



Evaluation of different approaches for modeling *Escherichia coli* O157:H7 survival on field lettuce



Robin C. McKellar^a, Fernando Pérez-Rodríguez^b, Linda J. Harris^c, Anne-laure Moyne^{c,d}, Burton Blais^e, Ed Topp^f, Greg Bezanson^g, Susan Bach^a, Pascal Delaquis^{a,*}

^a Agriculture and Agri-Food Canada, Pacific Agri-Food Research Centre, 4200 Highway 97 South, Summerland, BC, Canada V0H 1Z0

^b Department of Food Science and Technology, University of Córdoba, Córdoba, Spain 14014

^c Department of Food Science and Technology, University of California, One Shields Avenue, Davis, CA, USA 95616-8598

^d Western Center for Food Safety, University of California, 1477 Drew Avenue, Davis, CA, USA 95618

^e Canadian Food Inspection Agency, Building 22, 960 Carling Avenue, Ottawa, ON, Canada K1A 0Z2

^f Agriculture and Agri-Food Canada, Southern Crop Protection and Food Research Centre, 1391 Sandford Street, London, ON, Canada N5V 4T3

^g Agriculture and Agri-Food Canada, Atlantic Food and Horticulture Research Centre, 32 Main Street, Kentville, NS, Canada B4N 1J5

ARTICLE INFO

Available online 2 May 2014

Keywords:

E. coli O157:H7
field lettuce
predictive model
quantitative risk assessment
biphasic model

ABSTRACT

The ability to predict the behavior of *Escherichia coli* O157:H7 on contaminated field lettuce is essential for the development of accurate quantitative microbial risk assessments. The survival pattern of the species was assessed from several data sets derived from field-based experiments, which were analyzed by regression analysis fitting one monophasic model (log-linear) and two biphasic (Weibull and Cerf's model) models. Probabilistic models were also simulated with @RISK™, integrating the fitted monophasic and biphasic models in order to analyze their impact on the estimate of the extent of die-off subsequent to a contamination event in the field. Regression analysis indicated that *E. coli* O157:H7 followed a biphasic decay pattern in most cases, with the Weibull and Cerf's model showing similar good fit to individual and pooled survival data. Furthermore, results from the stochastic analysis demonstrated that using the log-linear model could lead to different risk estimates from those obtained with biphasic models, with a lower prevalence in the former scenario as no tailing is assumed in this model. The models and results derived from this work provide the first suitable mathematical base upon which to build probabilistic models to predict the fate of *E. coli* O157:H7 on field-grown leafy green vegetable.

Crown Copyright © 2014 Published by Elsevier B.V. All rights reserved.

1. Introduction

Recurring outbreaks of foodborne illness caused by the verocytotoxin-producing bacterium *Escherichia coli* O157:H7 have motivated attempts to define the public health risks associated with consumption of lettuce using a quantitative microbial risk assessment (QMRA) approach. QMRA has been defined as the process of estimating the risk from exposure to microorganisms by combining dose–response information for the infectious agent with information on the distribution of exposures (Haas, 1999). The Codex Alimentarius Commission (1999) has proposed a framework for performing QMRA which consists of four components including hazard identification, hazard characterization (i.e. dose–response models), exposure assessment, and risk characterization. The accuracy of QMRAs is improved by the availability of measurements related to the behavior of hazards in specific foods and the development of mathematical models able to predict exposure due to contamination at the various stages along the farm-to-consumption

chain. QMRAs have been proposed for *E. coli* O157:H7 in leafy vegetables delivered through specific distribution systems (Franz et al., 2010; Tromp et al., 2010; Pérez Rodríguez et al., 2011) or the complete farm-to-consumption chain (Danyluk and Schaffner, 2011). In each case, models were constructed to simulate behavior of the pathogen at specific stages. The required predictive mathematical equations were derived using data from a variety of sources that included previously published work, laboratory studies or measurements performed in commercial settings. There also has been an attempt to develop a mathematical model to predict the fate of *E. coli* O157:H7 in unprocessed lettuce from harvest through retail display using survival or growth data derived from laboratory investigations and temperature profiles measured in a commercial distribution chain (Koseki and Isobe, 2005).

Although these collective works have contributed important advances toward the development of QMRAs for leafy vegetables, all are hampered by the lack of models that can simulate pathogen behavior before harvest. Consequently, the risk of exposure associated with consumption of field-grown crops is difficult to assess. This is a critical gap given the anticipated significance of contamination and pathogen

* Corresponding author. Tel.: +1 250 494 6367.
E-mail address: pascal.delaquis@agr.gc.ca (P. Delaquis).

survival at this stage along the farm-to-consumption chain. The dearth of field-based research on the behavior of specific human pathogens on the edible portions of leafy vegetables has undoubtedly led to this deficiency. However, recent investigations conducted with field lettuce have yielded kinetic data that could be used to develop suitable quantitative approaches in order to predict the fate of *E. coli* O157:H7 following a contamination event (Erickson et al., 2010; Moyne et al., 2011; Bezanson et al., 2012). A joint analysis of the available information allows that more reliable and representative results could be generated given the limitations inherent to these studies when treated individually (i.e. few data or restricted and variable experimental conditions) (Viallette et al., 2005). Therefore, the purpose of this work was to use the available data in the development of quantitative approaches and models describing the fate of the *E. coli* O157:H7 species in field lettuce to be included in QMRA studies.

2. Material and methods

2.1. Data selection

Data sets derived from published reports and additional data sets collected by the authors through experimentation on the survival of *E. coli* O157:H7 in field lettuce were used in the development of the models described in the present work. To the best of our knowledge, this comprises the sum of accessible field data available at the time of writing. It should also be noted that data derived from controlled environment chamber or greenhouse trials were not considered due to apparent differences in behavior under such experimental conditions (see, for example, Ottoson et al., 2011). Literature data were taken from tables or digitized from graphs using Engauge digitizing software (<http://digitizer.sourceforge.net/>). The data sets were derived from trials conducted in different geographic regions, using different lettuce cultivars and by variable experimental procedures. For example, the study by Moyne et al. (2011) was carried out in California (USA) using different inoculation procedures resulting in initial inoculum densities ranging from 5–7 log cfu/g, in different seasons and on plants that were either 2 or 4 weeks old. In contrast, Erickson et al. (2010) conducted field experiments in Georgia (USA) using inoculation procedures that ensured inoculation of both abaxial and adaxial leaf surfaces. Hence the data sets incorporate survival data obtained under a wide range of environmental conditions.

2.2. Data analysis and modeling

Survival data for *E. coli* O157:H7 were collected, tabulated and standardized to represent $\Delta S = S_0 - S_t$ where S_0 is the cell density (log cfu/g) at $t = 0$ and S_t is the cell density at a given time, t , in days. Each independent data set was submitted to a regression analysis to fit different mathematical functions describing survival or log decrease over time to detect common or reproducible survival patterns. A similar approach was followed after pooling of the data. Regression was performed using the curve fitting tool provided in MATLAB 7.7.0 Software (The MathWorks Inc. Natick, 2008). Constraints were applied to regression parameters to optimize the regression process. In order to assess goodness of fit of each model, sum of square error (SSE), root mean square error (RMSE), adjusted coefficient of determination (adj-R^2), F-test and the corrected Akaike Information Criterion (AICc) were used. The mathematical functions selected to be fitted corresponded to two

inactivation models describing different inactivation kinetics. On the one hand, the log-linear model describing first-order kinetics with parameter k_3 (Bigelow and Esty, 1920), and on the other hand, the biphasic model consisting of two log-linear phases: an initial log-linear decrease (first-order kinetic; k_1) undergone by the sensitive microbial population, and an asymptote in which a more resistant microbial population ($1 - f$) decreases at a lower decay rate (k_2) (Cerf, 1977). In addition, the Weibull model defined by the parameters a and b (Peleg, 2006) was fitted to data. As its mathematical function is able to represent for biphasic and monophasic ($b \sim 1$) kinetics, it was deemed interesting to assess the performance of the model with respect to its potential application to represent different inactivation patterns. The mathematical functions used for the fitting process are shown in Table 1. The RMSE and the R^2 were determined to assess the goodness of fit of the estimated models. In order to enable comparison between both fitted models, the AICc was computed. This parameter is particularly suitable for comparing non-nested models with a different number of parameters such as the models studied here. In order words, AICc determines the model with the fewest parameters that still provides an adequate fit to the data, and the lowest AICc value corresponds to the most adequate model.

2.3. Monte Carlo simulation

Simulations of the survival models were obtained from 10,000 iterations of the appropriate probability distributions for survival time, initial cell concentration and uncertainty on the regression parameters from the Cerf and log-linear models, performed using @RISK™ 5.7 in Microsoft® Excel (Palisade Corp., Ithaca, New York, USA, http://www.palisade.com/decisiontools_suite/). A normal distribution was chosen for initial cell concentration since bacterial counts in non-laboratory systems are usually described by a bell-like distribution when values are log-transformed. Time in the field was represented by an exponential distribution that is often used to describe time between events in simulations. Uncertainty in the fitted overall Cerf model was represented through a triangular distribution defining the 95% prediction limits which had been obtained during fitting of pooled data with the non-linear regression procedure of GraphPad Prism version 5.04 for Windows, GraphPad Software, San Diego, California, USA, www.graphpad.com. Uncertainty on k_{max} from the log-linear model was determined by fitting a normal distribution to the array of k_{max} values obtained from the different individual data sets. Results from simulations were examined by comparing the frequency distributions for surviving *E. coli* O157:H7 populations under different modeling assumptions.

3. Results and discussion

3.1. Quantitative assessment of the survival pattern of *E. coli* O157:H7 on field lettuce based on individual data sets

The fit of the models to the different data sets and goodness of fit indices are shown in Fig. 1 and Table 2, respectively. Although experimental conditions differed between studies, statistical indexes such as RMSE, SSE and adj-R^2 as well as graphical representation indicated that in most cases the Cerf and Weibull models, representing a biphasic behavior, better described *E. coli* O157:H7 survival. To confirm this result, the F-test was used as the biphasic and linear models are nested. The F-test allows determination of the model best accounting for the

Table 1

Mathematical functions and parameters of the survival models applied to describe *E. coli* O157:H7 survival on leafy green vegetables in field.

Model Name	Model	Model Parameters	References
Log-linear	$\log(N/N_0) = k_3 \cdot t$	k_3	Bigelow and Esty (1920)
Biphasic model	$\log(N/N_0) = \log[f \cdot \exp(-k_1 \cdot t) + (1 - f) \cdot \exp(-k_2 \cdot t)]$	f, k_1, k_2	Cerf (1977)
Weibull	$\log(N/N_0) = -[(t/a)^b]$	a, b	Mafart et al. (2002)

observed survival pattern. The null hypothesis that the simpler model (linear model) is correct, i.e. none of the parameters added in the extended model (biphasic model) significantly improve the fit, was assessed at a significance level of 0.05. Since $p \leq 0.05$ leads to rejection

of the null hypothesis, the best model was the biphasic model. However, limitations in the number of data points ($n \leq 4$) precluded application of the F-test in several cases. For some data sets with very few data points (i.e., $n = 4$) the computed F-test provided evidence that the linear

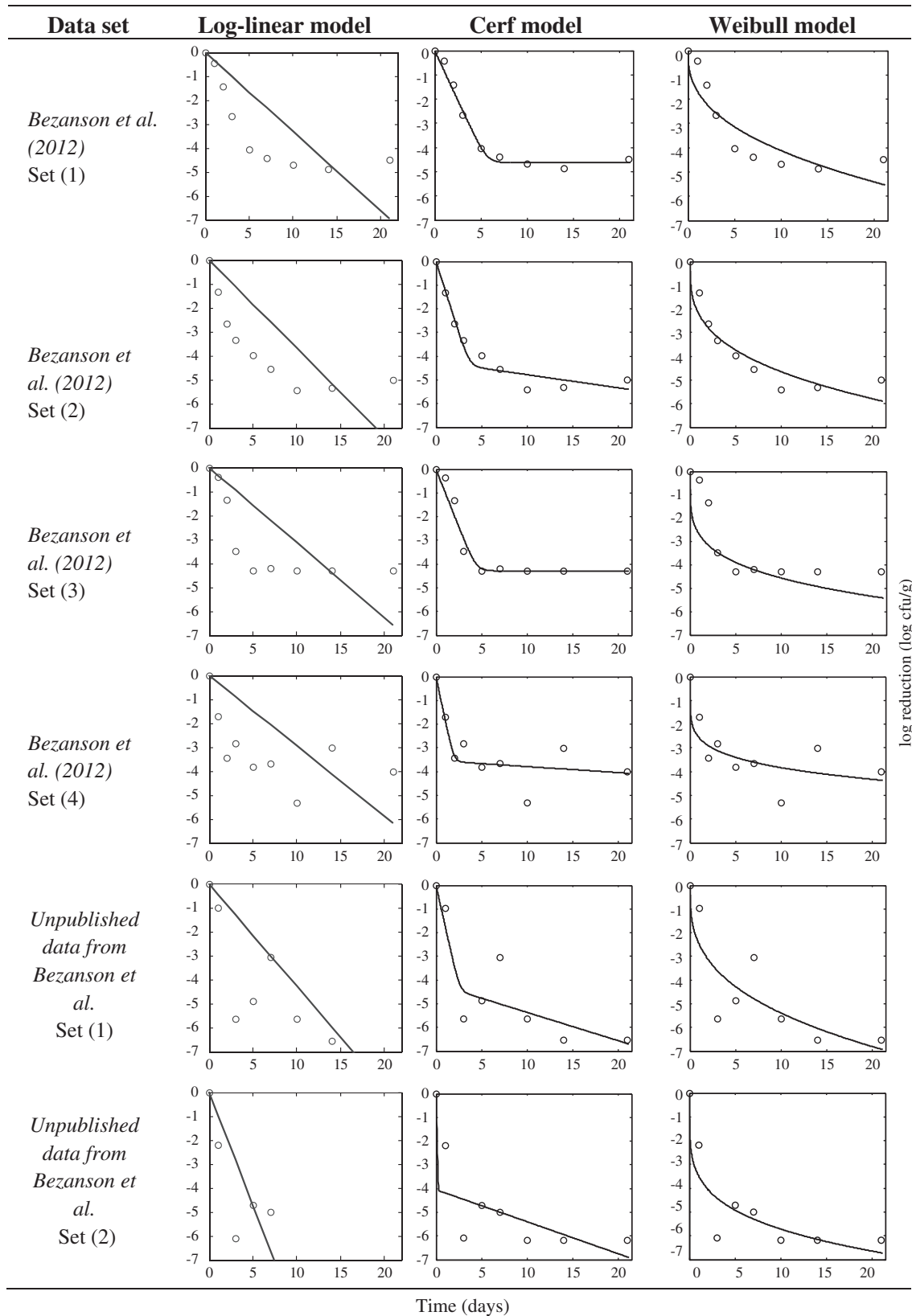


Fig. 1. The log-linear model, Cerf model and Weibull model to different individual data sets describing *E. coli* O157:H7 survival on leafy vegetables in field.

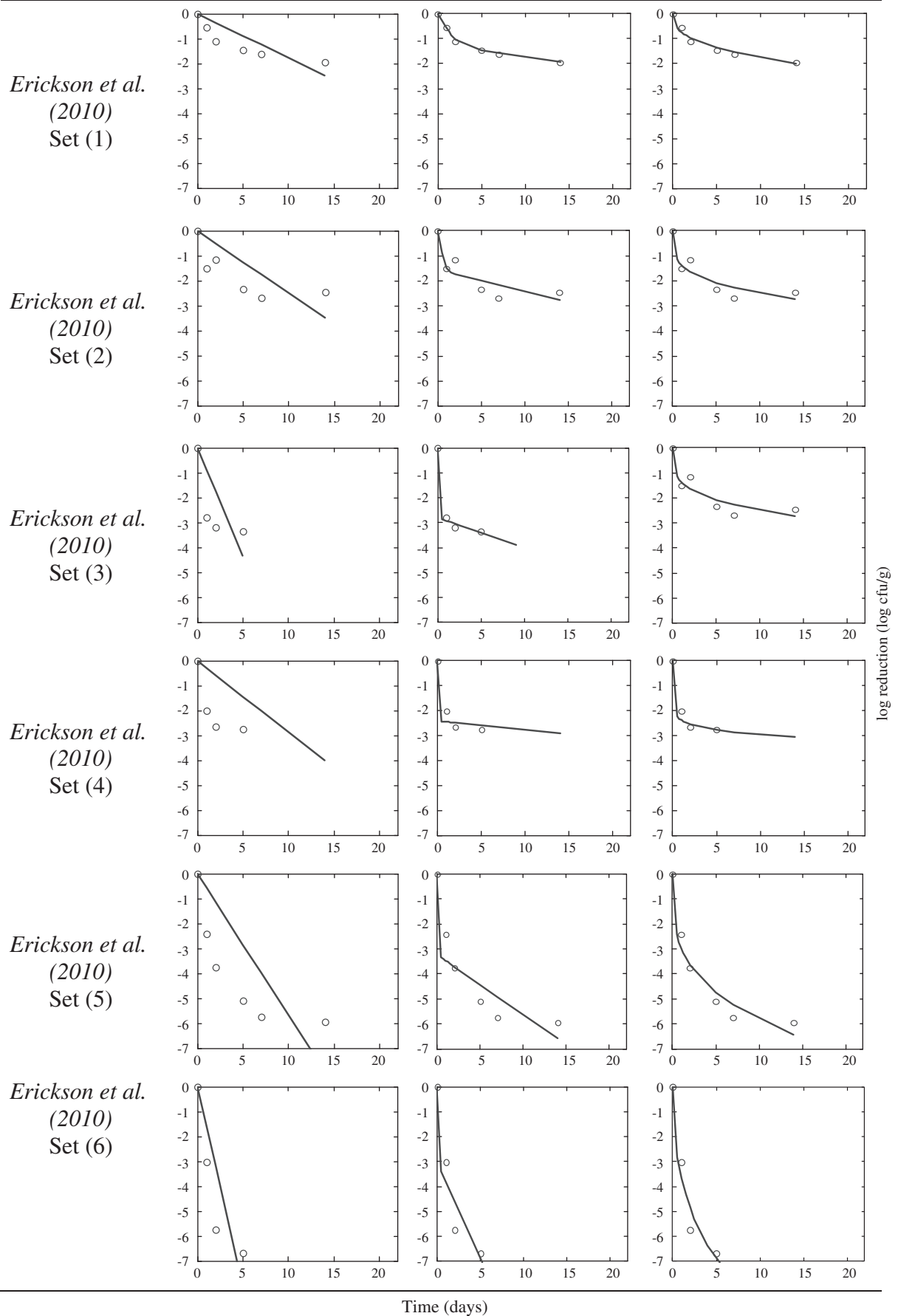


Fig. 1 (continued).

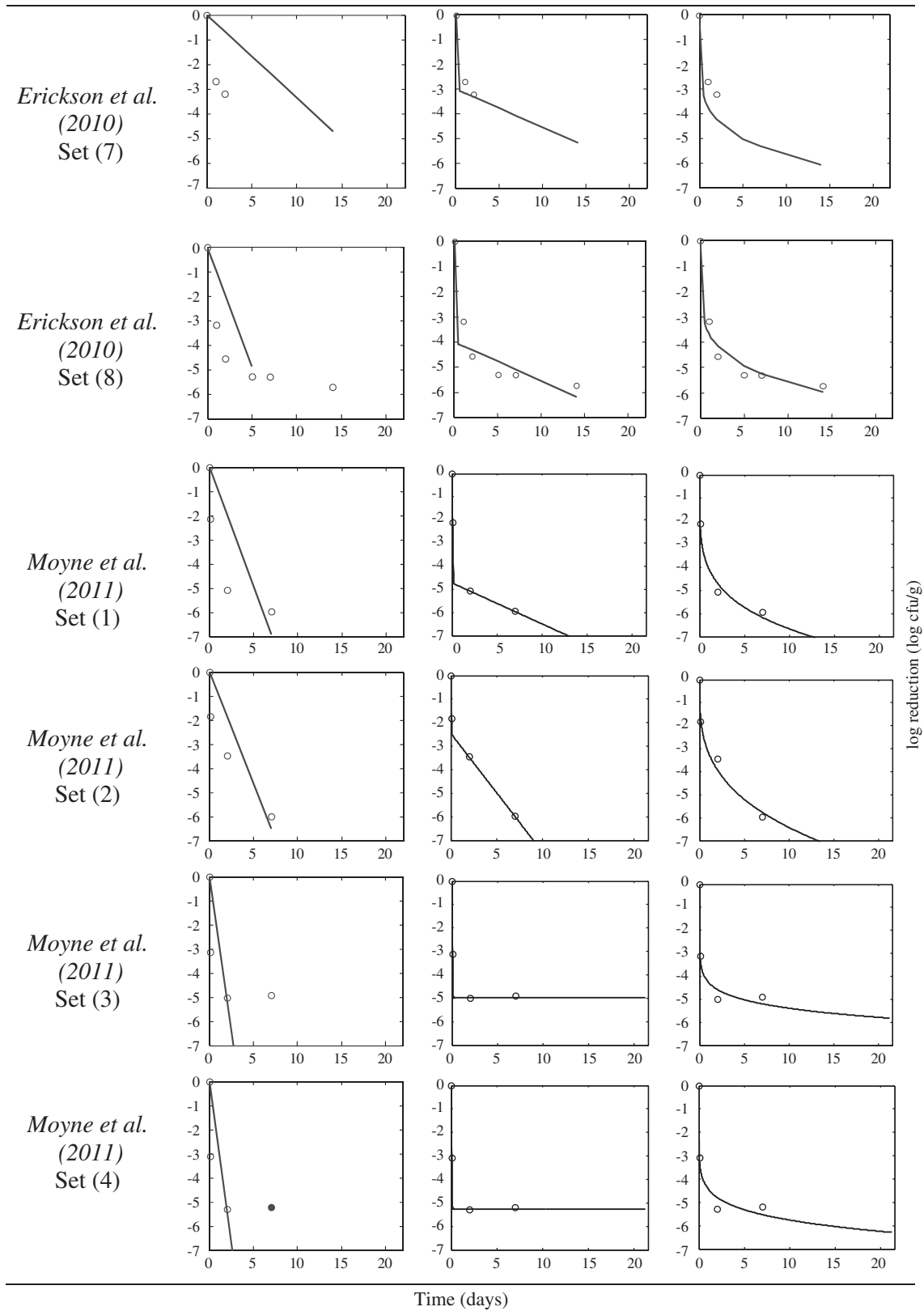


Fig. 1 (continued).

model was more correct. Furthermore, examination of the AICc values enabled comparison of the biphasic model with the Weibull model, since neither is nested. The results generally indicated that the Weibull model was a better choice on the basis of a lower fitting error and the

parsimony principle. The Weibull model is often applied in predictive microbiology as means to characterize different survival patterns such as upward concavity, downward concavity and an initial shoulder (Peleg and Cole, 1998; Fernández et al., 2002; Van Boekel, 2002;

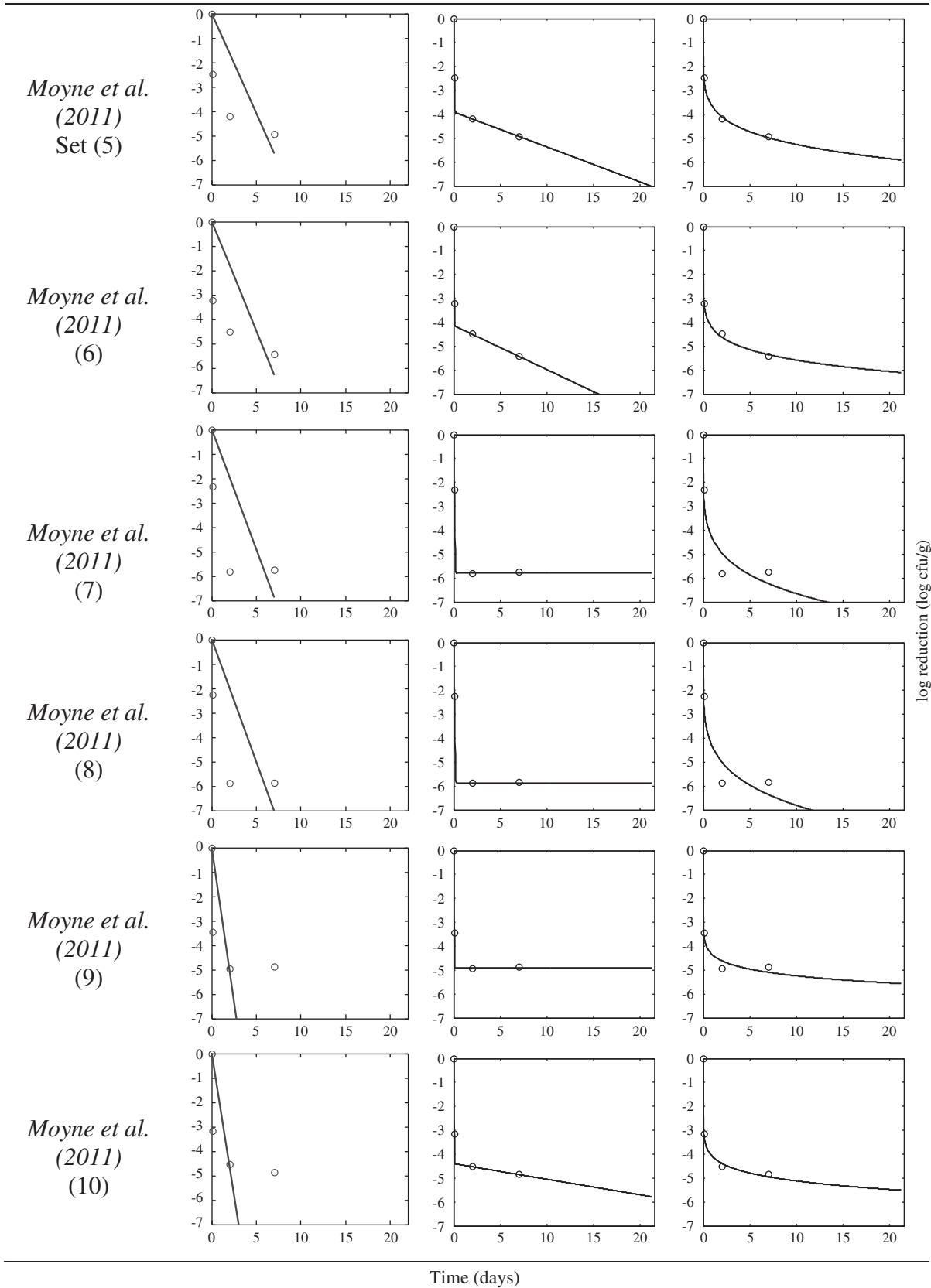


Fig. 1 (continued).

Aragao et al., 2007; Angelidis et al., 2010; Izquierdo and Gómez-López, 2011). Microbial inactivation as characterized by the Weibull distribution function may be viewed as a failure phenomenon where the

fraction of bacterial survivors progressively decreases in response to cumulative lethal effects caused by harsh environmental conditions (Peleg, 2006). Furthermore, this model can account for the existence

Table 2
Regression parameters and goodness-of-fit statistics obtained from the fitting of the Cerf model (1), Weibull model (2) and log-linear model (3) to different individual data sets describing *E. coli* O157:H7 survival on leafy green vegetables in field.

Reference	Data Set	Inoculum	Model	k_3	a	B	f	k_1	k_2	SSE	R2	Adj-R ²	RMSE	fit parameters	n	AICc	F-ratio	p-value	
Bezanson et al. (2012)	1	5.2	1				1.000	1.88	0.00	0.37	0.99	0.98	0.25	3.00	9.00	-10.75	172.16	0.00	
			2		0.277	0.394				0.37	0.988	0.983	0.85	2.00	10.00	-23.00			
			3	-0.33							21.55	0.81	0.78	1.64	1.00	9.00	13.86		
	2	5.4	1					1.000	2.77	0.13	0.88	0.97	0.96	0.38	3.00	8.00	3.69	78.15	0.00
			2		0.08	0.317					2.51	0.91	0.90	0.60	2.00	8.00	2.73		
			3	-0.37							28.45	0.80	0.76	1.89	1.00	8.00	16.55		
	3	5.0	1					1.000	2.23	0.00	1.28	0.95	0.94	0.46	3.00	9.00	0.47	55.22	0.00
			2		0.18	0.342					6.27	0.77	0.73	0.95	2.00	9.00	7.55		
			3	-0.31							24.90	0.76	0.73	1.76	1.00	9.00	15.16		
	4	5.3	1					1.000	4.16	0.06	3.71	0.80	0.73	0.79	3.00	8.00	15.18	20.08	0.00
			2		0.00	0.174					4.72	0.74	0.71	0.82	2.00	8.00	7.77		
			3	-0.29							33.49	0.68	0.62	2.05	1.00	8.00	17.85		
Unpublished data from Bezanson et al.	1	6.7	1				1.000	3.94	0.28	6.99	0.84	0.78	1.18	3.00	8.00	20.25	9.96	0.01	
			2		0.07	0.336					10.21	0.77	0.73	1.31	2.00	8.00	13.95		
			3	-0.43							34.84	0.81	0.78	2.23	1.00	8.00	18.17		
	2	6.2	1					1.000	5.00	0.13	1.98	0.95	0.93	0.63	3.00	5.00	NC	6.46	0.14
			2		0.00	0.217					5.12	0.86	0.84	0.92	2.00	5.00	30.12		
			3	-0.94							14.78	0.83	0.78	1.92	1.00	5.00	15.42		
Erickson et al. (2010)	1	2.3	1				0.943	1.54	0.12	0.01	1.00	1.00	0.05	3.00	4.00	-57.27	98.96	0.07	
			2		2.16	0.376					0.08	0.97	0.96	0.14	2.00	4.00	NC		
			3	-0.18							1.43	0.86	0.78	0.54	1.00	4.00	11.90		
	2	4.5	1					0.973	4.48	0.20	23.33	0.67	0.62	1.34	3.00	6.00	56.15	-1.17	NC
			2		0.29	0.259					0.56	0.89	0.86	0.37	2.00	6.00	3.75		
			3	-0.25							5.12	0.77	0.71	1.01	1.00	6.00	7.05		
	3	5.5	1					0.998	40.56	0.28	0.04	0.99	0.98	0.20	3.00	6.00	18.17	243.25	0.00
			2		0.00	0.107					0.02	1.00	1.00	0.10	2.00	6.00	-16.47		
			3	-0.86							6.79	0.77	0.71	1.50	1.00	6.00	8.74		
	4	2.7	1					0.996	30.41	0.08	0.27	0.95	0.92	0.30	3.00	4.00	-42.75	19.85	0.16
			2		0.00	0.094					0.16	0.97	0.97	0.20	2.00	4.00	NC		
			3	-0.29							11.07	0.67	0.50	1.49	1.00	4.00	20.07		
5	6.0	1					0.999	40.60	0.55	2.67	0.90	0.83	0.94	3.00	6.00	43.13	16.02	0.02	
		2		0.02	0.292					0.92	0.97	0.96	0.48	2.00	6.00	6.73			
		3	-0.57							31.13	0.74	0.67	2.50	1.00	6.00	17.88			
6	7.1	1					0.999	30.21	1.79	2.01	0.93	0.78	1.42	3.00	4.00	-34.76	2.05	0.47	
		2		0.04	0.393					1.32	0.95	0.93	0.81	2.00	4.00	NC			
		3	-1.60							10.21	0.88	0.82	1.84	1.00	4.00	19.75			
7	3.2	1				0.999	37.79	0.35	1.49	0.94	0.89	0.70	3.00	4.00	-35.96	5.25	0.31		

Moyne et al. (2011)	8	6.3	2		0.00	0.184				0.58	0.98	0.97	0.38	2.00	4.00	NC			
			3	-0.33						17.11	0.64	0.46	1.85	1.00	4.00	21.81			
			1			0.999	40.60	0.35		1.48	0.94	0.89	0.70	3.00	6.00	39.60	30.05	0.01	
	1	5.0	2		0.00	0.184			0.58	0.98	0.97	0.38	2.00	6.00	3.98				
			3	-0.57					31.13	0.74	0.67	2.50	1.00	6.00	17.88				
			1			1.000	5.00	0.00		0.00	1.00	1.00	0.00	3.00	4.00	NC	NC	-	
	2	5.2	2		0.00	0.218			0.23	0.99	0.98	0.34	2.00	4.00	NC				
			3	-0.98					14.68	0.78	0.66	2.21	1.00	4.00	21.20				
			1			0.997	5.00	1.13		0.00	1.00	1.00	0.00	3.00	4.00	NC	NC	-	
	3	4.9	2		0.02	0.306			0.39	0.98	0.97	0.44	2.00	4.00	NC				
3			-0.92					5.93	0.88	0.83	1.41	1.00	4.00	17.57					
1					1.000	5.00	0.00		0.00	1.00	1.00	0.00	3.00	3.00	NC	NC	-		
4	5.2	2		0.00	0.102			0.31	0.98	0.98	0.32	2.00	3.00	-24.80					
		3	-2.57					8.50	0.76	0.51	2.06	1.00	3.00	NC					
		1			1.000	5.00	0.00		0.00	1.00	1.00	0.00	3.00	3.00	NC	NC	-		
5	5.0	2		0.00	0.116			0.41	0.98	0.97	0.45	2.00	3.00	-23.99					
		3	-2.71					8.26	0.78	0.56	2.03	1.00	3.00	NC					
		1			1.000	5.00	0.09		0.00	1.00	1.00	0.00	3.00	4.00	NC	NC	-		
6	5.2	2		0.00	0.155			0.01	1.00	1.00	0.08	2.00	4.00	NC					
		3	-0.81					9.42	0.80	0.71	1.77	1.00	4.00	19.42					
		1			1.000	5.00	0.00		0.00	1.00	1.00	0.00	3.00	4.00	NC	NC	-		
7	5.5	2		0.00	0.119			0.02	1.00	1.00	0.10	2.00	4.00	NC					
		3	-0.89					13.85	0.77	0.65	2.15	1.00	4.00	20.97					
		1			1.000	5.00	0.00		0.00	1.00	1.00	0.00	3.00	4.00	NC	NC	-		
8	5.5	2		0.00	0.184			1.17	0.95	0.93	0.76	2.00	4.00	NC					
		3	-0.98					21.08	0.71	0.56	2.65	1.00	4.00	22.65					
		1			1.000	5.00	0.00		0.00	1.00	1.00	0.00	3.00	4.00	NC	NC	-		
9	4.9	2		0.00	0.193			1.24	0.95	0.93	0.79	2.00	4.00	NC					
		3	-1.00					20.97	0.72	0.57	2.64	1.00	4.00	22.63					
		1			1.000	5.00	0.00		0.00	1.00	1.00	0.00	3.00	3.00	NC	NC	-		
10	4.9	2		0.00	0.080			0.19	0.99	0.99	0.25	2.00	3.00	-26.36					
		3	-2.54					10.48	0.71	0.42	2.29	1.00	3.00	NC					
		1			53.000	5.00	0.00		0.00	1.00	1.00	0.00	3.00	3.00	NC	NC	-		
			2		0.00	0.097			0.03	1.00	1.00	0.11	2.00	3.00	-31.53				
			3	-2.33	-	-	-	-	-	8.81	0.71	0.42	2.10	1.00	3.00	NC	-	-	

of several bacterial subpopulations of variable sensitivity to lethal stresses which could arise due to innate heterogeneity in resistance within the microbial population or to phenotypic responses that enable long-term fitness under stressful conditions, for example the ability to successfully colonize protective microenvironments on the plant surface. Additional factors which could influence persistence include stress adaptation during the survival process or physical protection within inner leaves or by internalization as suggested by Ibekwe et al. (2009) and Erickson et al. (2010). Examination of individual data sets presented in Fig. 1 hinted at two distinct decay phases representative of a biphasic survival pattern. However, variability in the data precluded the development of a valid deterministic model and stochastic models are suggested in their stead.

3.2. Quantitative assessment of the survival pattern of *E. coli* O157:H7 on leafy vegetables based on the pooled data sets

One of the motivations for pooling data from different studies is to overcome the limitations and consequent bias inherent to individual experiments, such as limited data points (Garg et al., 2008). Consequently, the available data sets were pooled and submitted to regression analysis to examine overall trends in survival (Borenstein et al., 2009). The models and parameters used in the analysis are shown in Fig. 2 and Table 3. Multiple regressions carried out by different weighting methods using standard error of the model, number of data, etc., yielded similar results (data not shown) and weighting was not applied in the final analyses presented in this work. Both graphical analysis and calculation of goodness-of-fit indices hinted at a two-phase kinetic model characterized by a pronounced initial decay followed by a more gradual secondary decay and/or tailing phase (Fig. 2). In addition, results given by the F-test and AICc indices clearly confirmed the biphasic pattern, which hinted that the Weibull model was the most appropriate mathematical function to represent survival of *E. coli* O157:H7 on field lettuce (Table 3). This result was expected since the Weibull model is able to represent biphasic patterns with only two regression parameters, unlike the three regression parameter Cerf model yielding lower goodness of fit indices due to the parsimony principle (Peleg, 2006). Nevertheless, both models could be used to represent two-phase kinetics in this case. The k_1 and k_2 parameters in the Cerf model provide more explicit information about the reduction rate in each phase, which could make this model preferable for stochastic modeling in quantitative risk assessment studies when variables are coupled or certain dependences between variables and factors are relevant to final risk estimate.

The tailing effect observed in bacterial populations undergoing environmental stresses in the phyllosphere is a poorly understood phenomenon (see preceding section) but with important implications given the low infectious dose for *E. coli* O157:H7. Tailing is defined in the Cerf model through the term $1 - f$ whose value corresponds to the proportion of resistant cells in the population, and where f is the proportion of cells in the sensitive population. Calculations for individual data sets used in the present work yielded values for $1 - f$ that were always <5% and in several cases \ll 1%, while for the overall model $1 - f$ was <0.01%. The existence of a relationship between k_2 and the residual population could be statistically confirmed for the data sets taken from Erickson et al. (2010) through a positive correlation between $\log(N_0 * (1 - f))$ vs k_2 according to the Pearson correlation coefficient ($r = 0.92$).

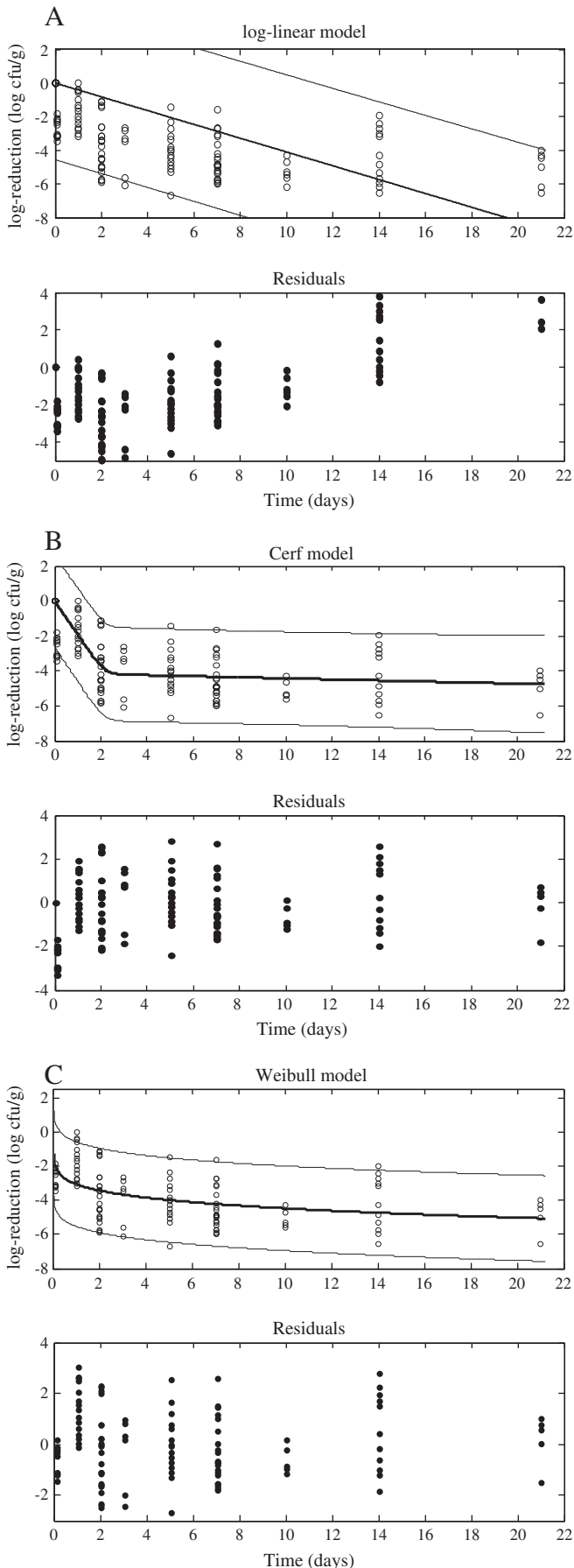
3.3. Rate of population decay of *E. coli* O157:H7 on field lettuce is affected by initial inoculum level

The observed correlation between k_2 and proportion of resistant cells (i.e. $1 - f$) prompted a more detailed search for possible environmental influences on decay rates. The influence of initial inoculum levels on the rate of population decay for *E. coli* O157:H7 on field lettuce was examined by plotting the initial inoculum level against the slope of

the linear portion of the decay curve for each set of survival data. For this simulation, the log-linear model was applied only to the initial linear portion of the individual decay curves to give a series of k_{\max} values (compared to k_3 parameter values which were derived from fitting to complete individual data sets [Table 2]). As shown in Fig. 3 rates of population decay increased with inoculum level. A positive correlation was confirmed by the application of Spearman's rank correlation coefficient, which yielded a value of 0.61 ($p < 0.00$). Such correlation is often apparent in decay data derived from studies carried out with many bacterial species, including *E. coli*, although the cause(s) of this effect remains uncertain (Bazin and Prosser, 1998). Nevertheless, the importance of this phenomenon should not be overlooked in quantitative risk assessment since available data suggest that levels of contamination with *E. coli* O157:H7 in fresh produce are very low. Consequently, decay rates in commercial production systems may be considerably lower than those reported in studies carried out with high inoculum levels. A similar analysis applied to k_1 and k_2 did not show any significant dependence except for the data sets from Erickson et al. (2010), in which a positive correlation between $\log(N_0 * (1 - f))$ vs k_2 was observed, as described in the preceding section.

3.4. Impact of using log-linear or biphasic models on the probabilistic estimation of *E. coli* O157:H7 populations in field

Analysis of individual and pooled data sets indicated that *E. coli* O157:H7 was able to persist on field lettuce at low cell densities leading to a tailing effect described by the Cerf and Weibull models (Figs. 1 and 2). For this reason, it was also important to consider the effect of tailing on the predictive power of the stochastic model. Risk assessors may be more inclined to use log-linear models to describe microbial reduction in the field because they are less complex, or because the data suggest monophasic rather than biphasic decay behavior, which can occur when limited data points are available. However, the selection of one model over another could result in different risk estimates. In this section, log-linear (k_{\max}) and biphasic models were compared in a stochastic environment to assess the impact of both hypotheses on estimates of final *E. coli* O157:H7 cell concentrations on field lettuce. The prediction limits generated from the fitted models were included in this stochastic analysis as truncated distributions to take into account, in the output, the variability and uncertainty derived from models. In QMRA studies, this type of information is crucial to more accurately assess and compare models and scenarios in relation to their impact on the final risk. The Cerf model developed in previous sections was simulated to achieve this end, considering time and initial concentration as stochastic variables as well as variability and uncertainty from both deterministic models using the 95% prediction limits for the Cerf model and the distribution of decay rates (k_{\max}) truncated at the minimum and maximum values of k_{\max} . As correlation was detected between initial concentration and values of k_{\max} , both variables were coupled in the simulation according to the Spearman's rank correlation coefficient described above. In the case of the Cerf model, no correlation was applied since only data from Erickson et al. (2010) showed certain correlation for k_2 in the second decay phase. Variables and models are shown in Table 4. Results indicated that the log-linear model under-predicted *E. coli* O157 survival, showing a pronounced left-tailing effect that yielded a left-skewed distribution as observed in Fig. 4. In turn, for the biphasic model, the distributions showed a more symmetric pattern, with a minimum and maximum value of -8.9 and 7.1 log cfu/g, in contrast with the log-linear model values of -61.4 and 5.3 log cfu/g, respectively. The two simulation approaches also gave different mean values: -2.6 and -1.0 log cfu/g for the log-linear and biphasic models, respectively. Percentile analyses also indicated that the biphasic model predicts higher cell concentrations, as illustrated by the following example. Assuming a head of lettuce weighs 400 g, a positive unit of contamination would result when cell concentration is ≥ -2.6 log cfu/g (i.e. ≥ 1 cfu/400 g). On the basis of this assumption, by applying the frequency values from



the simulated final concentration distributions, the biphasic model would estimate a total of 76% positive units while the log-linear model would predict approximately 65% positive units.

These results suggest that using a log-linear model leads to a major number of iterations with very low levels of the pathogen and consequently a higher likelihood of predicting non-contaminated product. Since the log-linear model was simulated with dependence between the initial concentration and k_{\max} while the Cerf model was done without correlation, some discrepancies could arise when comparing these different approaches. Therefore, an additional simulation was performed to assess the effect of disregarding the correlation for the log-linear model and enable comparison between models in similar simulation conditions. Although the final concentration distributions for log-linear model varied with respect to the simulation with correlation, notably on the minimum value which corresponded to -87.0 , the maximum and mean were practically similar with values of 5.7 and -2.6 log cfu/g. In addition, following the same assumption for lettuce size, the number of positive units was 64%, only one point below the result obtained in the simulation with correlation. We might have predicted a greater influence of correlation, but under the simulation conditions and assumptions here, correlation had a less than expected influence on the outcome of the simulation. Nevertheless that could be different in other models or for other assumptions; hence this aspect should be specifically evaluated in each case when stochastic models are built.

The survival time simulated in the model reached 9 days at 95th percentile, however if longer periods were considered the divergence between results from both models would be even greater. In order to illustrate this effect, a set of simulations were performed with different values for the mean time that lettuce was held in field (1–6 days). Note that according to the exponential distribution properties (distribution used for modeling time) a mean time of 6 days implies that 95th percentile is equal to $3 \times 6 = 18$ days. The percentage (per 1) of contaminated units, assuming a weight of 400 g as before, are represented in Fig. 5A, where it can be observed that as time increased differences between both models became greater, yielding lower percentages of contaminated units in the log-linear model. These results are not unexpected since the absence of a tailing effect in the log-linear model together with an increasing time unavoidably result in a higher log-decrease in the *E. coli* O157:H7 population, thus reducing the number of positive units. However, when an analogous analysis was performed for the mean initial concentration, the effect of changing concentration yielded less evident differences between both modeling approaches as observed in Fig. 5B. These observations illustrate that the selection of a realistic model is crucial to obtain accurate estimates of *E. coli* O157:H7 survival on field lettuce.

4. Conclusions

The present work provides statistical evidence that *E. coli* O157:H7 survival on field lettuce follows a biphasic pattern that is revealed by the application of two quantitative approaches for the analysis of individual survival data sets and pooled data from different studies. Results from the stochastic analysis of the different predictive models indicated that using the monophasic model (i.e. log-linear model) could lead to different risk estimates from those obtained with biphasic models (the Weibull and Cerf models). Thus, a log-linear approach yielded left-skewed distributions for the survivor numbers, in addition to a lower prevalence as no tailing is assumed in this model. Therefore, biphasic models are proposed as suitable predictive models to describe *E. coli*

Fig. 2. The log-linear model (A), Cerf model (B) and Weibull model (C) fitted to the pooled data sets describing *E. coli* O157:H7 survival on leafy vegetables in field and prediction limits associated to each model and residuals obtained from each model fitting.

Table 3
Regression parameters, confidence intervals and goodness-of-fit statistics obtained from the fitting of the Cerf model, Weibull model and log-linear model to the pooled data sets from different studies describing *E. coli* O157:H7 survival on leafy green vegetables in field.

Model	Model parameters						SSE	R ²	Adj-R ²	RMSE	AICc	F-ratio
	K ₃ (95% CI)	a (95% CI)	b (95% CI)	f (95% CI)	k ₁ (95% CI)	k ₂ (95% CI)						
Log-linear	0.41 (0.46, 0.36)						734	-	-	2.31	235.39	
Cerf				0.9999 (1.0000, 0.9998)	4.45 (3.61, 5.29)	0.06981 (-0.08, 0.22)	236.28	0.57	0.56	1.33	83.27	56.13 (0.00)
Weibull		0.00124 (-0.00, 0.01)	0.1663 (0.11, 0.22)		-	-	208.63	0.62	0.62	1.25	63.78	

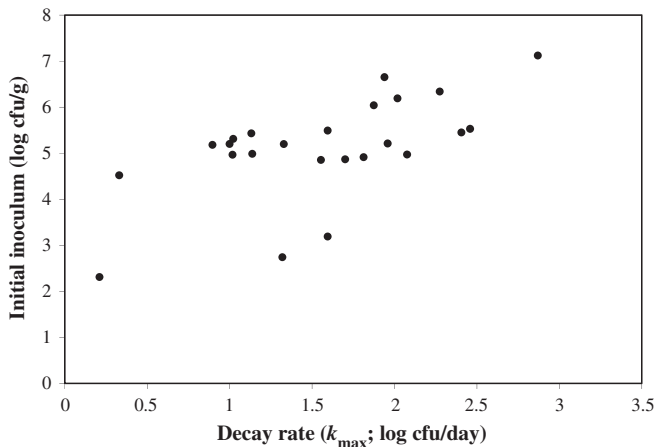


Fig. 3. Graphical representation of the relationship between the decay rate (k_{max}) of the linear region of the survival curves (i.e. slope) and the initial concentration obtained from different analyzed data sets.

O157:H7 survival on field leafy vegetables in quantitative risk assessments.

Acknowledgements

This work was carried out in the BASELINE project. BASELINE is funded by the Seventh Framework Program of the European Commission (grant agreement number 222738).

Table 4
@RISK™ model structure implemented in Excel spreadsheet for a stochastic model simulating *E. coli* O157:H7 survival on leafy vegetables in field.

Cell	Structure	Formula or Value	Description
B4	Model inputs	2	Mean initial (log cfu/g)
B5		1	Standard deviation (log cfu/g)
B6		9	Maximum initial (log cfu/g)
B7		RiskNormal(B4,B5, risktruncate (.B6), RiskCorrmat(NewMatrix2,1))*	Initial concentration distribution (log cfu/g)
B12		3	Average time in field (days)
B13		RiskExpon(B12)	Time distribution in field (days)
G19	Cerf model (biphasic model)	0.00007642	f: proportion of sensitive cells
H19		4.45	k ₁ : rate constant for sensitive cells (d ⁻¹)
I19		0.06981	k ₂ : rate constant for resistant cells (d ⁻¹)
J19		2.82	Prediction limit (log cfu/g)
B14		Log(1 - G19)*EXP(-H19*B13) + G19*EXP(-I19*B13))	Calculation of log reduction
B15		min(0,B14 + RiskTriang(-J19,0,J19))	Addition of prediction limit
B16		B7 + B15	Calculation of cells remaining (log cfu/g)
B17	Log-linear model	RiskNormal(-1.564,0.655,RiskTruncate(-2.87, -0.211,),RiskCorrmat(NewMatrix2,2))*	Distribution of k_{max} : reduction rate (log cfu/g)
B18	(monophasic model)	B17*B13	Calculation of log reduction
		B7 + B18	Calculation of cells remaining (log cfu/g)

* RiskCorrmat(NewMatrix2,1) and RiskCorrmat(NewMatrix2,2) are @RISK™ functions used to correlate k_{max} and initial concentration distribution based on the Spearman's rank correlation coefficient calculated on pooled survival data sets (-0.61).

References

- Angelidis, A.S., Boutsouki, P., Papageorgiou, D.K., 2010. Loss of viability of *Listeria monocytogenes* in contaminated processed cheese during storage at 4, 12 and 22 °C. *Food Microbiol.* 27, 809–818.
- Aragao, G.M.F., Corradini, M.G., Normand, M.D., Peleg, M., 2007. Evaluation of the Weibull and log normal distribution functions as survival models of *Escherichia coli* under isothermal and non isothermal conditions. *Int. J. Food Microbiol.* 119, 243–247.
- Bazin, M.J., Prosser, J.I., 1998. Physiological Models in Microbiology, vol. II. CRC Press, Inc., Boca Raton, FL pp. 1–44.
- Bezanson, G., Delaquis, P., Bach, S., McKellar, R., Topp, E., Gill, A., Blais, M., Gilmour, M., 2012. Comparative examination of *Escherichia coli* O157:H7 survival on Romaine lettuce and in soil at two independent experimental sites. *J. Food Prot.* 75, 480–487.
- Bigelow, W.D., Esty, J.R., 1920. The thermal death point in relation to typical thermophilic organisms. *J. Infect. Dis.* 27, 602–617.
- Borenstein, M., Hedges, L.V., Higgins, J.P.T., Rothstein, H.R., 2009. Meta-regression. *Introduction to Meta-Analysis*. John Wiley & Sons, Ltd.
- Cerf, O., 1977. A review. Tailing of survival curves of bacterial spores. *J. Appl. Bacteriol.* 90, 1–19.
- Codex Alimentarius Commission, 1999. Principles and guidelines for the conduct of microbiological risk assessment. Reference no. CAC/GL 30. Available at: http://www.codexalimentarius.net/download/standards/357/CXG_030e.pdf.
- Danyluk, M.D., Schaffner, D.W., 2011. Quantitative assessment of the microbial risk of leafy greens from farm to consumption: preliminary framework, data, and risk estimates. *J. Food Prot.* 74, 700–708.
- Erickson, M.C., Webb, C.C., Diaz-Perez, J.C., Phatak, S.C., Silvo, J.J., Davey, L., Payton, A.S., Liao, J., Ma, L., Doyle, M.P., 2010. Surface and internalized *Escherichia coli* O157:H7 on field-grown spinach and lettuce treated with spray-contaminated irrigation water. *J. Food Prot.* 73, 1023–1029.
- Fernández, A., Collado, J., Cunha, L.M., Ocio, M.J., Martínez, A., 2002. Empirical model building based on Weibull distribution to describe the joint effect of pH and temperature on the thermal resistance of *Bacillus cereus* in vegetable substrates. *Int. J. Food Microbiol.* 77, 147–153.
- Franz, E., Tromp, S.O., Rijgersberg, H., van der Fels-Klerx, H.J., 2010. Quantitative microbial risk assessment for *Escherichia coli* O157:H7, *Salmonella*, and *Listeria monocytogenes* in leafy green vegetables consumed at salad bars. *J. Food Prot.* 73, 274–285.
- Garg, A.X., Hackam, D., Tonelli, M., 2008. Systematic review and meta-analysis: when one study is just not enough. *Clin. J. Am. Soc. Nephrol.* 3, 253–260.
- Haas, Charles, 1999. Quantitative Microbial Risk Assessment. John Wiley & Sons, Inc., New York.

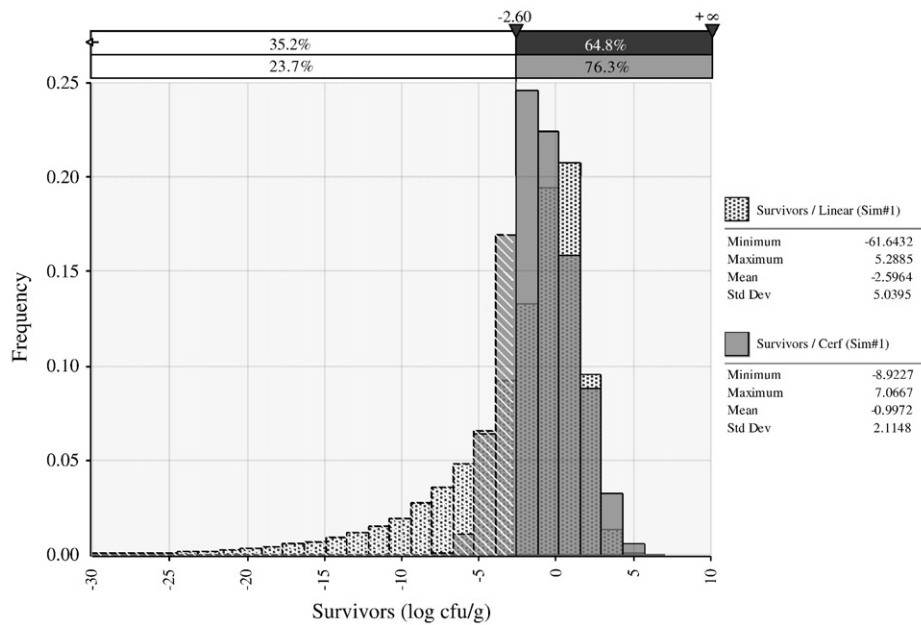


Fig. 4. Simulated survival of *E. coli* O157:H7 populations on leafy vegetables in field obtained from the stochastic model implemented in @RISK™ based on the log-linear model (i.e. monophasic pattern) and Cerf model (biphasic pattern).

Ibekwe, A.M., Grieve, C.M., Papiernik, S.K., Yang, C.H., 2009. Persistence of *Escherichia coli* O157:H7 on the rhizosphere and phyllosphere of lettuce. *Lett. Appl. Microbiol.* 49, 784–790.

Izquierdo, A., Gómez-López, V.M., 2011. Modeling the pulsed light inactivation of microorganisms naturally occurring on vegetable substrates. *Food Microbiol.* 28, 1170–1174.

Koseki, S., Isobe, S., 2005. Prediction of pathogen growth on iceberg lettuce under real temperature history during distribution from farm to table. *Int. J. Food Microbiol.* 104, 239–248.

Mafart, P., Couvert, O., Gaillard, S., Leguerinel, S., 2002. On calculating sterility in thermal preservation methods: application of the Weibull frequency distribution model. *Int. J. Food Microbiol.* 72, 107–113.

Moyne, A., Sudarshana, M., Blessington, T., Koike, S., Cahn, M., Harris, L., 2011. Fate of *Escherichia coli* O157:H7 in field-inoculated lettuce. *Food Microbiol.* 28, 1417–1425.

Ottoson, J.R., Nyberg, K., Lindqvist, R., Albin, A., 2011. Quantitative microbial risk assessment for *Escherichia coli* O157 on lettuce, based on survival data from controlled studies in a climate chamber. *J. Food Prot.* 74, 2000–2007.

Peleg, M., 2006. Isothermal microbial heat inactivation. *Advanced Quantitative Microbiology for Foods and Biosystems*. CRC Press, Boca Raton, USA, pp. 1–48.

Peleg, M., Cole, M.B., 1998. Reinterpretation of microbial survival curves. *Crit. Rev. Food Sci.* 38, 353–380.

Pérez Rodríguez, F., Campos, D., Ryser, E.T., Buchholz, A.L., Posada-Izquierdo, G.D., Marks, B.P., Zurera, G., Todd, E., 2011. A mathematical risk model for *Escherichia coli* O157:H7 cross-contamination of lettuce during processing. *Food Microbiol.* 28, 694–701.

Tromp, S.O., Rijgersberg, H., Franz, E., 2010. Quantitative microbial risk assessment for *Escherichia coli* O157:H7, *Salmonella enterica*, and *Listeria monocytogenes* in leafy green vegetables consumed at salad bars, based on modeling supply chain logistics. *J. Food Prot.* 73, 1830–1840.

Van Boekel, M.A.J.S., 2002. On the use of the Weibull model to describe thermal inactivation of microbial vegetative cells. *Int. J. Food Microbiol.* 74, 139–159.

Viallette, M., Pinon, A., Leporq, B., Dervin, C., Membré, J.-M., 2005. Meta-analysis of food safety information based on a combination of a relational database and a predictive modeling tool. *Risk Anal.* 25, 75–83.

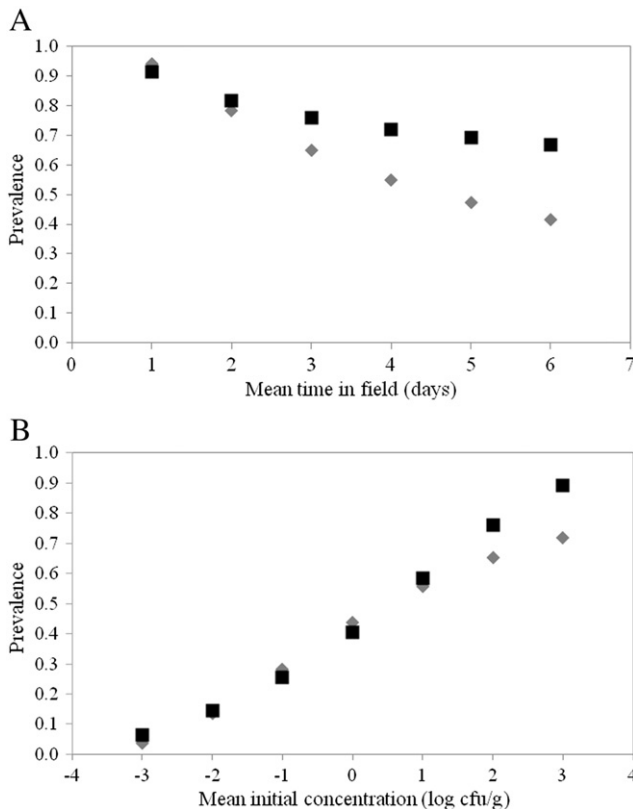


Fig. 5. Results from a scenario analysis applied to the stochastic model showing the effect of different exposure times (A) and initial concentrations of *E. coli* O157:H7 (B) on the survival of the pathogen on leafy vegetable in field when log-linear model (rhombus) and Cerf model (squares) are used.