Food Control xxx (2017) 1–7

Contents lists available at ScienceDirect

Food Control

journal homepage: www.elsevier.com/locate/foodcont

Application of Bayesian Networks in the development of herbs and spices sampling monitoring system

Yamine Bouzembrak [a, *], Louise Camenzuli [a], Esmée Janssen [b], H.J. van der Fels-Klerx [a]

[a] RIKILT Wageningen University and Research, Akkermaalsbos 2, 6708 WB, Wageningen, The Netherlands
[b] Wageningen University and Research, Department of Social Sciences, Hollandseweg 1, 6706 KN, Wageningen, The Netherlands

ARTICLE INFO

Article history:
Received 30 November 2016
Received in revised form
10 April 2017
Accepted 12 April 2017
Available online xxx

Keywords:
Prediction
Food chain
Food safety hazards

ABSTRACT

Knowing which products and hazards to monitor along the food supply chain is crucial for ensuring food safety. In this study, we developed a model to predict which types of herbs and spices products and food safety hazards should preferentially be monitored at each level of the supply chain (suppliers, border inspection points, market and consumers). A Bayesian Network method was used to develop a model based on notifications reported in the Rapid Alert System for Food and Feed and the database of the Dutch national monitoring program for chemical contaminants in food and feed over the period 2005–2014. The model was constructed by randomly selecting ca. 80% of the 3126 data records and validated using the remaining ca. 20% of the records. Model validation showed that the prediction accuracy was higher than 85%. Results showed that the sampling plan is closely related to the place where the products are checked along the supply chain, the products and the country of origin. Our approach of integrating different data sources and considering the entire supply chain can support industry and authorities at border inspection points and at all control points along the herbs and spices supply chain in setting priorities for their monitoring program.

1. Introduction

The herb and spice supply chain has developed from local or national production and trade to more global relationships between producers, manufacturers, processors and retailers across worldwide. The global market provides the consumer with access to a variety of products, at cheap prices. However, this market accessibility could increase the probability for product contamination and associated health risks. Herbs and spices may be contaminated with pathogens like *Salmonella*, chemicals like mycotoxins, pesticides and heavy metals; and with physical contaminants like pieces of glass. For example, contamination of spices with aflatoxins has been reported in several countries worldwide, namely Turkey (Aydin, Emin Erkan, Başkaya, & Ciftcioglu, 2007), Hungary (Fazekas, Tar, & Kovács, 2005), Malaysia (Jalili, Jinap, & Radu, 2010), Spain (Santos, Marín, Sanchis, & Ramos, 2010), India (Reddy, Mayi, Reddy, Thirumala-Devi, & Reddy, 2001), Pakistan (Paterson, 2007) and Sri Lanka (Yogendrarajah et al., 2014).

In the European Union (EU), official food and feed controls for Member States are defined in Regulation (EC) 882/2004, which ensures the effective implementation of Regulation (EC) 178/2002. According to Regulation (EC) 882/2004, each Member State designs annual plan for their official controls. These plans should be carried out with a “risk-based” approach at all stages of production, processing, and distribution. Controls at the “point of entry” from third countries are included in Regulation (EC) 882/2004; however, additional controls are defined in Regulation (EC) 669/2009. According to Regulation (EC) 669/2009, the frequency of official control of a particular food (or feed) item from a specific country of origin is defined and is enforced in every Member State. For example, 50% of all consignments of mint imported into the EU from Vietnam should be subjected to official controls for pesticide residues, irrespective of the point of entry. Regulation (EC) 669/2009 is updated periodically in accordance with new information on known (or emerging) risks.

Furthermore, Regulation (EC) 396/2005 Art. 29 sets up a community control program for the official controls of pesticide residues in food, which are to be included in national control programs. In these national control programs, specific pesticide/food combinations including the number of samples that should be analysed

* Corresponding author. RIKILT Wageningen University and Research, Akkermaalsbos 2, 6700 AE, Wageningen, The Netherlands.
E-mail address: yamine.bouzembrak@wur.nl (Y. Bouzembrak).

http://dx.doi.org/10.1016/j.foodcont.2017.04.019
0956-7135/

per Member State are annually defined. As of the end of 2016, spices and/or herbs have not been included in Member State national control programs. Additionally, according to Regulation (EC) 396/2005 Art. 30, the criteria for national control programs for pesticide residues are defined.

Regulation (EC) 178/2002 establishes a Rapid Alert System, which aims to support food and feed safety on the European market. Under the Rapid Alert System for Food and Feed (RASFF), members exchange notifications regarding measures taken with regards to food safety. The notifications reported in RASFF are available through the RASFF portal, which features an interactive searchable online database (RASFF, 2017).

The aim of this study was to prioritize which herb and spice products and hazards should be monitored along the supply chain (suppliers, border inspection points, market and consumers). Notifications reported in RASFF and data from the database of the Dutch national monitoring program for chemical contaminants in food and feed (KAP) were used in order to construct a BN model. The BN model was validated using an independent dataset and the validation showed that the prediction accuracy was high.

2. Materials and methods

For the objectives of this study, a BN model was developed to predict the presence of food safety hazards in herbs and spices, per product and country of origin, in order to guide risk-based monitoring. The proposed approach (Fig. 1) consists of three steps: (i) data collection on the main factors (in this case: hazard, product, and country of origin) from the following data sources: including RASFF, KAP, Eurostat, the EU pesticide database and legislation; (ii) construct a BN model including nodes, arrows, states, and the parameters for each node in the form of conditional probability tables (CPTs); and (iii) validate the model using an independent dataset.

2.1. Data collection

All notifications reported in the RASFF database under the hazard category "all hazards" and product category "herbs and spices" were extracted for the period 01/01/2005 to 31/12/2014. In the period between 2005 and 2014, 2223 notifications were reported in RASFF under the product category "herbs and spices" for all hazards.

In RASFF a notification may cite multiple hazards within the same hazard category, may report different levels of contamination, or may cite multiple products within the same notification (i.e. in this case the notification was duplicated for each product). The difference between the total number of notifications reported in RASFF portal (i.e. 1672 notifications) and total number used in the model (i.e. 2223 notifications) was due to such overlapping hazards/products/results within a notification.

Each RASFF record contained the following information: notification ID, date of notification, notifying country, classification, notification type, action taken, distribution status, product, product category, substance/hazard, control point (Table S1), country concerned (distribution, origin) and compliance to the legislation (RASFF, 2017).

Fig. 1. Method steps.

Similarly, records in the national monitoring results from the Netherlands (KAP) under the product category "Herbs and spices" and hazard category "mycotoxins" for the period 01/01/2005 to 31/12/2011 were retrieved. KAP is the database of the Quality of Agricultural Products program, which is hosted by the National Institute for Public Health and the Environment (RIVM) in Bilthoven in the Netherlands (van der Fels- Klerx & Bouzembrak, 2016). This database contains the results of the national monitoring program for chemical contaminants in agricultural products in the Netherlands (RIVM, 2017).

Each record contained the following information: reference number, date, product category, product name, hazard, concentration, country of origin, country of notification, control point (Table S2), legislation level, and compliance to legislation (Table S3).

Each notification was included in the model with additional data on trade volumes and the related legislation. Trade volume data of herbs and spices were retrieved from the EUROSTAT database. The maximum levels (MLs) for certain contaminants in herbs and spices, namely cadmium and lead for herbs and Aflatoxins and Ochratoxin A (OTA) for spices, were collected from Regulation (EC) 1881/2006 (consolidated version 11/03/2016). The maximum residue levels (MRLs) of pesticides in herbs and spices were listed in Regulation (EC) 396/2005 (consolidated version 28/10/2015). Also, the EU pesticide database was used to identify the active substances approved in Europe, references to EU legislation for each substance and MRLs in herbs and spices (European Commission, 2016).

2.2. Bayesian network

BNs are a class of probabilistic models originating from Bayesian statistics and decision theory combined with graph theory (Bonafede & Giudici, 2007; Nielsen and Jensen 2007). The structure of a BN consists of nodes (*i.e.*, random variables) connected by directed arcs reflecting a dependence structure between the nodes. BNs are applied in many domains such as hazard and risk assessment (Billoir, Delignette-Muller, Péry, & Charles, 2008; Iorio, Tagliaferri, & di Bernardo, 2009), fraud detection (Kirkos, Spathis, & Manolopoulos, 2007; Ngai, Hu, Wong, Chen, & Sun, 2011), accident prevention (Mohammadfam, Ghasemi, Kalatpour, & Moghimbeigi, 2017), food fraud prediction (Bouzembrak & Marvin, 2016; Marvin et al., 2016), nanomaterials (Marvin et al., 2017; Murphy et al., 2016) and medical image analysis (Arias, Martínez-Gómez, Gámez, Seco de Herrera, & Müller, 2016). BNs have the ability to integrate different data sources and types such as expert knowledge, measurement data and feedback experience (Buriticá & Tesfamariam, 2015) using Bayes theorem. The probability of event X_i at the condition of event X_j, $P(X_i|X_j)$, is expressed as:

$$P(X_i|X_j) = \frac{P(X_j|X_i) \times P(X_i)}{P(X_j)}$$

where $P(X_i)$ is the prior probability of event X_i and $P(X_j|X_i)$ is the probability of X_j under the condition of a known event X_i and $P(X_j)$ is the probability of X_j. BNs contain: (i) a set of variables (nodes) $U = \{X_1, \ldots, X_n\}$ and a set of directed edges between variables; (ii) each discrete variable with a finite set of states; and (iii) and a set of conditional and unconditional probabilities. If there is an edge from node X_i to node X_j, the node X_i is called the parent of node X_j and the node X_j is called the child of node X_i.

The joint probability distribution of all variables, $P(U) = P(X_1, \ldots, X_n)$, given by the product of all conditional probability tables specified in BN: