PROBABILITY OF *ESCHERICHIA COLI* CONTAMINATION SPREAD IN GROUND BEEF PRODUCTION

Petko M. Kitanov

School of Mathematical Sciences
Rochester Institute of Technology
Rochester, NY 14623, USA

Allan R. Willms*

Department of Mathematics & Statistics
University of Guelph
Guelph ON N1G 2W1, Canada

(Communicated by Hao Wang)

Abstract. Human illness due to contamination of food by pathogenic strains of *Escherichia coli* is a serious public health concern and can cause significant economic losses in the food industry. Recent outbreaks of such illness sourced from ground beef production motivates the work in this paper. Most ground beef is produced in large facilities where many carcasses are butchered and various pieces of them are ground together in sequential batches. Assuming that the source of contamination is a single carcass and that downstream from the production facility ground beef from a particular batch has been identified as contaminated by *E. coli*, the probability that previous and subsequent batches are also contaminated is modelled. This model may help the beef industry to identify the likelihood of contamination in other batches and potentially save money by not needing to cook or recall unaffected batches of ground beef.

1. **Introduction.** The bacterium *Escherichia coli* is commonly found in the intestine of warm-blooded organisms. Most of its strains are harmless and are a beneficial part of the gut fauna. Pathogenic *E. coli* strains, in particular O157:H7, cause illness in humans. Scallan et al. [12] indicate that a significant part of all cases of acquired food-borne illness in the U.S.A. is caused by the pathogenic strains of *E. coli*. These strains are primarily transmitted from cattle to humans by consumption of meat and especially under-cooked ground beef, although infection can also occur after consumption of dairy products [15, 16] or from other sources.

In the western world, most meat production is concentrated in large meat processing plants, and any outbreak of *E. coli* contamination may affect many people over a wide area. In addition to the health hazard, outbreaks cause large economic losses and have a negative impact on the beef industry. The existing food safety regulations in Canada and the U.S.A. require the removal of all production and raw sources associated with an identified contamination event. This leads to the recalling of large amounts of beef, much of it likely uncontaminated, and consequently

---

big losses for the beef industry. The goal of this paper is to present a probabilistic model for estimation of the likelihood that sequential batches of ground beef produced in a large plant are contaminated with pathogenic *E. coli* given that one batch is contaminated.

E. coli and its impact on human health have being studied extensively in the last several decades. There is a large body of literature dealing with risk assessment for *E. coli* in ground beef and burgers in different countries [4, 6, 7, 8, 13]. A risk assessment model for *E. coli* O157:H7 in ground beef and beef cuts in Canada is presented in [14]. The difficulties in risk assessment of contamination stems from the fact that it depends on many factors, such as production conditions in the meat processing plants, distribution networks, cooking methods, etc. In [17], using data from a large *E.coli* outbreak in the U.S.A., possible sources of contamination are investigated. The main method used in these studies is statistical analysis of empirical data.

Stochastic models for outbreak and transmission of *E. coli* O157:H7 infection in cattle, are presented in [18] and [20].

Most of the existing deterministic mathematical models about *E. coli*, published in recent years, are devoted to the study of the bacterium itself, rather than the study of the illness caused by it or the estimation of the risk of contamination [5, 10, 11].

The sources of contamination by pathogenic *E. coli* in a beef-packing plant has been studied by Aslam et al. [1, 2] and Bell [3]. *E. coli* in meat products, originates mainly from the hides of the incoming animals and is transferred to the trimmings, and subsequently the ground beef, during the dressing of carcasses and carcass breaking. There are still uncertainties regarding the exact relationship between the *E. coli* found on skinned carcasses and that found on meat at later stages of processing. In recent years, in order to reduce the number of bacteria on the skinned carcasses, North American beef processing plants have implemented different decontamination treatments, such as treatments with antimicrobial solutions and pasteurization. The effect of these treatments was investigated in [19]. Although this study found that these treatments significantly reduce the risk of contamination, the hazard can not be completely removed. Also, there are still no effective methods for quickly screening large amounts of ground beef in big production facilities. All of these uncertainties justify using probabilistic methods for control and estimation of *E. coli* contamination in the production of ground beef.

2. **Materials and methods.** The context of this model is a large ground beef production facility. Given that a particular batch of ground beef has been identified as contaminated, this model assigns probabilities of contamination due to the same origin for the other batches in the production cycle. A preliminary, less general, version of this model was presented in a short conference proceedings [9].

2.1. **Basic assumptions.** The primary assumption is that the contamination is due to the presence of the contaminant on a single carcass. Typically, only a portion of the carcass, often part of the fat layer, would be contaminated, but the model treats all portions of this one carcass as contaminated. Spread of the contaminant in the production process is assumed to be due to division and dispersion of the contaminated carcass portions; transfer via physical contact with other pieces and machinery surfaces is assumed to be negligible. Further, due to the temperature

at which production occurs, it is assumed that the contaminant does not grow appreciably.

Ground beef is produced in batches. Each batch has input from several raw sources, typically one or more "lean" fresh sources and "fat" fresh sources, but also often frozen sources and other sources such as Boneless Lean Beef Trimmings (BLBT), also called Lean Finely Textured Beef (LFTB), which is extracted from trimmings via a centrifuge at temperatures around 38 C. BLBT is usually free of bacterial contamination since it is typically treated to kill bacteria before being used. The meat in a batch is well-mixed and ground together, so that, if any contamination is present on any of the raw source material that is input to the batch, the entire batch is deemed to be contaminated.

As a carcass is processed, parts of it are trimmed off and put into raw source bins to be used as input to the ground beef batches. Typically a carcass gets spread over a region in the raw source. The size of this region and the probability of a piece of the carcass being present at any point in that region is highly dependent on the production process. If carcasses are spread *across* sources, measures of carcass overlap between any two sources are necessary for the model.

Some simplifying assumptions are made regarding how carcasses are spread in a particular raw source and how a raw source is used in ground beef production, as illustrated in Figure 1:

- For each raw source, every carcass is the same. The number of pieces contributed by each carcass, the various masses of those pieces, and the manner of the spread of those pieces throughout a region of the raw source, are the same for each carcass present in the raw source.
- Material within a raw source is ordered and used as input to the ground beef production batches in that order.
- The carcasses are sequentially processed. The regions of a raw source through which pieces from sequential carcasses are spread overlap one another but are shifted forward in the ordering. (Boundary effects for carcasses near the beginning and end of the raw source do alter the spread distributions of pieces from these carcasses; see Section 2.2.)

The second assumption allows us to define a "mass location" in each raw source; mass from a particular raw source used in batch number b comes from locations just prior to those for the mass from the same source used in batch number $b + 1$. In other words, the carcass pieces are points and the production geometry is linear. Although these assumptions are likely not valid for a real production facility, they are useful simplifying assumptions for the modelling process. The likely mixing that occurs within the raw source can be partly accommodated by the carcass piece distribution that is adopted in the model. If there is heterogeneity of the carcasses within the source and this aspect was deemed important, then detailed information about the character of this heterogeneity would be required before it could be built into the model.

Suppose there are B batches of ground beef produced in a production cycle and S raw sources used as input to these batches. Not all sources need be used in all batches. It is assumed that a particular batch of ground beef has been identified as being contaminated. This batch is referred to as the "hot" batch, and it is identified as batch number h. The origin of this contamination is due to a single hot raw source and in turn, to a single hot carcass in this raw source.
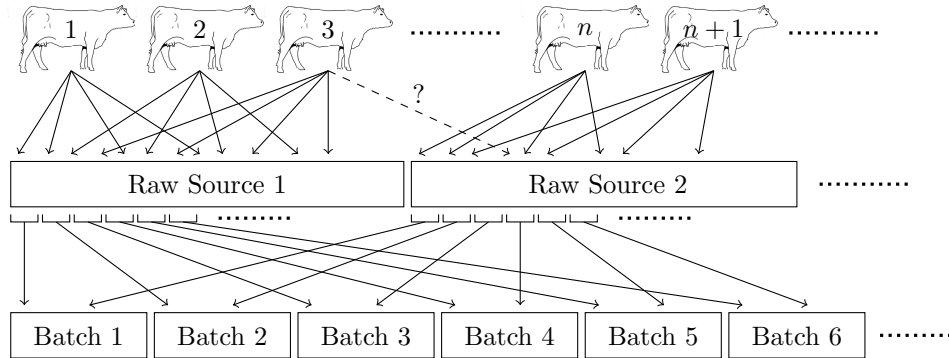
FIGURE 1. Schematic diagram showing spread of meat in ground beef production. Each carcass is spread in a like manner within a region of a raw source. The centres of regions from sequential carcasses are shifted forward in the raw source. The model can account for carcasses being spread across raw sources (dashed line). Material from the raw sources is sequentially input to consecutive batches of ground beef.

The following subsections describe in detail how carcasses are spread in a raw source, which determines their likelihood of appearing in a given batch of ground beef, and how, given a particular hot batch h, the model computes 1) the probability that any particular source is the hot source, 2) the probability that a particular carcass in the hot source is the hot carcass, and 3) the probability that the hot carcass is also present in other batches. These probabilities are then combined in Equation (13), at the end of this section, to determine the probability that other batches are contaminated. Table 1 lists the symbols used in the model.

2.2. **Spread of carcasses in a raw source.** It is assumed that each carcass present in a particular raw source s contributes p_s pieces and that the average mass of these pieces is a_s. Let C_s be the total number of carcasses in the raw source, then the total mass in the raw source is $M_s = C_s p_s a_s$. The pieces from each carcass are distributed throughout this raw source as illustrated in Figure 2 and described below.

For all pieces from a particular carcass c, the "centre" of the piece distribution is located in the source at the mass location

$$\mu_c = \left(c - \frac{1}{2}\right) p_s a_s, \qquad c \in \mathbb{Z}, \quad 1 \le c \le C_s = \frac{M_s}{p_s a_s}. \tag{1}$$

Thus the carcass centres are spread evenly across the source. Let F_s be the common *base* probability density function for all pieces from all carcasses in raw source s. Then, excluding boundary effects, the probability density function for pieces from carcass c in this raw source would be $F_s(x - \mu_c)$, where x is the mass location in the source. However, since the source is finite, carcasses with centres near the boundary would typically have F_s being nonzero beyond the beginning or end of the source. To prevent this, the distribution function is reflected at the boundaries ($x = 0$ and $x = M_s$) and added to the base. Therefore, the probability density function for

(a)

Base piece probabilty density, F_s

(b)

raw source mass location
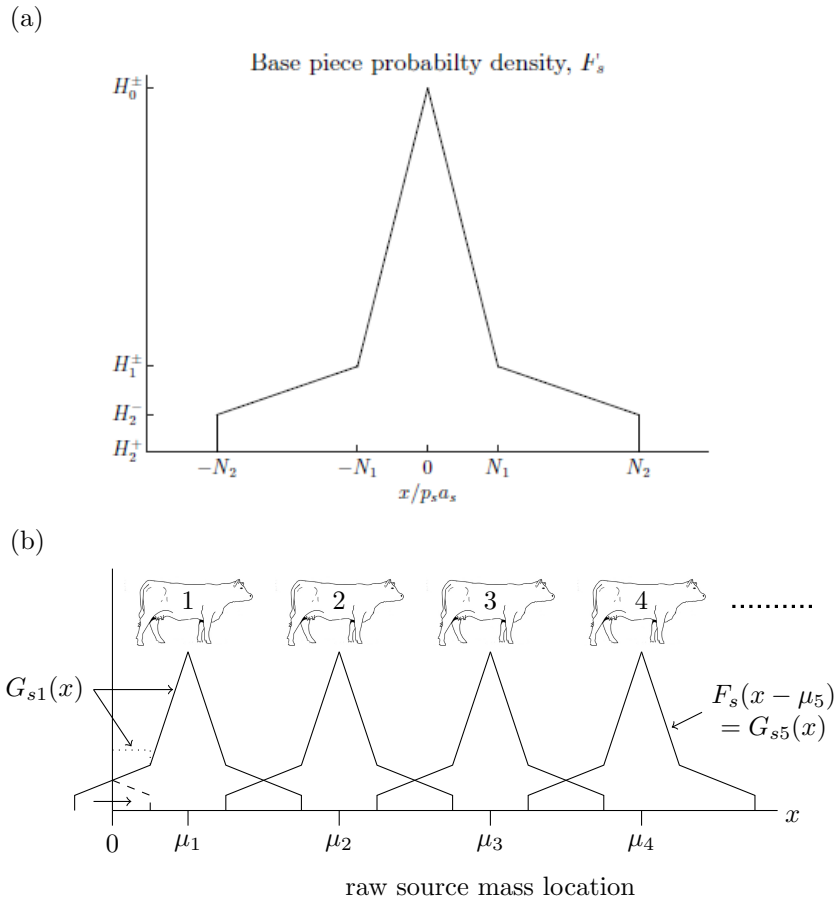
FIGURE 2. Spread of carcasses in a raw source. (a): Example base probability density function for pieces from a carcass. The density is symmetric, and piece-wise linear. In this example there are $K = 2$ linear segments both to the left and right of zero, and there are discontinuities at $\pm N_2$. The parameters K, N, and H must be chosen so that the area under the curve equals one. (b): Distributions of carcasses in a raw source. The centres, μ_c, of the distributions of consecutive carcasses are evenly spaced along the raw source. The individual distributions overlap (more than depicted in the figure). For carcasses in the middle of the source, the probability density function G_{sc} is just a shifted version of the base base probability F_s, as illustrated by G_{s5}. At the ends, distributions that extend beyond the boundary are reflected back, as indicated by the arrow and dashed line at the bottom left, and the reflected portion is added to the distribution already there yielding the dotted line distribution G_{s1}.

TABLE 1. List of Symbols.

S	Total number of raw sources.
B	Total number of ground beef batches.
h	The contaminated (hot) batch.
C_s	Number of carcasses in raw source s.
p_s	Number of pieces supplied by each carcass in raw source s.
a_s	Average size of pieces from each carcass in raw source s.
M_s	Total mass in raw source s.
x	Mass location in raw source.
μ_c	Mid point of piece distribution for carcass c (in source s).
F_s	Base probability density function for piece distribution in source s.
$G_{sc}(x)$	Probability density function for piece distribution for carcass c in source s.
$Q_{sc}(R)$	Probability that a piece from carcass c is located in region R in source s.
K	Half the number of piece-wise linear segments of F_s.
N_i	Boundaries of piece-wise linear segments of F_s, measured in number of carcasses from centre, μ_c, of distribution.
$H_i^{\pm}$	Values of F_s at boundary N_i, approaching from the left, $-$, or the right, $+$.
m_{sb}	Mass from source s input to batch b.
M_{sb}	Mass from source s input to batches 1 through $b-1$.
B_{sb}	Interval of mass locations in source s input to batch b.
$A_{sc}(B_{sb})$	Probability that carcass c is absent from the set B_{sb}, that is, carcass c contributes no pieces to batch b through source s.
f_s	Fraction of fat in raw source s.
g_s	Relative susceptibility to contamination factor for source s.
$V_{s_1 s_2}$	Fraction of carcasses present in both raw sources s_1 and s_2.

pieces from carcass c in source s is

$$G_{sc}(x) = F_s(x - \mu_c) + F_s(-x - \mu_c) + F_s(-x - \mu_c + 2M_s), \qquad x \in [0, M_s]. \quad (2)$$

The first term on the right side of (2) is just the base probability function shifted to the centre, μ_c, for this carcass; the second and third terms are the reflections around $x = 0$ and $x = M_s$, respectively. For carcasses near the middle of the raw source, the reflection terms will not contribute anything unless the breadth of the base probability function F_s is very large. In addition, $F_s(x)$ is restricted to be zero for $|x| > M_s$, so that a single reflection at each end incorporates all of the support of F_s. For each piece from carcass c the probability that it is located between mass locations x_1 and x_2 is

$$Q_{sc}([x_1, x_2]) = \int_{x_1}^{x_2} G_{sc}(x) \, dx. \quad (3)$$

More generally, define $Q_{sc}(R)$ for some set R as the integral of $G_{sc}(x)$ over R.

The above assumptions put some restrictions on the density function F_s. In particular, the total expected mass of all carcasses in a particular set must add to

the size of the set. That is, for any set $R \subseteq [0, M_s]$,

$$\sum_{c=1}^{C_s} p_s a_s Q_{sc}(R) = |R|.$$

A number of functions F_s may satisfy this constraint, but here only a class of even (around 0), piece-wise linear functions, where each piece has a length being a multiple of $p_s a_s$, is considered. An example of such a function is shown in Figure 2(a). In the following, a technical description of this class of functions is given. A function F_s in this class may be defined in terms of parameters K, N, and H as below, where the subscript s on the these parameters has been suppressed for readability:

$$F_s(x) = \begin{cases} \frac{\left(\frac{|x|}{p_s a_s} - N_{i-1}\right)H_i^- + \left(N_i - \frac{|x|}{p_s a_s}\right)H_{i-1}^+}{N_i - N_{i-1}} & \text{if } N_{i-1} \leq \frac{|x|}{p_s a_s} < N_i,\ 1 \leq i \leq K, \\ 0 & \text{if } \frac{|x|}{p_s a_s} \geq N_K, \end{cases} \tag{4}$$

where the N_i are integers (number of carcasses) satisfying $0 = N_0 < N_1 < N_2 < \cdots < N_K \leq C_s$, and $H_i^{\pm} \geq 0$, $0 \leq i \leq K$, are the limiting values of F_s at $N_i p_s a_s$ from the left $(-)$ and right $(+)$, respectively. Necessarily, $H_0^- = H_0^+$ and $H_K^+ = 0$. Also, F_s must itself be a valid probability density, hence

$$\sum_{i=1}^{K} (N_i - N_{i-1}) p_s a_s (H_i^- + H_{i-1}^+) = 1.$$

This is just the statement that the area under F_s must equal one. This class of functions includes the uniform distribution, obtained with $K = 1$, and $H_0^{\pm} = H_1^- = \frac{1}{2N_1 p_s a_s}$. If the parameters are chosen such that $H_i^- = H_i^+$, $1 \leq i \leq K$, then the function F_s is continuous. Each raw source s may have a different distribution function F_s, and thus a different set of parameters K, N, and H.

The realized distribution of pieces from a particular carcass is dependent on the locations of all other pieces from *all* carcasses, because piece locations cannot overlap. In particular, the locations of pieces from a given carcass c are not independent selections from the density function $G_{sc}(x)$, nor are they only dependent on previous selections for the same carcass. However, assuming independence greatly simplifies the selection modelling and should introduce minimal error especially when there are a large number of carcasses. Independent selection is therefore employed in this model.

2.3. **Probability of carcass presence in given batches.** The probability that a carcass c from source s is input to batch b is computed as follows. The mass in each raw source is labeled according to the sequential batches for which it is used. Let m_{sb} be the mass input from source s to batch b, and let M_{sb} be the mass input from source s to batches prior to batch b, that is, $M_{sb} = \sum_{i=1}^{b-1} m_{si}$. The range of mass locations in raw source s that are input to batch b is then the interval

$$B_{sb} = (M_{sb}, M_{sb} + m_{sb}], \tag{5}$$

see Figure 3. For source s, denote the probability of carcass c being *absent* from the set B_{sb} as $A_{sc}(B_{sb})$. Since the selection of the p_s pieces for carcass c is modelled as being independent of the selection of other pieces, A_{sc} is given by

$$A_{sc}(B_{sb}) = (1 - Q_{sc}(B_{sb}))^{p_s}. \tag{6}$$
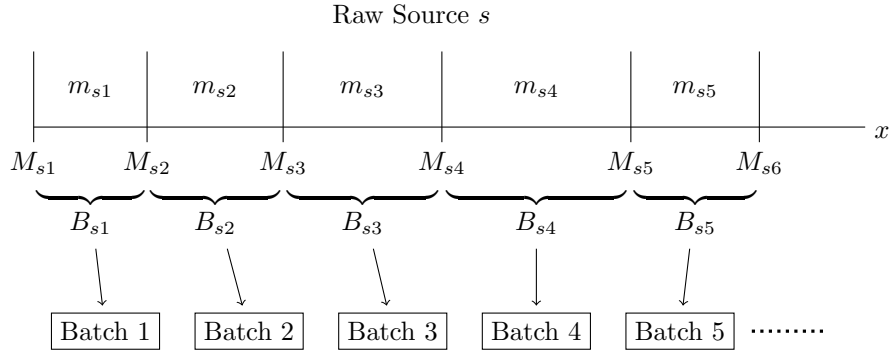
Raw Source s

FIGURE 3. Mass input from raw source s to batches. Batch b receives a mass of m_{sb} from source s. This mass is located in the interval $B_{sb} = (M_{sb}, M_{sb} + m_{sb}]$, where M_{sb} is the total mass used from this source in batches prior to b.

Therefore, the probability of carcass c from source s being present in batch b is

$$\text{Prob}(c \text{ from } s \text{ in } b) = 1 - A_{sc}(B_{sb}). \tag{7}$$

Consider two batches, h and j, that receive input from the same source s, and a particular carcass c from that source. The probability that this carcass has at least one piece that is input to batch h and at least one that is input to batch j is

$$\text{Prob}(c \text{ from } s \text{ in } h \text{ \& } j) = 1 - [A_{sc}(B_{sh}) + A_{sc}(B_{sj}) - A_{sc}(B_{sh} \cup B_{sj})]. \tag{8}$$

The last term is present because it is included in both the previous terms but should only be counted once.

2.4. **Identification of the hot source.** The contamination in the hot batch, h, may be due to any of the raw sources that provided input to this batch, and these raw sources may have varying degrees of relative susceptibility to being contaminated. It is known that the contamination is often carried on the fat, so the fraction of fat, f_s, in raw source s is an important factor. However, even sources with the same fat content may have differing likelihood of being contaminated. For example, a frozen source may be much less likely to be contaminated than a fresh source. To account for these differences, a relative susceptibility factor, g_s, must be assigned to each source by the user. The absolute size of these factors is not important, only their sizes relative to one another. Thus if the user specifies $g_1 = 0.5$, $g_2 = 1.0$, and $g_3 = 1.5$, then source 2 is twice as likely to be contaminated than source 1, source 3 is three times more likely to be contaminated than source 1 and 1.5 times more likely to be contaminated than source 2.

The probability that raw source s is the origin of the contamination in batch h is taken as

$$\text{Prob}(s \text{ is hot} \mid h) = \frac{g_s f_s m_{sh}}{\sum_{i=1}^{S} g_i f_i m_{ih}}. \tag{9}$$

Equation (9) is simply a weighted fractional contribution of mass from source s to the hot batch h. The larger the fraction of mass from source s in batch h, the more likely the contamination came from source s. The weighting is the product of the fat fraction f_s and the susceptibility factor g_s.

2.5. **Identification of the hot carcass.** Given that batch h is the hot batch and assuming the contamination is due to a single hot carcass in raw source s (which contributes mass to batch h) then, with no additional information, the probability that a particular carcass is the hot one is simply the relative probability of that carcass being present in batch h:

$$\text{Prob}(c \text{ from } s \text{ is hot} \mid h) = \begin{cases} \dfrac{\text{Prob}(c \text{ from } s \text{ in } h)}{\sum_{k=1}^{C_s} \text{Prob}(k \text{ from } s \text{ in } h)} & \text{if } m_{sh} > 0, \\ 0 & \text{if } m_{sh} = 0. \end{cases} \quad (10)$$

The right side of the above is computed using (7).

2.6. **Presence of the hot carcass in other batches.** Assuming that a particular carcass c is the hot carcass and that this carcass contaminated batch h via input from source s, then the final step is to compute the probabilities that this carcass is also present in each of the other batches. It is possible that a particular carcass may be present in more than one raw source. This could occur, for example, in a production facility where certain trim from each carcass is placed in one raw source, and other trim from the same carcass, is placed in another source. Presence of carcass c from source s in another batch j, may therefore be due to either input from the same source or input from pieces of the same carcass present in another source.

The model presented here uses a conditional probability of carcass overlap. The probability that carcass c_1 from source s_1 is the same as carcass c_2 from source s_2, given that c_1 from source s_1 is in the hot batch h, is denoted

$$\text{Prob}(c_1 \text{ from } s_1 \equiv c_2 \text{ from } s_2 \mid c_1 \text{ from } s_1 \text{ in } h).$$

This probability must be assigned by the user. If sufficiently detailed information on the trimming process were available, this probability could reflect that, giving different values for different carcasses within each raw source. Without such detailed information, a more coarse approximation of this probability can be assigned using a measure of source overlap, $V_{s_1 s_2}$, defined as the number of carcasses present in both sources divided by the total number of distinct carcasses in both sources. If n carcasses are present in both sources s_1 and s_2, then this overlap fraction is

$$V_{s_1 s_2} = \frac{n}{C_{s_1} + C_{s_2} - n},$$

which may be re-arranged for n giving

$$n = \frac{V_{s_1 s_2} (C_{s_1} + C_{s_2})}{1 + V_{s_1 s_2}}.$$

The probability of carcass overlap can then be approximated by the probability of selecting one of these n carcasses from source s_1 and then selecting the exact same carcass from source s_2:

$$\text{Prob}(c_1 \text{ from } s_1 \equiv c_2 \text{ from } s_2 \mid c_1 \text{ from } s_1 \text{ in } h)$$
$$= \left(\frac{n}{C_{s_1}}\right)\left(\frac{1}{C_{s_2}}\right) = \frac{V_{s_1 s_2} (C_{s_1} + C_{s_2})}{(1 + V_{s_1 s_2}) C_{s_1} C_{s_2}}. \quad (11)$$

This coarse approximation, due to lack of further detail, disregards the condition that carcass c_1 from source s_1 is in batch h.

The probability that carcass c is present in batch j given that carcass c from source s is present in the hot batch h can now be determined. This probability is computed as one minus the product of the probabilities that carcass c is *not* in batch j via equivalence with any other carcass from any other source:

$$\text{Prob}(c \text{ in } j \mid c \text{ from } s \text{ in } h) = 1 - \text{Prob}(c \text{ not in } j \mid c \text{ from } s \text{ in } h)$$

$$= 1 - \prod_{i=1}^{S} \prod_{k=1}^{C_i} \Big[1 - \text{Prob}\big((c \text{ from } s \equiv k \text{ from } i) \,\&\, (k \text{ from } i \text{ in } j) \mid c \text{ from } s \text{ in } h \big) \Big]$$

$$= 1 - \left[1 - \frac{\text{Prob}(c \text{ from } s \text{ in } h \,\&\, j)}{\text{Prob}(c \text{ from } s \text{ in } h)} \right] \prod_{\substack{i=1 \\ i \neq s}}^{S} \prod_{k=1}^{C_i} \Big[1$$

$$- \text{Prob}\big(c \text{ from } s \equiv k \text{ from } i \mid c \text{ from } s \text{ in } h \big) \text{Prob}(k \text{ from } i \text{ in } j) \Big] \quad (12)$$

In the last pair of lines of the above equation the factor with $i = s$ is separated out from the i product since its product over k collapses to a single factor. Also the conditional probability rule

$$\text{Prob}(A|B) = \text{Prob}(A\&B) / \text{Prob}(B)$$

has been used in this factor. Equation (12) may be computed using Equations (7), (8), and (11).

2.7. Probability of other batches being contaminated. The probability that batch j is also contaminated given that batch h is contaminated is then found by 1) summing over all raw sources, computing the probability that that source is hot, 2) summing over all carcasses in that source, computing the probability that that carcass is hot, and 3) multiplying by the probability that that carcass is also present in batch j.

$$\text{Prob}(j \text{ hot} \mid h \text{ is hot}) = \sum_{s=1}^{S} \text{Prob}(s \text{ is hot} \mid h)$$

$$\times \sum_{c=1}^{C_s} \Big[\text{Prob}(c \text{ from } s \text{ is hot} \mid h) \, \text{Prob}\left(c \text{ in } j \mid c \text{ from } s \text{ in } h \right) \Big] \quad (13)$$

This expression can be evaluated using (1)–(12).

3. **Results and discussion.** The following two examples illustrate the above model. The first is a synthetic full set of data for a fictitious beef processing plant, the second is based on a partial data set from genetic typing experiments done at an actual production plant. The data for this final example are used to provide estimates for the spread of a carcass within a raw source and to inform the model about masses of raw source input to batches. Unfortunately, it is not possible to directly verify the model with data since a processing facility clearly cannot deliberately contaminate their ground beef supply with *E. coli*. However, further experiments would be beneficial to help determine the spread of carcasses within a raw source, which will depend very much on the processes in place at any given facility.

TABLE 2. Model parameters for the raw sources in the synthetic data set.

Source	g_s	f_s	p_s	a_s (kg)	N_{1s}	C_s	M_s (kg)
I (frozen lean)	0.2	0.05	25	0.5	15	160	2000
II (frozen lean)	0.2	0.09	25	0.5	15	160	2000
III (frozen lean)	0.2	0.07	25	0.5	15	160	2000
IV (fresh lean)	0.8	0.10	20	0.25	20	500	2500
V (fresh lean)	0.8	0.08	20	0.25	20	600	3000
VI (fresh fat)	1.0	0.40	40	0.2	30	250	2000
VII (fresh fat)	1.0	0.45	40	0.2	30	250	2000

TABLE 3. Source input mass, m_{sb}, (kg) and total fat percentage for the synthetic data set.

	Source							
	frozen lean			fresh lean		fresh fat		
Batch	I	II	III	IV	V	VI	VII	fat %
1	312			136		552		25
2	384			52		564		25
3	114			404		260	222	25
4	262			239		231	268	25
5	201	205		89		293	212	25
6	320	180		292		100	108	15
7	407	105		284			204	15
8		390		456			154	15
9		300		205	325		170	15
10		209		211	543		37	10
11		293		132	536		39	10
12		318	94		540		48	10
13			479		454		67	10
14			701		226		73	10

3.1. **Synthetic data.** These data are fictitious but based on typical values one might encounter in a large ground beef production facility. In this example, there are a total of $S = 7$ sources (three frozen lean, two fresh lean, and two fresh fat), and a total of $B = 14$ one-tonne batches of ground beef are produced. Relevant information for the sources is provided in Table 2. The carcass overlap fractions between different sources were taken as zero except for the following overlap between the fresh sources:

$$V_{45} = 0.001 \quad V_{46} = 0.02 \quad V_{47} = 0.005 \quad V_{56} = 0.007 \quad V_{57} = 0.01 \quad V_{67} = 0.001$$

The uniform distribution for F $(K = 1, H_0^{\pm} = H_1^{-} = 1/(2N_{1s}p_s a_s))$ was used for all raw sources in this example. The mass from each source used in each batch is provided in Table 3.

Before displaying the final results, we first illustrate some of the intermediate model computations for this example. The probability that a particular source is

Table 4. Probability (%) that sources are hot, given hot batch h, for the synthetic data set. Prob(s_1 is hot | h) is computed from Equation (9) and the data from Tables 2 and 3. Blank entries indicate zero probability due to no mass input.

| hot | Source | | | | | | |
| | frozen lean | | | fresh lean | | fresh fat | |
Batch	I	II	III	IV	V	VI	VII
1	1.3			4.6		94.0	
2	1.6			1.8		96.6	
3	0.5			13.6		43.8	42.1
4	1.1			8.2		39.4	51.4
5	0.9	1.6		3.2		52.0	42.3
6	2.7	2.7		19.7		33.8	41.1
7	3.4	1.6		18.9			76.2
8		6.2		32.3			61.4
9		4.5		13.8	17.5		64.2
10		5.2		23.4	48.2		23.1
11		7.8		15.6	50.7		25.9
12		9.1	2.1		54.7		34.2
13			10.2		44.1		45.7
14			17.2		25.3		57.5

the hot source given that batch h is hot, is determined by Equation (9). Given the values in Tables 2 and 3, this probability is easily computed and is reported in Table 4 for each source and each batch being the hot batch. As can be seen from this table, the fresh fat sources typically have the highest probability of being the hot source, followed by the fresh lean sources. This is due to both the fact that the contamination factor g_s in Table 2 is highest for the fat sources, and also because the fat content f_s in these sources is highest. For example, if Batch 11 is the hot batch, the probability that Source VII is the hot raw source from which the contaminated carcass came is above 25% even though Source VII only contributed 39 out of 1,000 kg to the batch.

The probability that a particular carcass within a raw source is the contaminated carcass given that batch h is the hot batch is given by Equation (10) and computed using Equations (1)–(7). For this example, using the uniform distribution as indicated, the results of this calculation for several hot batches h and Sources II and VII are shown in Figure 4. Note from this figure that as the hot batch number increases, the carcass numbers with nonzero probability of being the hot carcass also shift to the right. The flat section of each probability curve is due to the fact that the chosen carcass distribution functions for this example are fairly narrow, and hence most of the central mass contributed to the batch from this source has the same probability of being contaminated. Wider curves correspond to batches that take larger amounts of mass from the source. Results for other sources and other hot batches are similarly computed.

The probability that a particular carcass from a particular raw source that is present in batch h is also present in batch j is given by (12). Consider carcass number 100 in Source VI. From Tables 2 and 3 this source consists of 250 carcasses
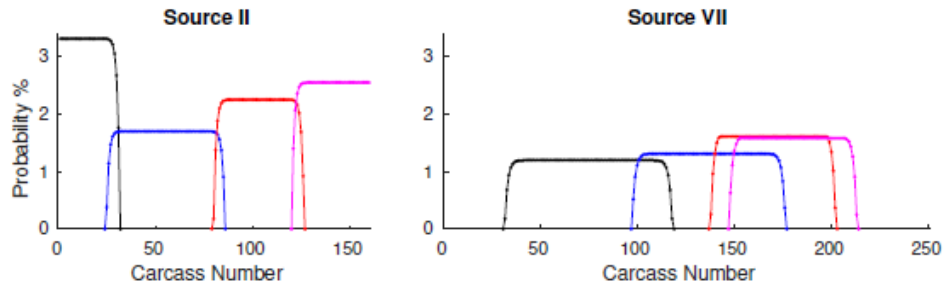
FIGURE 4. Probability (%) that a particular carcass in a hot source is hot given that batch h is the hot batch, Equation (10), for the synthetic data set. The four separate curves in each plot correspond to hot batch $h = 5, 8, 10$, and 12, left to right, respectively.
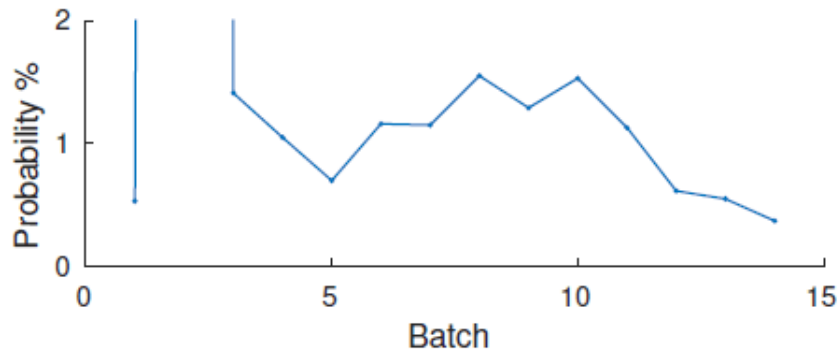
FIGURE 5. Probability (%) that carcass number 100 in Source VI is present in other batches given that it is present in Batch 2, Equation (12), for the synthetic data set.

comprising 2000 kg, about the second quarter of which is input to Batch 2. It is not difficult to show given the piece distribution function, that if there were no overlap of carcasses between sources, this carcass would only be present in Batch 2. However, since Source VI overlaps with Sources IV, V, and VII in this example (V_{46}, V_{56}, and V_{67} are nonzero), there will be nonzero probability of this carcass appearing in other batches. Figure 5 displays this probability. Since V_{46} is considerably larger than either V_{56} or V_{67}, the nonzero probability of carcass 100 from Source VI being present in other batches is larger for the batches to which Source IV contributes significantly, namely Batches 3, 4 and 6 through 11.

Using the full model, and letting each batch be the hot batch in turn, probabilities of contamination for each batch were computed. Complete results are given in Table 5 and a selection of these are plotted in Figure 6.

The likelihood of other batches being contaminated is highly dependent on the source input. In the example, if one of the early batches is the hot batch, then the likelihood of contamination is large only for batches near the hot batch. Conversely,

TABLE 5. Probability of contamination for each batch in percent. Each column corresponds to a different hot batch.

Batch	hot batch 1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	100	46	3	0	0	0	0	1	0	1	0	0	0	0
2	60	100	33	12	0	1	0	1	0	1	1	0	0	0
3	4	46	100	63	32	1	1	1	1	1	1	1	1	0
4	1	20	75	100	70	40	16	1	1	1	1	1	1	0
5	1	1	32	63	100	78	41	16	1	1	1	1	0	0
6	1	1	1	29	61	100	63	28	10	1	0	0	0	0
7	1	1	1	10	23	44	100	58	31	7	5	4	1	0
8	1	1	1	1	10	20	61	100	56	14	13	15	15	11
9	1	1	1	1	1	8	35	58	100	48	24	26	29	29
10	1	1	1	1	1	1	16	29	62	100	53	28	31	32
11	1	1	1	1	1	1	11	25	43	47	100	50	35	36
12	1	1	1	1	1	1	7	21	39	20	41	100	56	42
13	1	1	1	1	1	1	2	17	34	18	22	47	100	67
14	0	0	0	0	0	0	0	10	27	14	18	27	57	100
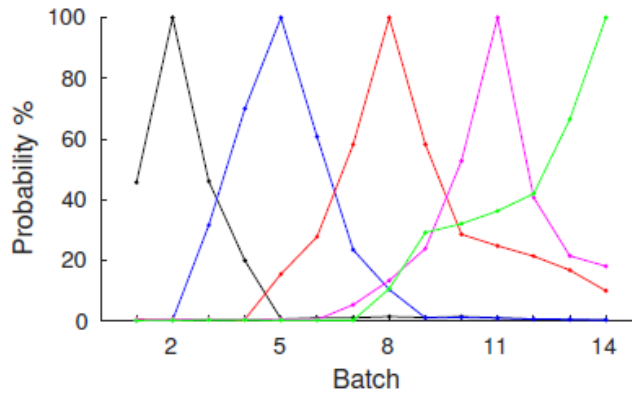
FIGURE 6. Probability of contamination for each batch given that a fixed batch is contaminated (hot). The five separate curves correspond to Batches 2, 5, 8, 11, and 14 being the hot batch.

if one of the last batches is the hot batch, then a greater number of other batches have large probability of being contaminated. This is due to the source configuration shown in Tables 2 and 3. Since the contamination is most likely to be carried in the fatty sources (VI and VII) followed by the lean fresh sources (IV and V), the distribution of these sources across the batches is a primary contributor to the contamination probability distribution. The other two factors that were found to be very important were the values of N_{1s} and Vs_1s_2. If these spread indicators were small, then the contamination was much more confined to nearby batches to the hot batch. The overall result then, is the obvious observation that if one wishes to restrict contamination, then one should restrict the spread of each carcass within the raw source, and restrict the spread of the raw source across batches. In other words, try to make it so that one carcass is present in as few batches as possible.

3.2. **Real data.** These data are from genetic typing experiments performed at an industrial ground beef production facility in 2012. The primary goal of these experiments was to see if genetic typing could be used to accurately determine the number of carcasses present in a ground beef batch and to estimate the amount of overlap of genetic material from batch to batch. Samples of beef were taken from the final ground beef product and genetically analyzed to determine the number of different "profiles" (or distinct animals) present. These profiles were then compared against the profiles found from sampling other batches and the number of shared samples was reported. In addition, this data set provides some information on the sources input to each batch but only partial information on mass amounts. The company where these genetic typing experiments were performed has not provided permission to be named and the experimental results are not publicly available, hence these data are provided without reference.

Although these experiments were not designed specifically to help construct and validate this model, they do provide some information that can be utilized. Our model requires a knowledge of the spread of a carcass across a raw source, but the samples for these data were taken from the final ground beef batches, not raw sources. Therefore these data do not distinguish spread differences between raw sources. Nonetheless, the information available is used to estimate the spread of carcasses in all sources assuming all sources have identical spreads.

There were eight raw sources: four frozen lean, two fresh lean, and two fresh fat. In this experiment 30 of 45 sequential batches were sampled and genetically profiled. Table 6 gives the number of distinct genetic profiles in each batch and the number that match profiles from other batches. The data in this table are interpreted as the weighted average of the conditional probability that a particular carcass is present in batch j given that it is present in batch h, the weights being the probability that that carcass is present in h. That is, if T_{jh} is the entry from row j and column h of the table (although the table does not show the bottom triangle, it is symmetric so that $T_{jh} = T_{hj}$), then

$$\frac{T_{jh}}{T_{hh}} = \frac{\text{\# matched profiles in batches } h \text{ and } j}{\text{\# profiles in batch } h}$$
$$= \frac{\sum_{s_1=1}^{S} \sum_{c_1=1}^{C_{s_1}} \text{Prob}(c_1 \text{ from } s_1 \text{ in } h) \text{Prob}(c_1 \text{ in } j \mid c_1 \text{ from } s_1 \text{ in } h)}{\sum_{s_1=1}^{S} \sum_{c=1}^{C_{s_1}} \text{Prob}(c_1 \text{ from } s_1 \text{ in } h)}. \quad (14)$$

The above can be calculated using (7) and (12). Since the data are limited, providing only information on carcass spread through sums of the sources rather than spread in individual sources, in order not to have too many parameters, the spread in all sources is assumed to be described by the same two-piece ($K = 2$) continuous ($H_1^- = H_1^+$, $H_2^\pm = 0$) distribution function F. The data indicated that there were four suppliers from whom Source I, Sources II-IV, Sources V-VI, and Sources VII-VIII originated, respectively. For this reason, only source pairs II and III, II and IV, III and IV, V and VI, and VII and VIII were allowed to have overlap of carcasses, and the overlap fraction was assumed the same for II and III, II and IV, and III and IV. The number of pieces, p, and the average piece mass, a, were also taken as the same for each source. The parameters a, p, H, N, and V were chosen so that the computed values from the right side of (14) were as close as possible to the data values from the left side of (14) using a sum of squared differences as the minimization function. These results are shown in Table 7. The best fit to the data

TABLE 6. Number of profiles (on the diagonal) and profile matches between batches.

Batch	1	2	3	4	5	6	7	8	9	10	11	12	13	15	16	20	21	22	23	27	28	29	30	33	40	41	42	43	44	45
1	87																													
2	7	56																												
3	15	10	64																											
4	12	2	6	61																										
5	4	3	0	4	57																									
6	7	2	2	8	14	62																								
7	2	1	4	3	1	5	66																							
8	2	1	0	4	4	6	12	50																						
9	1	1	2	0	2	3	6	3	56																					
10	1	0	0	2	1	2	4	2	9	50																				
11	1	1	0	2	4	1	3	6	6	4	58																			
12	1	1	0	2	3	4	3	5	4	4	9	61																		
13	1	1	0	1	0	3	1	0	0	0	2	3	67																	
15	1	1	0	0	0	3	0	1	0	0	0	1	3	63																
16	1	0	0	0	0	0	0	1	1	0	0	3	3	11	57															
20	0	1	0	0	1	1	0	0	1	0	1	2	0	0	0	49														
21	0	0	0	0	0	1	0	0	0	0	1	0	0	3	1	7	87													
22	1	0	0	0	0	1	0	0	1	0	0	2	0	0	2	7	11	45												
23	1	1	0	1	0	1	0	0	2	0	0	1	0	3	3	0	4	3	56											
27	0	0	0	0	0	4	0	0	1	0	2	1	0	2	0	0	2	1	2	52										
28	0	0	0	0	0	0	0	0	1	0	0	0	0	1	3	1	2	1	2	3	51									
29	0	0	0	0	0	0	0	0	3	0	0	1	0	1	0	1	3	3	1	4	3	63								
30	0	0	0	0	0	0	1	1	0	0	2	0	0	0	1	0	0	3	4	2	1	10	71							
33	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2	2	5	0	62						
40	0	0	0	0	0	0	1	0	2	0	1	2	0	0	1	2	5	0	4	4	0	2	0	0	53					
41	0	0	0	0	0	0	0	0	1	0	1	2	0	0	0	1	1	0	0	0	0	1	1	5	5	50				
42	1	0	0	0	0	0	1	0	1	0	2	1	0	0	0	0	1	0	0	1	1	0	1	4	7	4	78			
43	0	0	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	3	6	27	81		
44	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	6	8	15	13	65	
45	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2	0	0	1	1	0	0	0	2	1	3	11	5	14	62

TABLE 7. Fit carcass distribution parameters for the combined raw
sources. The overlap fractions V_{24} and V_{34} are equal to V_{23}.

p	a (kg)	N_1	N_2	$H_0^{\pm}$	$H_1^{\pm}$	V_{23}	V_{56}	V_{78}
8	0.045	27	6383	4.351×10^{-2}	1.177×10^{-5}	0.08	0	0.20

TABLE 8. Estimated source input mass, m_{sb}, (kg) for real data set.

					Source			
		frozen lean			fresh lean		fresh fat	
Batch	I	II	III	IV	V	VI	VII	VIII
1–10	355	355			264		255	
11–13		355	355		264		255	
14–17		355	355		132	132	255	
18–20		355	355			264	255	
21–22		355	355			264		255
23–40			709			264		255
41–45				709		264		255

indicated that each carcass in a raw source contributes 8 pieces of average size 0.1
lbs. (0.045 kg) to that source, totalling 0.8 lbs (0.364 kg). The carcass distributions
are concentrated fairly tightly around their centres, but with long, low tails. Over
94% of the carcass distribution lies within about 26 carcasses (21 lbs., 9.545 kg) on
either side of its centre, with the remaining spreading out as far as 6202 carcasses
(4,962 lbs., 2,255 kg). These parameters for the carcass distributions were used in
all eight of the model's raw sources.

The data from Table 6 indicate profile matches between Batches 1–10 and 23–45
even though Table 8 shows there are no common raw sources for these two sets of
batches. Consequently, there must be some overlap of carcasses across raw sources.
The fitted overlap fractions indicate a substantial carcass overlap in Sources II–IV
(three of the frozen lean sources) and between Sources VII and VIII (the two fresh
fat sources), but no overlap between Sources V and VI (the two fresh lean sources).

Unfortunately, in the data available to us, only the mass input from each raw
source to the first batch was recorded as well as an indication (without mass
amounts) of which raw sources were used in subsequent batches from which samples
were taken. Based on this, the mass input was estimated as shown in Table 8 for
all 45 batches. No further information about the raw sources was provided by this
data set even though much of it would have been available to have been recorded.
The remaining parameters needed for our model have been estimated (or derived
from previous assumptions) and are given in Table 9. Of the remaining parameters,
the susceptibility, g_s, and the fat content, f_s, were given reasonable but arbitrary
values and so that the fat content of the ground beef batches was about 10%. The
mass of each source, M_s, and the number of carcasses in each source were calculated
using the assumed mass inputs of Table 8 and the estimated values of p and a from
Table 7.

Complete results are given in Tables 10 and 11 and a selection of these are plotted
in Figure 7. As can be seen in these tables, the probability of contamination of

Table 9. Estimated and derived model parameters for the raw sources.

Source	g_s	f_s	C_s	M_s (kg)
I (frozen lean)	0.2	0.05	1,950	3,545
II (frozen lean)	0.2	0.05	4,290	7,800
III (frozen lean)	0.2	0.05	9,360	17,018
IV (frozen lean)	0.2	0.05	1,950	3,545
V (fresh lean)	0.8	0.10	2,175	3,955
VI (fresh lean)	0.8	0.10	4,350	7,909
VII (fresh fat)	1.0	0.24	2,800	5,091
VIII (fresh fat)	1.0	0.24	3,500	6,364
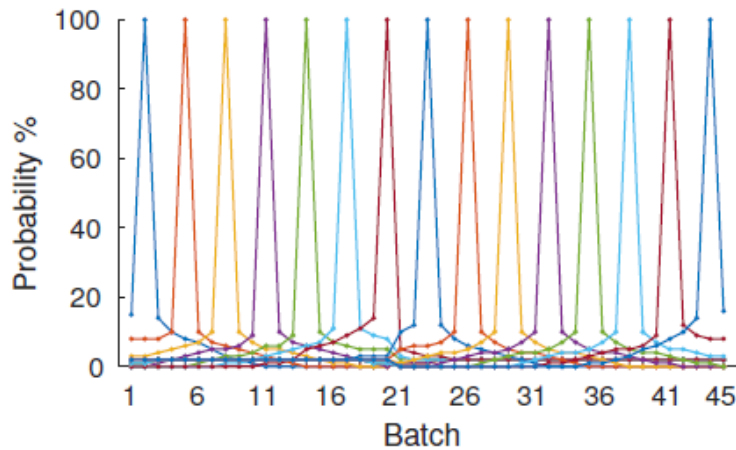
Figure 7. Probability of contamination for each batch given that a fixed batch is contaminated (hot). The fifteen separate curves correspond to Batches 2, 5, 8, ..., 41 and 44 being the hot batch.

batches immediately adjacent to the hot batch is typically about 10–15%. Batches further away have a decreasing chance of contamination but the probability does not drop to zero until about 8-10 batches away from the hot batch. This is primarily due to the spread distribution of carcasses in the raw sources, which, as noted above, is very concentrated near the centre. The fresh fat source changes from Source VII for batches 1–20 to Source VIII for batches 21–45, as shown in Table 8. Since much of the contamination probability is due to these fat sources (high in fat f_s and relatively high susceptibility factors g_s, Table 9), there is a noticeable asymmetric effect when the hot batch is near this switch. When batch 20 is the hot batch, batches prior to 20 show greater probability of also being contaminated than do batches after 20. This effect is present to a lesser degree when the hot batch is number 17, 18, or 19. Similarly, when the hot batch is 21, batches after 21 show a greater probability of contamination than those prior, and this effect is also seen to a lesser degree for hot batches 21, 22, and 23. There is also evidence of "long-range" contamination spread. As shown in Tables 10 and 11 and indicated in their captions, Batches 21–45 have a

TABLE 10. Probability of contamination for Batches 1–24 in percent for hot batches (columns) 1–20. Batches 25–45 (not shown) all had probability of contamination of 2 percent for hot Batches 1–20.

Batch	hot batch																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	100	15	12	10	8	6	5	3	1	1	0	0	0	0	0	0	0	0	0	0
2	16	100	14	10	8	7	5	3	2	1	0	0	0	0	0	0	0	0	0	0
3	12	14	100	12	8	7	5	4	3	2	1	0	0	0	0	0	0	0	0	0
4	10	10	12	100	10	7	6	5	4	3	2	1	0	0	0	0	0	0	0	0
5	9	8	8	10	100	10	7	5	5	4	3	2	1	0	0	0	0	0	0	0
6	7	7	7	7	10	100	10	7	6	5	4	3	2	1	0	0	0	0	0	0
7	5	5	5	6	7	10	100	10	7	6	5	4	3	2	1	0	0	0	0	0
8	3	3	4	5	5	7	10	100	10	7	5	5	4	3	2	1	1	0	0	0
9	2	2	3	4	5	6	7	10	100	10	6	5	5	4	3	2	1	1	0	0
10	1	1	2	3	4	5	6	7	10	100	9	9	9	7	6	6	5	5	4	5
11	0	0	1	2	3	4	5	5	6	9	100	10	10	6	6	5	4	3	2	1
12	0	0	0	1	2	3	4	5	5	9	10	100	10	6	6	5	4	2	2	1
13	0	0	0	0	1	2	3	4	5	9	10	10	100	9	7	6	5	4	3	2
14	0	0	0	0	0	1	2	3	4	7	6	6	9	100	10	7	6	5	5	5
15	0	0	0	0	0	0	1	2	3	6	6	6	7	10	100	10	7	7	6	6
16	0	0	0	0	0	0	0	1	2	6	5	5	6	7	10	100	11	8	7	7
17	0	0	0	0	0	0	0	1	1	5	4	4	5	6	7	11	100	11	9	9
18	0	0	0	0	0	0	0	0	1	5	3	2	4	5	7	8	11	100	13	11
19	0	0	0	0	0	0	0	0	0	4	2	2	3	5	6	7	9	13	100	14
20	0	0	0	0	0	0	0	0	0	5	1	1	2	5	6	7	9	11	14	100
21	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	3	4	4	5
22	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	3	4	4
23	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	3	3	3
24	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3	3

TABLE 11. Probability of contamination for Batches 1–4 and 16–45 in percent for hot batches (columns) 21–45. Batches 5–15 (not shown) all had probability of contamination of 2 percent for hot batches 21–45.

Batch	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2
3	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2
4	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
...																									
16	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
17	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
18	4	3	3	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2
19	4	3	3	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2
20	5	4	3	3	2	2	2	2	2	2	1	0	0	0	0	0	0	1	1	1	2	2	2	2	2
21	100	14	10	9	7	6	4	3	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	14	100	12	9	7	6	6	4	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	10	12	100	12	8	6	6	5	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	9	9	12	100	11	7	6	6	4	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	7	7	8	11	100	10	7	7	5	4	3	2	1	0	0	0	0	0	0	0	0	0	0	0	0
26	6	6	6	7	10	100	10	7	5	5	4	3	2	1	1	0	0	0	0	0	0	0	0	0	0
27	4	6	6	6	7	10	100	10	7	7	5	4	4	3	2	1	0	0	0	0	0	0	0	0	0
28	3	4	5	6	7	7	10	100	10	7	7	5	4	4	3	2	1	0	0	0	0	0	0	0	0
29	1	3	4	4	5	5	7	10	100	10	7	7	5	4	4	3	2	1	0	0	0	0	0	0	0
30	1	1	1	2	4	5	7	7	10	100	10	7	7	5	4	4	3	2	1	0	0	0	0	0	0
31	0	0	0	1	3	4	5	7	10	10	100	10	7	7	5	4	4	3	2	1	0	0	0	0	0
32	0	0	0	0	2	3	4	5	7	7	10	100	10	7	7	5	4	4	3	2	1	0	0	0	0
33	0	0	0	0	1	2	4	4	5	7	7	10	100	10	7	7	5	4	4	3	2	1	0	0	0
34	0	0	0	0	0	1	3	4	4	5	7	7	10	100	10	7	7	5	4	4	3	2	1	0	0
35	0	0	0	0	0	0	2	3	4	4	5	7	7	10	100	10	7	7	5	4	4	3	2	1	1
36	0	0	0	0	0	0	1	2	3	4	4	5	7	7	10	100	10	7	7	5	4	4	3	2	2
37	0	0	0	0	0	0	0	1	2	3	4	4	5	7	7	10	100	10	7	7	5	5	4	3	3
38	0	0	0	0	0	0	0	0	1	2	3	4	4	5	7	7	10	100	10	7	5	5	4	3	3
39	0	0	0	0	0	0	0	0	0	1	2	3	4	4	5	7	7	11	100	10	9	6	6	5	5
40	0	0	0	0	0	0	0	0	0	0	1	2	3	4	4	5	7	7	11	100	7	6	6	6	6
41	0	0	0	0	0	0	0	0	0	0	0	1	2	3	4	4	5	7	9	7	100	12	13	10	8
42	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	4	5	6	6	6	12	100	12	10	9
43	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	4	5	6	6	13	12	100	14	12
44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	4	5	6	10	10	14	100	16
45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	2	3	5	6	8	9	12	16	100

small (about 2%) probability of contamination when the hot batch is in the range 1–20, and similarly Batches 1–20 have a small probability of contamination when the hot batch is in the range 21–45. This long-range contamination spread is primarily due to the carcass overlap probability between Sources VII and VIII, $V_{78} = 0.2$. As mentioned above, Source VII is input to the first 20 batches and Source VIII to Batches 21–45. Since these sources have a relatively high probability of being contaminated, their overlap causes long-range contamination spreading.

4. **Conclusions.** We believe that the proposed model may help to reduce economic losses in the beef industry. To our knowledge, this is a novel method for estimating of the probability of *E. coli* contamination in the production of ground beef. Most of the input to the model is easily obtained from production records, or easily estimated (such as the average size of pieces). The most difficult input values to estimate are the within-source spread parameters, K, L, and $H^{\pm}$, and the across-source overlap parameters $V_{s_1 s_2}$. However, some knowledge of these values could be obtained through detailed observations of the plant processes or by using genetic sampling experiments in the raw sources as well as the final ground beef product. The model has been implemented in C and is available on the author's web site http://www.uoguelph.ca/~awillms/grbeefcontam/. A useful next step would be to use an innocuous surrogate for pathogenic *E. coli* that could be applied to pieces from a particular carcass and then identified in the final product batches. This would allow for better determination of carcass spread within a raw source and would provide data for direct validation of the model.

As a general statement, results from this model indicate that to restrict the spread of E. coli contamination, ground beef production processes should be designed to restrict the spread of pieces from a single carcass within a raw source, and limit the number of batches to which a given raw source provides input.

REFERENCES

[1] M. Aslam, G. G. Greer, F. M. Nattress, C. O. Gill and L. M. McMullen, Genotypic analysis of *Escherichia coli* recovered from product and equipment at a beef-packing plant, *Journal of Applied Microbiology*, **97** (2004), 78–86.

[2] M. Aslam, F. M. Nattress, G. G. Greer, C. Yost, C. O. Gill and L. McMullen, Origin of contamination and genetic diversity of *Eschericia coli* in beef cattle, *Applied and Environmental Microbiology*, **69** (2003), 2794–2799.

[3] R. G. Bell, Distribution and sources of microbial contamination on beef carcasses, *Journal of Applied Microbiology*, **82** (1997), 292–300.

[4] M. H. Cassin, A. M. Lammerding, E. C. D. Todd, W. Ross and R. S. McColl, Quantitative risk assessment for *Escherichia coli* O157: H7 in ground beef hamburgers, *International Journal of Food Microbiology*, **41** (1998), 21–44.

[5] D. A. Drew, G. A. Koch, H. Vellante, R. Talati and O. Sanchez, Analyses of mechanisms for force generation during cell septation in *Escherichia coli*, *Bulletin of Mathematical Biology*, **71** (2009), 980–1005.

[6] G. Duffy, F. Butler, E. Cummins, S. O'Brien, P. Nally, E. Carney, M. Henchion, D. Mahone and C. Cowan, E. coli *O157: H7 in Beef burgers Produced in the Republic of Ireland: A Quantitative Microbial Risk Assessment*, Technical report, Ashtown Food Research Centre, Teagasc, Dublin 15, Ireland, 2006.

[7] G. Duffy, E. Cummins, P. Nally, S. O'Brien and F. Butler, A review of quantitative microbial risk assessment in the management of *Escherichia coli* O157:H7 on beef, *Meat Science*, **74** (2006), 76–88.

[8] E. Ebel, W. Schlosser, J. Kause, K. Orloski, T. Roberts, C. Narrod, S. Malcolm, M. Coleman and M. Powell, Draft risk assessment of the public health impact of *Escherichia coli* O157:H7 in ground beef, *Journal of Food Protection*, **67** (2004), 1991–1999.

[9] P. M. Kitanov and A. R. Willms, Estimating *Escherichia coli* contamination spread in ground beef production using a discrete probability model, in *Mathematical and Computational Approaches in Advancing Modern Science and Engineering* (eds. J. Bélair, I. F. I, H. Kunze, R. Makarov, R. Melnik and R. Spiteri), Springer, AMMCS-CAIMS 2015, Waterloo, Canada, 2016, 245–254.

[10] A. Rabner, E. Martinez, R. Pedhazur, T. Elad, S. Belkin and Y. Shacham, Mathematical modeling of a bioluminescent *E. Coli* based biosensor, *Nonlinear Analysis: Modelling and Control*, **14** (2009), 505–529.

[11] E. Salazar-Cavazos and M. Santillán, Optimal performance of the tryptophan operon of *E. coli*: A stochastic, dynamical, mathematical-modeling approach, *Bulletin of Mathematical Biology*, **76** (2014), 314–334.

[12] E. Scallan, R. M. Hoekstra, F. J. Angulo, R. V. Tauxe, M. A. Widdowson, S. L. Roy, J. L. Jones and P. M. Griffin, Foodborne illness acquired in the United States – Major pathogens, *Emerging Infectious Diseases*, **17** (2011), 7–15.

[13] M. Signorini and H. Tarabla, Quantitative risk assessment for verocytotoxigenic *Escherichia coli* in ground beef hamburgers in Argentina, *International Journal of Food Microbiology*, **132** (2009), 153–161.

[14] B. A. Smith, A. Fazil and A. M. Lammerding, A risk assessment model for *Escherichia coli* O157:H7 in ground beef and beef cuts in canada: Evaluating the effects of interventions, *Food Control*, **29** (2013), 364–381.

[15] J. Turner, R. G. Bowers, M. Begon, S. E. Robinson and N. P. French, A semi-stochastic model of the transmission of *Escherichia coli* O157 in a typical UK dairy herd: Dynamics, sensitivity analysis and intervention prevention strategies, *Journal of Theoretical Biology*, **241** (2006), 806–822.

[16] J. Turner, R. G. Bowers, D. Clancy, M. C. Behnke and R. M. Christley, A network model of *E.coli* O157 transmission within a typical UK dairy herd: The effect of heterogeneity and clustering on the prevalence of infection, *Journal of Theoretical Biology*, **254** (2008), 45–54.

[17] J. Tuttle, T. Gomez, M. P. Doyle, J. G. Wells, T. Zhao, R. V. Tauxe and P. M. Griffin, Lessons from a large outbreak of *Escherichia coli* O157:H7 infections: Insights into the infectious dose and method of widespread contamination of hamburger patties, *Epidemiology and Infection*, **122** (1999), 185–192.

[18] X. Wang, R. Gautam, P. J. Pinedo, L. J. S. Allen and R. Ivanek, A stochastic model for transmission, extinction and outbreak of *Escherichia coli* O157:H7 in cattle as affected by ambient temperature and cleaning practices, *Journal of Mathematical Biology*, **69** (2014), 501–532.

[19] X. Yang, M. Badoni, M. K. Youssef and C. O. Gill, Enhanced control of microbiological contamination of product at a large beef packing plant, *Journal of Food Protection*, **75** (2012), 144–149.

[20] X. S. Zhang, M. E. Chase-Topping, I. J. McKendrick, N. J. Savill and M. E. J. Woolhouse, Spread of *Escherichia coli* O157:H7 infection among Scottish cattle farms: Stochastic models and model selection, *Epidemics*, **2** (2010), 11–20.

E-mail address: pmksma@rit.edu

E-mail address: AWillms@uoguelph.ca