parameters $\boldsymbol{w}_\beta, b_\beta, \boldsymbol{w}_\eta$ and b_η are found by minimising the likelihood function (5).

III. THE DATASET

The growth domain models for *C. botulinum* described here are based on the dataset described in Fernández and Peck [4]. Tubes containing a sterile meat-based medium containing lysozyme, an enzyme found to increase the measured heat resistance of spores of non-proteolytic *C. botulinum* (Lund and Peck [5], Peck [2]), were inoculated with a suspension of the spores of eight strains of non-proteolytic *C. botulinum*, at a final concentration of 10^6 spores per tube, and subjected to a range of heat treatments, shown in table III. The tubes were then cooled and incubated at temperatures of 5, 8, 12, 16 and 25°C for 90 days. Five replicates were performed at each incubation temperature, for each heat treatment regime. The tubes were inspected every 2–3 days for signs of growth, indicated by obvious formation of gas. At the end of the experiment, samples from each heat treatment regime, showing growth at the lowest incubation temperature and for the highest incubation temperature that did not show growth, were tested for toxin (Peck *et al.* [6], Stringer *et al.* [7], Carlin and Peck [8]). This type of dataset is known as *time to growth* data, as the results are presented in terms of a table showing the number of days after which each tube showed signs of growth. Full details of the experimental method are recorded in Fernández and Peck [4].

IV. EXAMPLE OF CURRENT PRACTICE

The paper by Whiting and Oriente [9] gives a representative example of current practice in modelling growth domain of microbial pathogens, we adopt this method as a benchmark against which to evaluate the kernel survival analysis technique proposed in this paper. The Whiting-Oriente model is constructed in two stages: First, the probability of growth within each set of five replicates sharing a common heat treatment regime and incubation temperature are modelled using a logistic model,

$$P(t) = \frac{P_{\max}}{1 + e^{\kappa(\tau - t)}}, \quad (7)$$

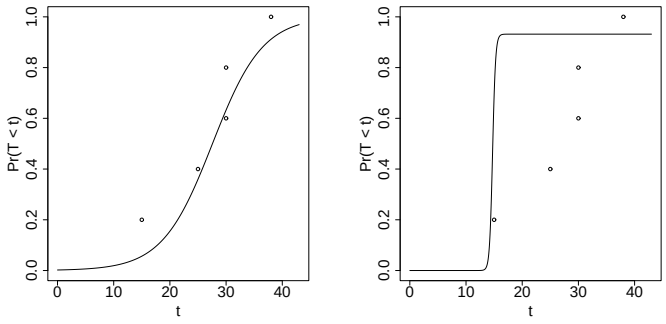
where t is the time in days, $P(t)$ is the probability of growth occurring by time t, $P_{\max}$ is the maximum probability of growth after the censoring time of 91 days, κ is scale parameter and τ is a location parameter. Following [9], the optimal values for $P_{\max}$, κ and τ for each of the 309 combinations of heat treatment regime and incubation conditions treatment conditions were determined by a least-square fitting procedure, minimising the squared difference between the model (7) and the empirical cumulative distribution function. Box constraints were imposed to ensure that $P_{\max}$ is constrained to lie between 0 and 1. In the second stage, a least-squares quadratic regression model is used to estimate the optimal values of $P_{\max}$, κ, and $\log(\tau)$ as a function of the explanatory variables (cooking time, incubation time and cooking temperature. This is implemented as a standard linear regression problem (the model including all main effects, squared main effects and all pairwise interaction terms). Backward elimination was then employed to remove terms not contributing significantly to the quality of the model fit.

Figure 2 (a) and (b) illustrate the major shortcoming of this approach, showing the stage one and stage two model fits for a set time to growth data for five test tubes under identical heat treatment and incubation conditions. The stage 1 model fit appears to be very good, however the stage two model fit is clearly very poor. This occurs because the second stage model fitting procedure is related to the data itself only via the parameters of the logistic models resulting from stage one. Indeed there is no guarantee that errors introduced by the stage two model fitting will not result in a non-zero probability of growth before time $t = 0$ or probabilities of growth outside the range $[0,\ 1]$. For example, out of 309 sets of replicates, on

TABLE III

HEAT TREATMENTS APPLIED TO A MEAT-BASED MEDIUM CONTAINING SPORES OF ***C. botulinum***

Temperature (°C)	Duration (min)		
70°C	104.9	529.1	998.9
	1596.3	2065.9	2544.5
75°C	284.6	463.1	734.2
	1071.5	1376.5	1793.0
80°C	11.4	69.7	98.0
	127.9	183.8	229.6
	294.9	362.7	
85°C	23.3	35.7	52.0
	57.8	83.8	
90°C	10.3	10.9	15.3
	23.5	33.5	



(a) (b)

Fig. 2. Example of a stage one model fit (a) and corresponding stage two model fit (b) using the Whiting-Oriente model.

69 occasions the $P_{\max}$ parameter was either less than zero or greater than one. This means of course that the log-likelihood statistic cannot be determined for this model. In order to form a consistent model of the data, the fitting of the survivor function and linear regression components of the model must be conducted in a single step.

V. Conventional Parametric Survival Analysis

The conventional approach to survival analysis seeks to form a linear model that accurately captures the relationship between the explanatory variables, $\boldsymbol{x}_i$, and the parameters of the survivor function used to model the corresponding survival times, t_i. The parameters of the linear model are normally determined by a maximum likelihood approach. In this study, we construct a polynomial regression model by performing a linear regression with all main effects, squared main effects and all interaction terms. Again a backward elimination procedure is used to remove redundant regressors.

TABLE IV
Log-likelihoods for linear parametric survival models.

Model	Log-likelihood
Exponential	-2153.1
Weibull	-2018.5
Log-logistic	-1893.8

Figure 3 gives examples of results obtained using exponential, log-logistic and Weibull survivor functions to model an observed set of five replicate with common heat treatment and incubation conditions. Again, the log-logistic and Weibull models appear to give a subjectively better fit to the observed data. This result is confirmed by the log-likelihood statistics for each mode, shown in table IV, in this case the log-logistic model outperforms the Weibull models.

VI. Kernel Survival Analysis

In recent years the "kernel trick" has been shown to provide a principled and effective strategy for constructing non-linear variants of existing linear statistical techniques, such as ridge regression [10, 11], principal component analysis and Fisher discriminant analysis [12, 13], as well as more modern methods such as the maximum margin classifier [14, 15] (for a comprehensive survey of kernel learning methods, see Schölkopf and Smola [16]). In this work, we use this approach to develop a non-linear survival analysis technique. Given a dataset consisting of observed survival times (which may or may not be censored) and explanatory variables,

$$\mathcal{D} = \{(\boldsymbol{x}_i,\ t_i)\}_{i=1}^{\ell}, \quad \boldsymbol{x}_i \in \mathcal{X} \subseteq \mathbb{R}^d, \quad t_i \in \mathbb{R}^+.$$

we construct linear survival models in a feature space $\mathcal{F}$ formed by a fixed transformation of the input variables, $\boldsymbol{\phi} : \mathcal{X} \rightarrow \mathcal{F}$. Rather than specifying the transformation $\boldsymbol{\phi}$ directly, the feature space is instead induced by a kernel function $\mathcal{K} : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$, obeying Mercers conditions [17], that defines the inner product between vectors in $\mathcal{F}$, i.e. $\mathcal{K}(\boldsymbol{x}, \boldsymbol{x}') = \langle \boldsymbol{\phi}(\boldsymbol{x}) \cdot \boldsymbol{\phi}(\boldsymbol{x}') \rangle$. A common kernel function is the radial basis function,

$$\mathcal{K}(\boldsymbol{x}, \boldsymbol{x}') = \exp\left\{-\gamma \|\boldsymbol{x} - \boldsymbol{x}'\|^2\right\},$$

used for all kernel models comprising this study. In the case of the Weibull survivor function, the kernel survival model is then defined by

$$\beta(\boldsymbol{x}) = \exp\left\{\langle \boldsymbol{w}_\beta \cdot \boldsymbol{\phi}(\boldsymbol{x}) \rangle + b_\beta\right\}$$

and

$$\eta(\boldsymbol{x}) = \exp\left\{\langle \boldsymbol{w}_\eta \cdot \boldsymbol{\phi}(\boldsymbol{x}) \rangle + b_\eta\right\}.$$

The exponential transformation ensures that the modelled parameters of the survivor function are strictly positive. A logistic transformation can be incorporated into the model if a parameter of the survivor function is required to lie in the range $[0,\ 1]$. The model parameters $\boldsymbol{w}_\beta, b_\beta, \boldsymbol{w}_\eta$ and b_η are then determined by minimising the regularised negative log-likelihood of the observed data,

$$\begin{aligned} E &= \sum_{\mathcal{U}} \log f(t_i \mid \boldsymbol{x}_i) + \sum_{\mathcal{C}} \log F(t_i \mid \boldsymbol{x}_i) \\ &+ \lambda_\beta \|\boldsymbol{w}_\beta\|^2 + \lambda_\eta \|\boldsymbol{w}_\eta\|^2, \end{aligned}$$

where λ_β and λ_η are regularisation constants [18], controlling the bias-variance trade-off [19] for the models used to estimate the conditional shape and scale parameters of the Weibull survivor function respectively. The representer theorem [20] suggests that the minimiser of regularised criteria of this form can be written as expansions over training data,

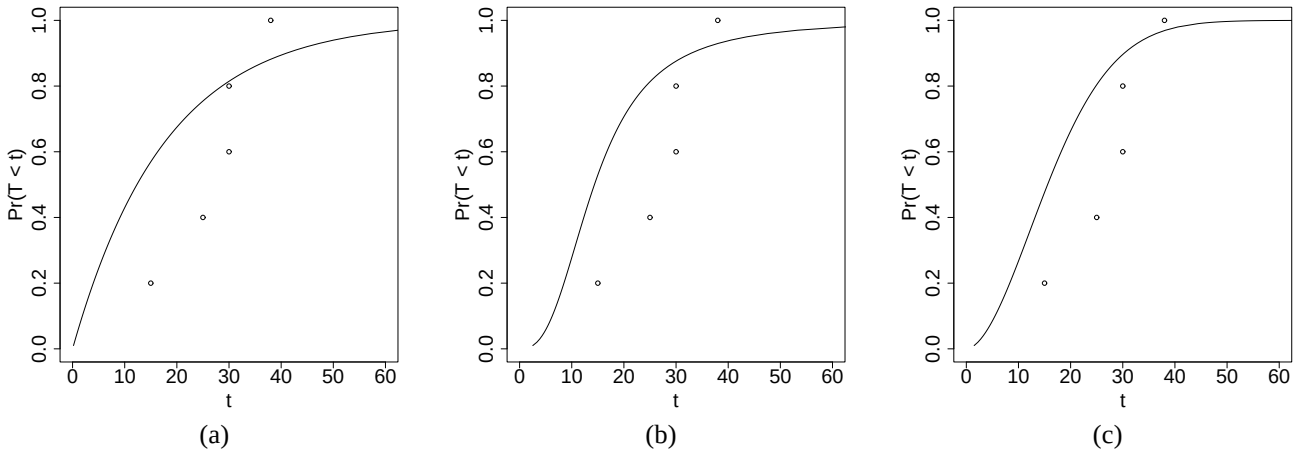
$$\boldsymbol{w}_\beta = \sum_{i=1}^{\ell} \alpha_i^\beta \boldsymbol{\phi}(\boldsymbol{x}_i), \quad \text{and} \quad \boldsymbol{w}_\eta = \sum_{i=1}^{\ell} \alpha_i^\eta \boldsymbol{\phi}(\boldsymbol{x}_i).$$

The kernel survival model can then be written in the form of kernel expansions:

$$\beta(\boldsymbol{x}) = \exp\left\{\sum_{i=1}^{\ell} \alpha_i^\beta \mathcal{K}(\boldsymbol{x}_i, \boldsymbol{x}) + b_\beta\right\},$$

and

$$\eta(\boldsymbol{x}) = \exp\left\{\sum_{i=1}^{\ell} \alpha_i^\eta \mathcal{K}(\boldsymbol{x}_i, \boldsymbol{x}) + b_\eta\right\}.$$



(a) (b) (c)

Fig. 3. Example output of linear parametric survival analysis models using exponential (a), log-logistic (b) and Weibull (c) survivor functions.

Like other kernel methods, the optimal parameters of the model are given by the solution of a convex optimisation problem, without local minima (in this study we use a second order Newton gradient descent training procedure). A second advantage of a kernel approach is that control of over-fitting is simply a matter of choosing the optimal values of the regularisation parameters, in this case λ_β and λ_η, via a k-fold cross-validation strategy [21].

Figure 4 (a-c) show example model fits for kernel survival analysis models based on exponential, log-logistic and Weibull survivor functions respectively. In each case the necessary kernel and regularisation parameters were optimised so as to minimise the 10-fold cross-validation estimate of the negative log-likelihood using a Nelder-Mead optimisation method [22]. As is the case for classical linear parametric survival analysis, the and log-logistic and Weibull distributions appear to be subjectively superior to the exponential distribution. Figures 4 (d-e) show contour plots giving the probability of growth as a function of incubation temperature and incubation time for each model. The contours of constant probability appear smooth in each case, demonstrating that the models have not over-fitted the data.

TABLE V

LOG-LIKELIHOODS FOR KERNEL PARAMETRIC SURVIVAL MODELS.

Distribution	Log-likelihood
Exponential	-2094.4724
Log-logistic	-2015.9851
Weibull	-1768.2970

Table V shows the log-likelihoods for kernel survival models based on exponential, log-logistic and Weibull survivor functions. Again the two-parameter log-logistic and Weibull functions are superior, the best performance being obtained using the Weibull survivor function. The kernel models generally outperform the conventional survival analysis methods based on linear regression. This is because the use of regularisation allows the use of flexible non-linear kernel functions whilst avoiding over-fitting the training data.

Contour plots, such as those shown in figure 4 (c-d), giving the probability of growth as a function of incubation temperature and incubation time provide a straight-forward means of determining the growth domain of a microbial pathogen such as *Clostridium botulinum*. A contour can be drawn representing a given probability of growth, say 0.05; the growth domain is then defined as the region of the plot above this contour, the region below the contour representing food storage conditions considered to be at low risk from hazards associated with *Clostridium botulinum*.

VII. CONCLUSIONS

In this paper we have shown that the use of the "kernel trick" provides a principled means of constructing flexible non-linear models for the analysis of survival data, based on the large body of existing theory from classical statistics. The proposed kernel survival analysis method clearly outperforms survival analysis based on linear regression and a rather more ad-hoc model representative of current practice in growth domain modelling on a dataset describing the growth domain of *Clostridium botulinum*. The improved performance lies in the ability to model arbitrary relationships between treatment/incubation regimes and the parameters of the survivor function, and the use of regularisation to avoid over-fitting.

ACKNOWLEDGEMENTS

The authors would like to thank Nicola Talbot for her assistance in typesetting this paper and Gary Barker, Richard Lee and Gareth Janacek for their helpful comments. This work was supported by the Biotechnology and Biological Research Council (grant number 83/D17534 and competetive strategic grant funding to MWP).

REFERENCES

[1] B. M. Lund and M. W. Peck, "Clostridium botulinum," in *The Microbiological Safety and Quality of Food*, B. M. Lund, A. C. Baird-Parker, and G. W. Gould, Eds., pp. 1057–1109. Aspen, Gaithersburg, USA, 2000.

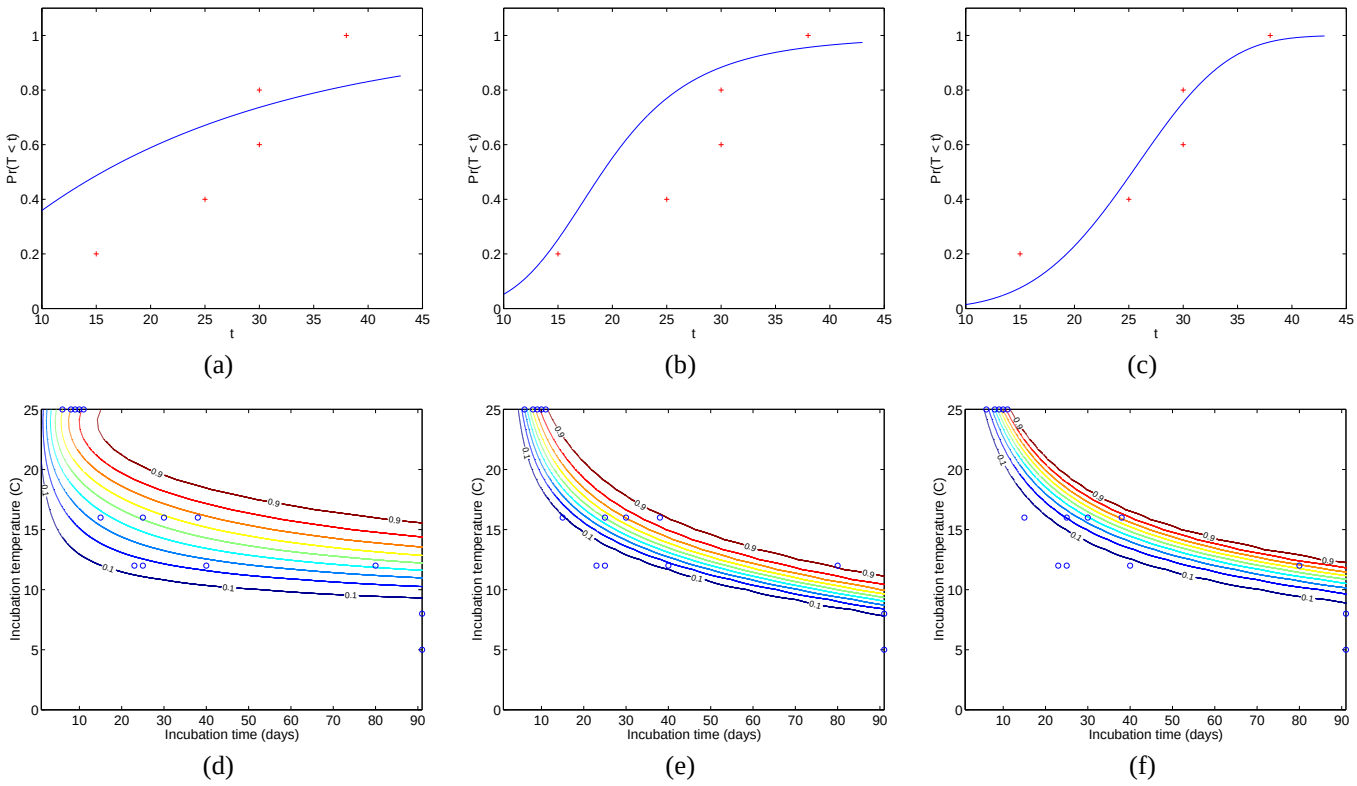


Fig. 4. Example output of kernel survival analysis models using exponential (a), log-logistic (b) and Weibull (c) survivor functions and corresponding contour plots for the probability of growth as a function of incubation time and temperature (d-f), following heating at $75°$C for 463.1 min.

[2] M. W. Peck, "*Clostridium botulinum* and the safety of refrigerated processed foods of extended durability," *Trends in Food Science and Technology*, vol. 8, pp. 186–192, 1997.

[3] D. R. Cox and D. Oakes, *Analysis of Survival Data*, Chapman and Hall, London, 1984.

[4] P. S. Fernández and M. W. Peck, "A predictive model that describes the effect of prolonged heating at 70°–90°C and incubation at refrigeration temperatures on growth and toxigenesis by non-proteolytic *Clostridium botulinum* in the presence of lysozyme," *Applied and Environmental Microbiology*, vol. 65, no. 8, pp. 3449–3457, Aug. 1999.

[5] B. M. Lund and M. W. Peck, "Heat resistance and recovery of spores of non-proteolytic *Clostridium botulinum* in relation to refrigerated, processed foods with extended shelf-life," *Applied Bacteriology*, vol. 76, pp. 115–128, 1994.

[6] M. W. Peck, B. M. Lund, D. A. Fairbairn, A. S. Kassperson, and P. C. Undeland, "Effect of heat treatment on survival of, and growth from, spores of non-proteolytic *Clostridium botulinum* at refrigeration temperatures," *Applied and Environmental Microbiology*, vol. 61, pp. 1780–1785, May 1995.

[7] S. C. Stringer, N. Haque, and M. W. Peck, "Growth from spores of non-proteolytic *Clostridium botulinum* in heat treated vegetable juice," *Applied and Environmental Microbiology*, vol. 65, no. 5, pp. 2136–2142, May 1999.

[8] F. Carlin and M. W. Peck, "Growth and toxin production by non-proteolytic and proteolytic *Clostridium botulinum* in cooked vegetables," *Letters in Applied Microbiology*, vol. 20, pp. 152–156, 1995.

[9] R. C. Whiting and J. C. Oriente, "Time-to-turbidity model for non-proteolytic type B *Clostridium botulinum*," *International Journal of Food Microbiology*, vol. 35, no. 1, pp. 49–60, 1997.

[10] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.

[11] C. Saunders, A. Gammermann, and V. Vovk, "Ridge regression in dual variables," in *Proceedings of the Fifteenth International Conference on Machine Learning (ICML-1998)*, J. Shavlik, Ed. 1998, Morgan Kaufmann.

[12] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, pp. 179–188, 1936.

[13] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, and K.-R. Müller, "Fisher discriminant analysis with kernels," in *Neural Networks for Signal Processing*, vol. IX, pp. 41–48. IEEE Press, New York, 1999.

[14] B. E. Boser, I. M. Guyon, and V. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth Annual ACM Workshop on Computational Learning Theory*, D. Haussler, Ed., Pittsburgh, PA, July 1992, pp. 144–152.

[15] C. Cortes and V. Vapnik, "Support vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.

[16] B. Schölkopf and A. J. Smola, *Learning with kernels - support vector machines, regularization, optimization and beyond*, MIT Press, Cambridge, MA, 2002.

[17] J. Mercer, "Functions of positive and negative type and their connection with the theory of integral equations," *Philosophical Transactions of the Royal Society of London, A*, vol. 209, pp. 415–446, 1909.

[18] A. N. Tikhonov and V. Y. Arsenin, *Solutions of ill-posed problems*, John Wiley, New York, 1977.

[19] S. Geman, E. Bienenstock, and R. Doursat, "Neural networks and the bias/variance dilema," *Neural Computation*, vol. 4, no. 1, pp. 1–58, 1992.

[20] G. S. Kimeldorf and G. Wahba, "Some results on Tchebycheffian spline functions," *J. Math. Anal. Applic.*, vol. 33, pp. 82–95, 1971.

[21] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *Journal of the Royal Statistical Society*, vol. B 36, no. 1, pp. 111–147, 1974.

[22] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Computer Journal*, vol. 7, pp. 308–313, 1965.