International food imports: Using visualization methods to identify vulnerabilities and risks

By Stephen M. Perrine
Adviser J Blanford

Abstract
Food vulnerabilities either through theft or contamination are on the rise and a growing concern. Geographic information systems and visual analytics software are tools that may be used to help detect patterns associated with food vulnerabilities and hence improve risk assessments of food imports. The objective of this study is to assess different methodologies that can be used to improve our understanding of food distribution patterns in the USA and predict port of entry. To test this, five years of Canadian exports of two food commodities, one perishable and one shelf stable, shipped by truck to the U.S. were examined to compare overall shipping patterns and to determine if least cost path analysis can predict the port of entry a given delivery route will use based on the source and destination locations. A least cost path is expected when shipping food items because of fuel costs and product quality loss for perishable items. Cost surfaces based on distance, time, and timeslope were created to run LCPA and were able to accurately predict the port of entry crossed 28.4%, 30.1% and 29.9% of the time, respectively, for perishable and non-perishable foods combined. The LCP routes do, however, closely follow known GPS tracked trucking routes (up to an 81% overlap).

Keywords
Bioterrorism, Imported Food, Information Visualization, Least Cost Path, Truck Transportation

1.0 Background

Imported food poses risks to the American economy and consumer in terms of food safety, fraud and terrorism. Total U.S. food imports increased 263% between 1999 and 2012 [1]. During 2012 the U.S. imported nearly $114 billion worth of food equating to approximately 62 million metric tons including fish/shellfish, fruits, beverages and vegetables [1,2]. Every year, an estimated 9.4 million people are affected by unintentionally caused foodborne illness in the U.S. and the number may be higher because many cases are not reported or diagnosed [3]. Food fraud poses significant public health risks because the adulterant or substituted product is often unconventional and difficult to detect [4], cannot be identified by reviewing documents and usually requires "state of the art" laboratory analysis [5], and health responses to the adulterant or substituted product is unpredictable.

It would be financially and physically impossible to inspect and sample every imported food shipment before releasing into U.S. commerce. In fact, less than three percent of food imports are routinely inspected for filth, decomposition, antibiotics, and pesticides [6]. With regard to invasive crop pests, the U.S. Department of Agriculture's Food Safety Inspection Service inspects all high risk products, but only inspects 2%-3% of the boxes on each truck [7]. This very low inspection and sampling rate means that there is a high likelihood that contaminated food will enter the U.S. undetected.

Anyone importing food for human or animal consumption, food additives or dietary supplements are required to provide the U.S. Food and Drug Administration (FDA) with advance notice ("prior notice") of what will be imported prior to arriving in the United States for risk assessment purposes ("Bioterrorism Act," 2002). One prior notice is required for each import line (i.e. unique food item) per entry and includes information about the consignee, physical location of the product manufacturer and the declared U.S. port of entry. Performing risk assessments on each prior notice submission takes time resulting in a small percentage of risk assessments being conducted (approximately 80,000 per year). For example during 2012, with an average of 30,000 prior notices per day to process, the FDA Division of Food Defense Targeting (DFDT) only performed risk assessments on a small percentage (0.71%) of all submissions.

Risk assessments of entry data are processed by the FDA with the aid of a computer program called PREDICT (Predictive Risk-based Evaluation for Dynamic Import Compliance Targeting) to target high-risk FDA-regulated shipments for sampling [8]. The benefit of this analysis is that the program evaluates food product risk and firm violation history to identify items that should be examined, sampled, referred for compliance action or cleared for entry into commerce [9]. Even if PREDICT was applied for prior notice screening, the program relies upon complete and accurate data and a history of shipping for pattern analysis. Prior notice submissions often contain missing and/or inaccurate data, first time entities, and non-commercial entities that make electronic screening difficult. Not only does incomplete data affect the quality of the risk assessment, but so does the allotted time required to conduct risk assessments. This is dependent on the mode of transportation by which the items arrive in the U.S., ranging from eight hours for sea shipments to only two hours for road shipments [10]. With such a high volume of prior notices submitted each year and extremely short times for making risk assessments, additional methods must be developed to ensure that the highest risk shipments are targeted for DFDT review and FDA/Customs and Border Protection (CBP) field inspection.

Understanding potential sources of risk of food contamination are important. Ideally, the use of geospatial technologies such as GPS and radio frequency identification (RDIF) tags [11] can provide real time monitoring of transportation routing of food products during all stages of the supply chain from point of origin (source) to the final destination [7,12]. Such intelligent transport systems would be useful for traceability of the producer [7,12-14], understanding transportation routing, provide security information by identifying vulnerabilities along routes (i.e. unattended cargo that may be vulnerable to opportunistic terrorist attack/contamination) [15-17], record unanticipated route deviations or extended stops [11] and identify multiple attempts to enter the U.S. after being denied (i.e. port-shopping [4]). However, this level of detail is not available in the existing data and therefore other methods are necessary to help understand movement of food products and vulnerabilities en route. Therefore improving risk assessment methods are vital to minimize food-related risks whether unintentional or not.

Although, methods that evaluate food product risk and firm violation history are already in use, additional methods that combine this information with visualization and spatial analytical methods are necessary to improve efficiency in screening of prior notices and enhance analysis of shipping patterns to detect abnormalities and is the key objective of this study. In this study, data of vegetable produce (representing perishable goods) and coffee/tea (representing non-perishable goods) imported into the U.S. from Canada was analyzed to (i) gain an overview of geographic diversity of products (ii) identify shipping patterns and (iii) determine most efficient routes to predict likely routes and port of entry.

2.0 Methods

In this section, details are provided on the data obtained for this analysis, any steps taken to prepare data for analysis, and specific steps taken to perform spatial analyses and visualizations.

2.1 Food Import data

Food import data for both perishable goods (e.g. vegetable/vegetable products that are raw, ambient and raw, refrigerated) and non-perishable goods (coffee/tea) were obtained from the FDA ORADSS (Office of Regulatory Affairs Reporting, Analysis and Decision Support System) database for trucks traveling from Canada to the U.S. between 2008 and 2012. A total of 138,693 unique origin-port of entry-destination (OPD) shipping routes representing 2,378,172 import lines were extracted. Data included Canadian manufacturers in a given city (source), the U.S. port of entry (POE) crossed, and the U.S. consignees in a given city (destination) for each food type in a given year. Once the data was extracted, it was necessary to remove 1,176 shipping routes representing 186,148 import lines for one or more of the following reasons: missing manufacturer city, missing consignee city, missing POE, manufacturer not in Canada, consignee was in Hawaii or Puerto Rico, manufacturer address is too ambiguous to determine, port of entry is not on the U.S. northern border. In addition, the data did not contain geographic coordinates therefore it was necessary to geocode the data. To prepare for geocoding, the data were cleaned by correcting misspellings, spelling out all abbreviations (i.e., Saint and Sainte), and changing Canadian Provinces when indicated by the city and postal code. A total of 721 Canadian source cities and 4,314 U.S. destination cities were run through Navteq address locaters (Source: [18]). Matches were accomplished by matching City, Province with Canadian Administrative Place names and City, State with U.S. Administrative Place names with a match success of 92.1% (Canada) and 95.5% (U.S.). Coordinates for unmatched cities were manually geo-referenced using Bing Maps (http://www.bing.com/maps).

2.2 Analysis

2.2.1 Data reduction

A total of 24,689 unique origin-destination (OD) city pairs were identified. Since it is impossible to analyze these in a timely manner the data was aggregated. To do so, all geocoded cities were assigned to a 100 km^2 grid of the Military Grid Reference System (http://earth-info.nga.mil/GandG/coordsys/grids/mgrs_100km_dloads.html) and the centroid of each grid was used for further analysis. Areas that overlapped bodies of water (10m ocean and 10m lakes dataset obtained from http://www.naturalearthdata.com/downloads/) were clipped and centroids assigned after these areas were removed. By using the grid reference system, the number of origin-destination pairs was reduced to 4,618. Data were separated by perishable foods (vegetable/vegetable products) and non-perishable foods (coffee/tea) and only those origin-destination grid pairs in common to both food types were retained for analysis. A total of 531 origin-destination pairs were used for the remainder of this analysis.

2.2.2 Visualization of food movement. Visualization methods ranging from cartography, time geography, information visualization and geovisualization (see [19] for overview) have become increasingly useful for pattern detection including understanding movement flows. For example, cartographic flow maps are useful for visualizing movement between locations using arrows of varying thickness to indicate direction and quantity. Although useful, when large amounts of data are involved, the arrows may obscure one another and make the map difficult to interpret. To overcome this, aggregation and summarization methods [20], with trajectories aggregated either by similarities in origin and destination or by route have been useful in reducing the display clutter of large movement datasets and highlighting key routes between locations. Information visualization techniques such as tree diagrams and origin-destination (OD) matrices can graphically represent flow map data but lack the benefit of spatial context [19]. Other visualization techniques such as interactive histograms, scatter plots, and parallel plots permit detailed exploration of data, particularly when linked to maps [21]. These methods have the advantage of reducing information complexity and revealing patterns and relationships that can be more easily interpreted and are not possible with static maps [22]. Geovisualization further supports geospatial analysis through interactive visualization tools that facilitates pattern detection, association, and analytical reasoning [23]. Since the goal of this study is to examine shipping patterns, diversity and establish a likely port of entry, a variety of visualization and spatial analysis methods were used.

Visual analytics were used to understand the port of entry associated with each source location in Canada as well as port of entry associated with destination locations in the U.S. Specifically, heat maps and treemaps were created using Spotfire 6.5.0 (TIBCO Software Inc., 2014 (spotfire.tibco.com)) and connections mapped using ESRI ArcGIS 10.2 (http://esri.com).

A treemap, like a tree graph is a two-dimensional space-filling approach where each node is represented by a rectangle whose area is proportional to the value of an attribute [24]. This method has been used to visualize network traffic [25] and for spatio-temporal visualizations [26]. To create the treemaps, the attribute data used represented the total number of trucks travelling from a source location in Canada to a destination in the USA via a port of entry. A hierarchy was created based on the sum of origin-port and port-destination occurrences resulting in a nested set of squares whose color, size, and position represent the number of times a particular port is found associated with a given grid in the data. Spotfire places the largest square in the upper left corner and the smallest square in the lower right corner through the use of a squarified algorithm.

Heat maps were used to provide a visual summary of the density of trucks moving through each space (origin, POE and destination). Similar in design to a spreadsheet, the value of each space is a color representing a quantity based on a gradient between the highest and lowest values in a dataset. Heat maps are useful for visualizing large datasets and identifying hotspots and attributes with similar values based on color. This method has successfully been used to capture space-time movements [19]. Similar to the treemap, the heat maps were created by aggregating the data to capture the number of times trucks shipped from a given location in Canada through a POE and the number of times trucks continued from each POE to a given U.S. destination.

2.2.3 Truck Routing and Predication of Port Crossings. The connections were also visualized on a map by capturing truck movements and connecting the source and destination points. To do so the total number of trucks traveling between each source-destination, source-port of entry and port of entry-destination was calculated.

Once we analyzed the general movement patterns of each food type between Canada and the U.S., we next estimated the likely port of entry by determining likely route choice between the source and destination. Several routing methods can be used to identify least-cost routes. Least Cost Path Analysis (LCPA), a common GIS method, has been successfully used to identify the most cost efficient route in a variety of studies ranging from transportation [27,28], invasive species management [29], biodiversity, pipeline and power line transmission routing [30] and trail planning [31] by integrating a road network that includes different road types (e.g. road types such as highways, major streets, minor streets, trails, etc.) with an impedance factor that captures factors that may affect travel time (e.g. slope and elevation). In transportation, although LCPA has frequently been used for planning purposes it has also successfully predicted (with 90% accuracy) the most-likely trucking routes that will be taken by a driver [32]. Another popular method used to model vehicle routes, particularly with the goal of reducing drive time and fuel consumption, is known as the vehicle routing problem (VRP) [33,34]. This method is more complex than the LCPA since it requires more detailed information such as traffic density, time of delivery, traffic lights, travel speed and stop signs. Therefore for the purpose of this study the LCPA was used.

To estimate travel routes three friction surfaces representing the cost of a truck moving across a surface were created that capture travel based on (i) the shortest distance, (ii) travel time and (iii) enhance travel time that includes a reduction in travel time due to slope. The cost-distance function in ArcGIS 10.2 Spatial Analyst was used to calculate the path with the least cost by calculating the linear and diagonal accumulated least cost of getting to the nearest source from each grid cell [35].

To examine potential truck routes based on distance alone, road data was obtained from the Homeland Security Infrastructure Program (HSIP) Gold data (2013) and contained 769,437 segments of interstates, highways, major roads and streets. Roads were rasterized using a cell size of 1000m where each cell represents 1km. Since we were only interested in distance, each cell was assigned the same value of 1 and used to find the shortest distance between each location. Throughout the rest of this paper, the friction surface and associated outputs will be referred to as the distance grid. Next we examined travel routes based on time. Each road type was assigned a travel speed value based on road type ranging from 1 to 8 (**Table 1A**), where 1 represents roads where travel speed is fastest (e.g. Highways) and 8 represents roads where travel speed is slowest (e.g. roundabout). To capture this, the roads dataset was rasterized using a cell size of 1000m and each road type was assigned the values in Table 1A such that each cell represented the time it takes to travel across 1km. This surface will be referred to as the time grid. Lastly, speed of travel can be affected by slope (see [36] and references therein) therefore to account for this the time grid was further enhanced by taking slope into account. To do so travel speed values of the time grid were multiplied with slope values (**Table 1B**, where roads with little to no slope (e.g. < 3%) were assigned a low value of 1 and steep slopes (e.g. > 12%) were assigned a value of 5. Thus, when combined with the travel speed value a steep slope would reduce the travel speed. For example, when a road speed of 1 was combined with a steep slope (value = 5) the new road speed value would become 5. The slope surface was created by calculating the percent slope for elevation for North

America (data obtained from the USGS (https://lta.cr.usgs.gov/GTOPO30 at 1km resolution) in ArcGIS 10.2 using the slope command. Each slope value was then reclassified into five categories using the values listed in Table 1B. The friction surface and associated outputs will be referred to as the timeslope grid. The least cost path was calculated between each origin-destination pair representing 20 Canadian origin grid centroids and each of its corresponding U.S. destinations using all three friction surfaces, resulting in the creation of 531 unique routes in common to both food types.

Table 1:(A) Relative travel speed and (B) slope values used to create the LCPA friction surfaces to calculate commercial truck shipping routes.

(A) Road type	Value	(B) Slope (%)	Value
Highway	1	**0-3**	1
Interstate	2	**3-6**	2
Major road	3	**6-9**	3
Street	4	**9-12**	4
Ramp	6	**> 12**	5
Ferry	7		
Roundabout	8		

2.2.4 Evaluation. To ascertain the ability for LCPA to capture potential truck routes and port crossings, outputs from each LCPA route were compared to the known port of entry and truck routes compiled from GPS trucking data [37]. The port of entry each LCPA path crossed, or came closest to (within 1 km) due to the rasterization of the road layer was recorded and compared to the port of entry recorded for that route by food type (e.g. all and perishable vs. non-perishable). Since we do not have actual routes taken by each truck, we assessed the accuracy of routes by comparing the routes with the actual truck routes using GPS trucking data. To do so we selected the output created using distance, time and timeslope for one route (source = Quebec, CA (18TWR); destinations = California (10SEG, 11SLT, 11SMS, 11SMT), Florida (17RLL, 17RLN, 17RMP, 17RNJ, 17RNK), Georgia (17RKQ), Illinois (16TDM), Maryland (18SUH, 18SUJ), Massachusetts (19TCG), Michigan (16TEN, 16TGN, 17TLG), Missouri (15SUD), New Hampshire (19TCH), New Jersey (18TWK, 18TWL), North Carolina (17SQV), Oklahoma (15SUA), Pennsylvania (17SLU, 18SUK, 18TVK), South Carolina (17SLU), Texas (14SPB, 14SQB), Vermont (18TXQ), Virginia (18STJ, 18SVG), Washington (10TET)) and intersected these with GPS truck routes. GPS trucking data were obtained from Transport Canada for one month (see [37] for details) and used to create known trucking routes. Points were rasterized to create routes and then converted to lines. Each route was buffered at 0.5 km and used to intersect with each route. To assess routing accuracy total distance of the route was compared with the total distance of the route that intersected with the buffered truck route.

3.0 Results

3.1 Import of perishable and nonperishable food items from Canada to the U.S. (2008-2012)

A total of 531 OD pairs represent 83,415 shipments and 1,293,380 import lines. The numbers of origin and destination cities were reduced from 721 to 20 and 4,314 to 199, respectively (**Figure 1**). The corresponding number of shipments originating from each of the 20 Canadian points of origin and arriving at one or more of the 199 U.S. destination grids passing through the U.S. ports of entry are illustrated in **Figures 1 and 2**. Only 28 of 127 available ports along the U.S./Canadian border were used by food shipments of vegetable/vegetable products (**Figure 2A**) and coffee/tea products entered the U.S. at 16 ports of entry (**Figure 2B**).

Figure 1: Number of shipments originating in Canada (red) and arriving at destinations in the U.S. (blue) following a data reduction step whereby all locations were assigned to one 100 km^2 MGRS grid (inset).

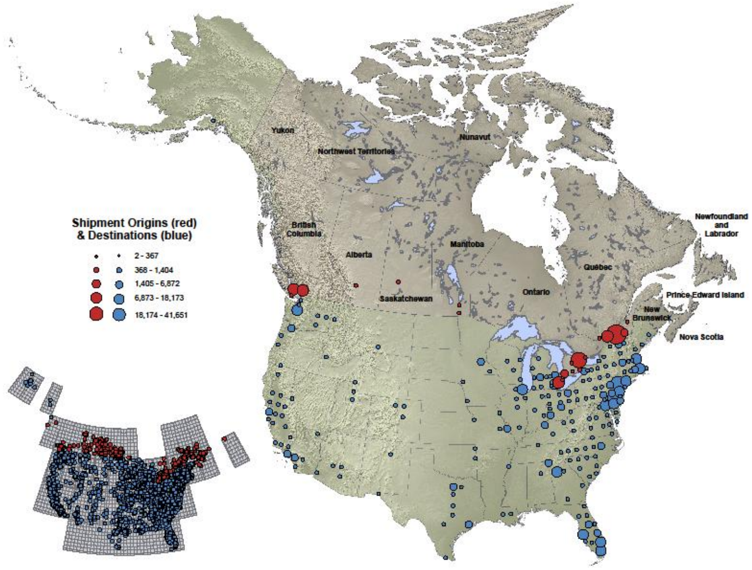
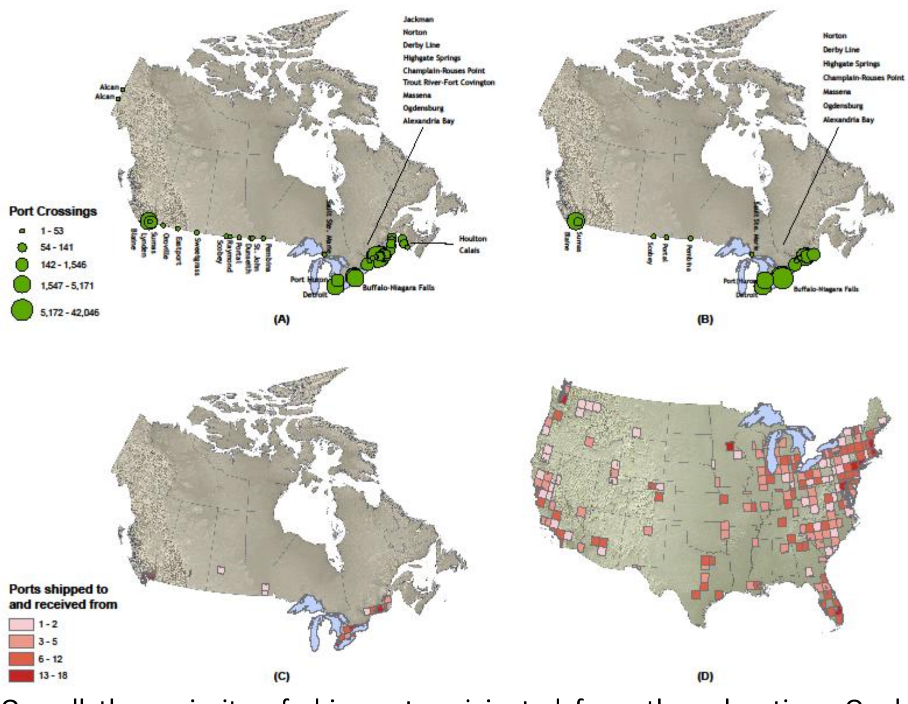


Figure 2: Distribtion of U.S. ports of entry for shipments of (A) vegetables/vegetable products and (B) coffee/tea. Number of ports of entry (C) shipped to from Canadian origins and (D) received from by U.S. destinations for both perishable and non-perishable food products.



Overall the majority of shipments originated from three locations Quebec (18TXR, 49.9%), Ontario (17TPJ, 21.8%), and British Columbia (10UDV, 8.2%) (**Table 2, Figure 1**) for both food types and entered the U.S. through 5 ports. These included 712-Champlain-Rouses Point, New York (52.5%), 901-Buffalo/Niagara Falls, New York (16.3%), 3004-Blaine, Washington (11.0%) and 3801-Detroit, Michigan (10.1%) (**Table 3, Figure 2**). Primary U.S. destinations included New Jersey (18TWL, 17.4%), Maryland

(18SUJ, 6.4%), Massachusetts (19TCG, 6.4%), Florida (17RNJ, 4.1%), and Connecticut (18TXL, 4.1%) (**Figure 1**).

Table 2: The total number of truck shipments and import lines exported from Canada between 2008 and 2012.

MGRS Grid & Province*	Vegetables/Veg Products		Coffee/Tea		Total	
	Shipments	Lines	Shipments	Lines	Shipments	Lines
10UDU-BC^{02}	9	30	9	12	18	42
10UDV-BC^{15}	2,292	65,320	4,576	26,384	6,868	91,704
10UEV-BC^{15}	3,559	142,740	711	12,358	4,270	155,098
11UQS-AB^{02}	1	1	1	2	2	3
13UCT-SK^{02}	2	4	4	6	6	10
14UPA-MB^{01}	1	1	-	-	1	1
14UPV-MB^{02}	3	3	1	1	4	4
17TLG-ON^{14}	3,167	203,725	59	624	3,226	204,349
17TMG-ON^{05}	41	1241	11	277	52	1,518
17TMH-ON^{09}	1,082	22,907	322	3463	1,404	26,370
17TNH-ON^{05}	182	5,084	43	407	225	5,491
17TNJ-ON^{10}	256	14,620	46	97	302	14,717
17TPH-ON^{08}	310	4,505	57	164	367	4,669
17TPJ-ON^{18}	4,646	61,548	13,522	242,611	18,168	304,159
17TQJ-ON^{01}	2	3	1	1	3	4
18TVR-QC^{02}	16	37	4	21	20	58
18TWR-QC^{12}	5,481	29206	113	390	5,594	29,596
18TXR-QC^{17}	38,142	281,440	3,507	144,050	41,649	425,490
19TBL-QC^{08}	625	11,723	391	16,992	1,016	28,715
19TCM-QC^{04}	211	1344	9	38	220	1,382
Sum	**60,028**	**845,482**	**23,387**	**447,898**	**83,415**	**1,293,380**

*Superscript indicates the number of ports used by shipments originating from the MGRS grid.

Table 3: The total number of truck shipments and import lines entering the U.S. by port of entry between 2008 and 2012.

U.S. Port of Entry Code	Vegetables/ Vegetable Products		Coffee/Tea		Total	
	Shipments	Lines	Shipments	Lines	Shipments	Lines
0104	141	1,959	-	-	141	1,959
0106	85	323	-	-	85	323
0115	6	20	-	-	6	20
0209	627	11,749	223	3,425	850	15,174
0211	131	779	158	5,686	289	6,465
0212	114	443	177	5,904	291	6,347
0701	13	18	4	21	17	39
0704	3	9	43	172	46	181
0708	622	2,039	395	3,891	1,017	5,930

0712	42,046	306,110	1,708	97,009	43,754	403,119
0715	17	21	-	-	17	21
0901	4,152	79,260	9,462	173,732	13,614	252,992
3004	4,062	133,063	5,111	37,871	9,173	170,934
3009	1,546	73,330	128	1,608	1,674	74,938
3019	19	33	-	-	19	33
3023	2	7	-	-	2	7
3104	1	1	-	-	1	1
3301	1	1	-	-	1	1
3302	1	1	-	-	1	1
3309	6	16	1	1	7	17
3310	11	35	-	-	11	35
3401	53	123	51	230	104	353
3403	32	93	1	1	33	94
3405	1	1	-	-	1	1
3422	1	2	-	-	1	2
3801	5,171	224,836	3,215	64,967	8,386	289,803
3802	1,156	11,197	2,704	53,346	3,860	64,543
3803	8	13	6	34	14	47
Sum	**60,028**	**845,482**	**23,387**	**447,898**	**83,415**	**1,293,380**

When food types were examined separately, there were noticeable differences in shipping patterns. Sixty-three percent of all vegetables/vegetable products originated from Quebec (18TXR), entering the U.S. through port 712-Champlain-Rouses Point, New York (70.0%), and were transported to three key areas that included New Jersey (18TWL, 20.9%), Massachusetts (19TCG, 8.4%), and Maryland (18SUJ, 7.5%) (**Table 2, Table 3, Figure 3a, 3c**). Coffee and tea shipments primarily originated from three key locations in Canada that include Ontario (17TPJ, 57.8%), British Columbia (10UDV, 19.6%), and Quebec (18TXR, 15.0%) and entered the U.S. through two key ports that include 901-Buffalo-Niagara Falls, New York (40.5%) and 3004-Blaine, Washington (21.9%), and were delivered to New Jersey (18TWL, 8.7%), Illinois (16TDM, 5.3%), and Connecticut (18TXL, 4.7%) (**Table 2, Table 3, Figure 3b, 3d**). This is further emphasized through the treemap and heat map (**Figure 4**) which clearly shows the primary source of origin for all imports is located in grid 18TXR (Quebec) (**Figure 4B**) and enters through port 712-Champlain-Rouses Point, New York (**Figure 4A**). Similarly, the primary U.S. destination is grid 18TWL (New Jersey) (**Figure 4D**) receiving most of its shipments through port 712 - Champlain-Rouses Point, New York (**Figure 4C**).

Figure 3: Movement of trucks between origin and destination locations for (A) perishable vegetable food products and (C) non-perishable tea/coffee food products (N=531). Total number of product lines of (B) perishable vegetable food products (N=1,367) and (D) non-perishable tea/coffee food products (N=1,106) between origin and port of entry (red) and between port of entry and destination (green) for 2008-2011.

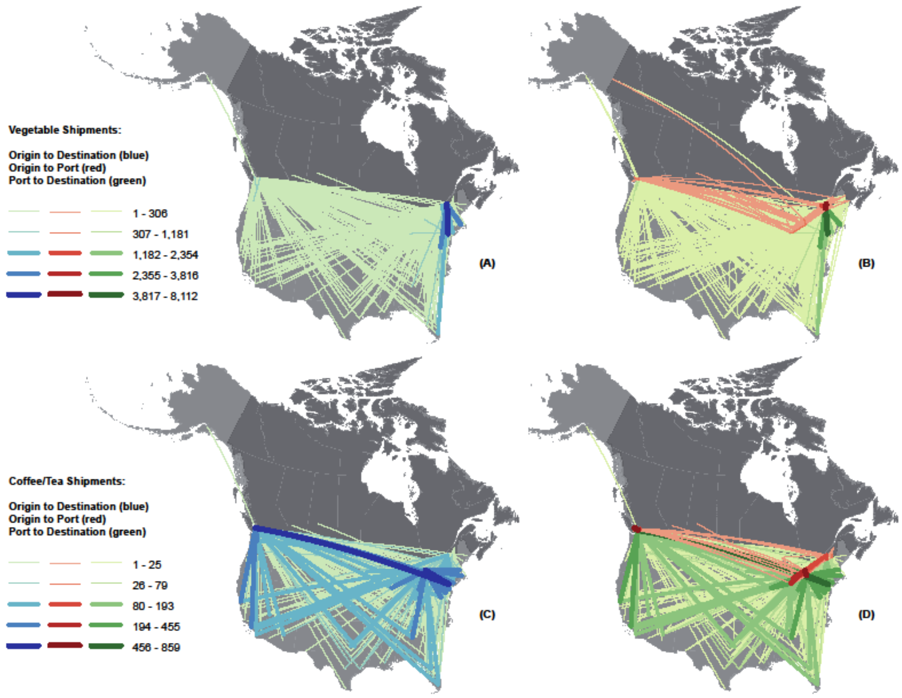
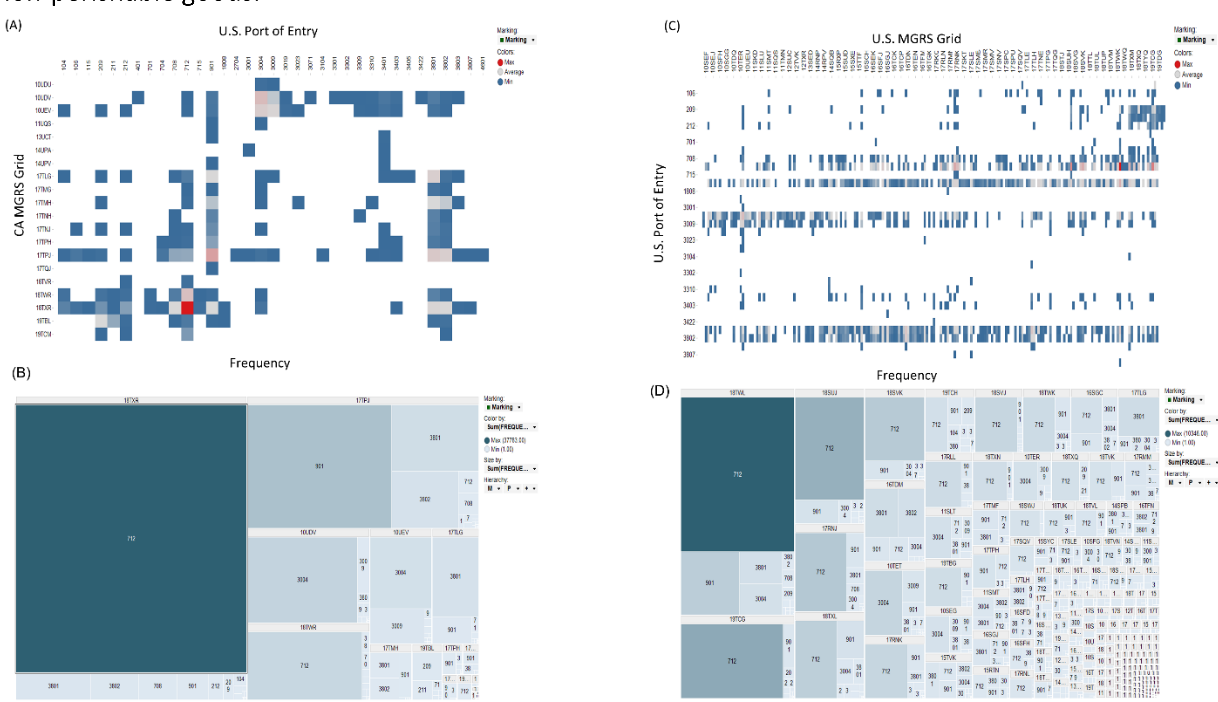


Figure 4: Visualizations showing port of entry associated with source locations in Canada using a (A) heat map and (B) treemap for perishable and non-perishable goods. And visualizations showing port of entry associated with destination locations in U.S. using a (C) heat map and (D) treemap for perishable and non-perishable goods.



3.2 Determining routes between origin-destination and port of entry

The routes created using each of the three LCPA friction surfaces (distance, time and timeslope) varied considerably as illustrated in **Figure 6** where we show the resulting outputs created for one point of origin in Quebec (MGRS grid 18TWR) to multiple destinations in the US (N=33). Distance routes performed the worst and timeslope the best when compared with GPS truck data (**Table 5**). Only 51% of the distance route overlapped with known GPS truck routes compared to 76% for time and 81% for timeslope.

Figure 6: Estimated routes for a single origin (MGRS grid 18TWR, Ontario, Canada, black square) to 33 destinations in the U.S. for all both vegetable and coffee/tea food types (black circles) using the least cost path analysis based on distance, time and time/slope travel.

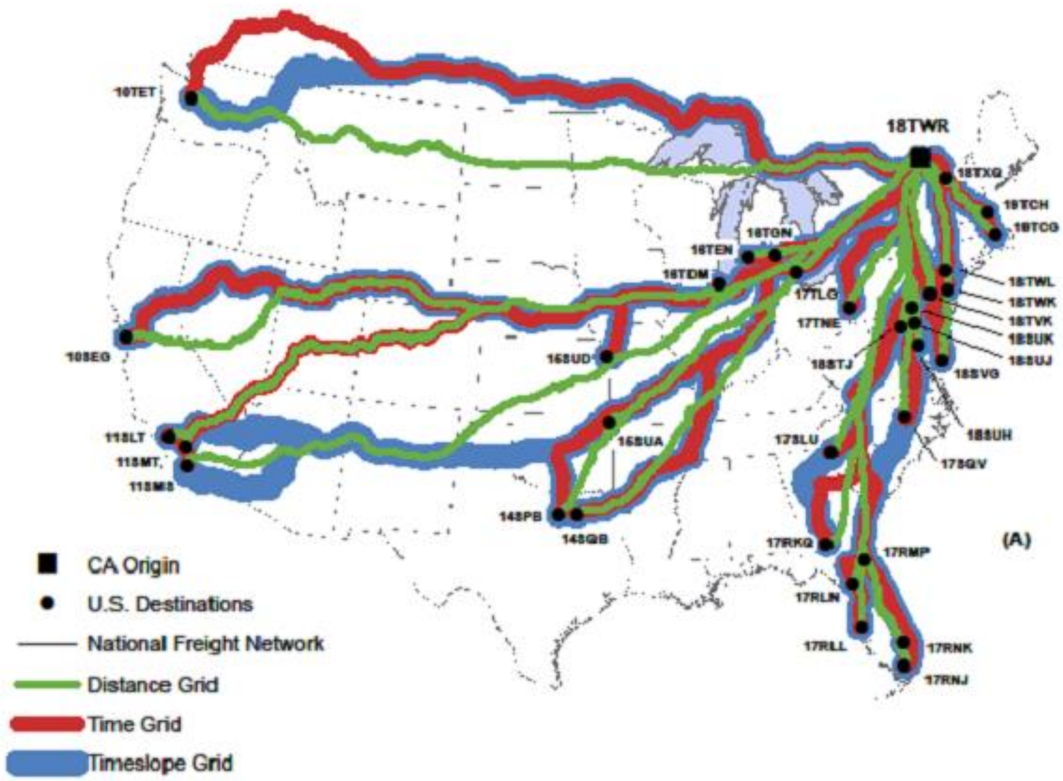


Table 5: Summary of LCPA comparisons to GPS tracked truck routes.

LCPA Friction Surface	Total Distance (km)	Total route length overlapping with GPS routes (km)	% Overlap
distance	26,783.755	13,784.313	51
time	22,400.613	17,096.532	76
timeslope	23,100.584	18,855.455	81

3.3 Evaluation

When the actual port of entry used during shipment was compared to those estimated for the same route we found that between 28-30% of matches were successful with between 26-29% matched successfully for vegetable products and 31% for coffee/tea products (**Table 6**). Matches were lowest based on distance and similar for both time and timeslope grids.

Table 6: Accuracy of predicted border crossing location by distance, time and timeslope.

FDA Food Industry Name	Source-POE-Destination Combinations	POE Matches: Distance Grid		POE Matches: Time Grid		POE Matches: Timeslope Grid	
		N	Percentage	N	Percentage	N	Percentage
Vegetables / Veg Products	1,369	361	26.4%	396	28.9%	388	28.3%
Coffee / Tea	1,115	345	30.9%	352	31.6%	354	31.7%
Combined	**2,484**	**706**	**28.4%**	**748**	**30.1%**	**742**	**29.9%**

4.0 Discussion

In this study origin-destination information was used to obtain an overview of geographic diversity of products, visualize shipping patterns and predict trucking routes using a least cost path analysis. We found that when speed and slope were used we were able to capture trucking routes, suggesting that goods were transported along routes with higher speeds (e.g. highways vs. other road surfaces) and through areas with least slope to ensure goods would reach destinations quickly and efficiently. Thus, as pertains to the international trade of goods, movement of goods follow a series of physical flows, many of which may not necessarily follow the most direct route (e.g. shortest distance), but instead follow a least cost path [38] between manufacturer and consignee. For example, the truck will take a path where the cost per mile to transport goods, such as fuel and oil costs, which during 2011 was the single greatest operating expense per mile and per hour [39] will affect route choice. Time can also affect route choice, since extended travel time will affect the freshness, quality, and safety of perishable foods [34] as well as increase costs.

Although potential routes taken by trucks were found to be a good fit when compared to GPS-tracked truck movement, we were not able to predict the actual route that may have been taken during a single journey. This became apparent when predicting border crossings since we were only able to successfully predict 31.7% of the crossings (**Table 3**). Many factors may affect truck routing over and above travel time and cost. These may include port congestion, port hours and number of lanes, traffic conditions, weather, avoiding known or suspected danger areas (e.g., cargo theft [40], extortion [41], and terrorism [42]), typology of transportation networks, supply chain logistics and international trade agreements and regulations [38].

For the case here, although Canada and the U.S. have a free trade agreement [43] that was designed to eliminate trade barriers, the routing of trucks may be limited due to restrictions placed on commercial vehicle length, width, and weight [44], port of entry restrictions and cabotage (the provision of transportation services by a foreign firm between point to point moves within the same country) [38]; Department of Homeland Security, 2012).

A second factor that may affect truck routing are related to supply chain logistics that include the physical flows associated with the transport chain as goods are moved between locations. These can be affected by several factors that include demand, supply, cost, load unit, regulation and ownership and distribution channels and centers [38]. One or several of these factors will no doubt play an important role in route selection. Selecting the least cost route based on transport costs is only one factor. Therefore, the LCPA, although good at capturing potential routes, as used here is too simplistic since it does not sufficiently capture the complexities associated with commercial truck movement and hence is unlikely to capture actual routes since it does not take into consideration any additional supply chain logistics (e.g. multiple stops between origin-destination [45]), multiple commodity shipments [46] intermodal freight movements [47], and re-routing due to weather [48].

Overall, three ports were mainly used to cross the border between Canada and the U.S. These included two ports in New York (712-Champlain/Rouses Point (N=43,754) and 901-Buffalo/Niagara Falls (N=13,614)) and one in Washington (3004-Blaine (N=9,173)). The greatest number of perishable food shipments mainly entered through ports in New York (712-Champlain/Rouses Point (N=42,046) and 901-Buffalo/Niagara Falls (N=4,152)) and Michigan (3801-Detroit (N=5,171)) while non-perishable foods entered through New York (901-Buffalo/Niagara Falls (N=9,462)), Michigan (3801-Detroit (N=3,215)) and Washington (3004-Blaine, (N=5,111)). When ports were analyzed individually (e.g. for a single origin-destination) it became apparent that for a single origin-destination pair multiple ports of entry may be used. For example when goods (both perishable and non-perishable) were transported from Ontario (17TPJ) to Massachusetts (19TCG) ten different ports were used. Understanding why different ports may be used for transporting the same good from the same location will be important for detecting changes in movement patterns and potential food risks by identifying outlier shipments that result in the need for port-shopping (the process whereby goods denied entry by CBP and/or FDA at one U.S. port are rerouted to other U.S. ports attempting to bypass inspection/sampling and enter U.S. commerce).

Future analysis and assessments can be improved by improving existing data issues. Some of these include:

(i) **Duplicate port codes:** The same port code was used for multiple ports (N= 28). For example, five ports were assigned port code 715-Trout River/Fort Covington (Fort Covington, Trout River, Burke, Chateaugay, and Churubusco) spanning a distance of approximately 46 km along the New York/Quebec border. Therefore ensuring unique identifiers are used for each port will improve accuracy and enhance future assessments.

(ii) **Incomplete and inaccurate data:** Several records contained incomplete place name information. The data was either missing or too ambiguous to decipher. Even though these records represent a small fraction of the total number used in this study, they still contain valuable information. Future efforts should ensure that place names are decipherable and include city-province and city-state information to enable locations to be properly geocoded since the same city name can be found in multiple states.

(iii) **Geolocation of placenames:** Currently the data cannot be viewed spatially since it does not include latitude and longitude and needs to be geo-coded. The most time consuming component of this study was cleaning and error checking the place name data and geo-coding the data. Standardizing place name requirements will help prevent duplication of places and multiple-variations in spelling of the same location. For example, many French-Canadian place names had 10 or more variations of spelling, capitalization, abbreviation, punctuation and hyphenation (e.g. Sainte Clotilde, Saint Clotilde, Ste-Clotilde, Ste Clotilde, St-Clothide, Ste Clotide, Saint Clotide, Sainte-clotilde-de-chateauguay, Sainte-Clotilde-De-Chateaugu, Clotilde, Sainte-Clotilde-D). Not only will this result in more accurate and complete data records but also enable continued visualization of these data as well as improve efficiency in decisions made by regulatory agencies.

The time allocated to conduct risk assessments on imported foods is extremely limited. While it is possible to determine the total number of shipments a firm has made over a specified period of time, which ports were used and how often, and how many different consignees that firm shipped to, having the ability to visualize the transport flows and how each of these components are connected through space and over time is useful for understanding existing transport patterns. In this study a variety of spatial and visualization methods were used to understand the transport patterns of perishable and non-perishable goods. Through the use of these visuals not only was it easy to see the existing trade patterns, identify key destinations and port of entry locations but also to compare how trade patterns differed between the two products. As the data is improved, having the ability to sort and compare current shipment patterns to historic shipment patterns by shipper, manufacturer and product line would allow for the development of "geographic transport footprints" that captures different flow patterns by product, shipper or manufacturer and be used to create individual "geographic transport profiles". Each profile would capture movement patterns over time and used for identifying deviations from historical routes and lead to further investigations.

References

1. USDA (2013) Summary data on food import values for 14 food categories, annual data since 1999. USA: United States Department of Agriculture.
2. USDA (2013) Locate a Port of Entry - Air, Land, or Sea. . USA: United States Department of Agriculture.
3. Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson M-A, et al. (2011) Foodborne illness acquired in the United States--major pathogens. Emerging infectious diseases 17: 7-15.
4. Spink J, Moyer DC (2011) Defining the public health threat of food fraud. J Food Sci 76: R157-163.
5. Shears P (2010) Food fraud – a current issue but an old problem. British Food Journal 112: 198-213.
6. FDA (2012) Global Engagement. U.S. Food and Drug Administration.
7. Nganje W, Richards T, Bravo J, Hu N, Kagan A, et al. (2009) Food safety and defense risks in US-Mexico produce trade. Choices 24: 1-8.
8. FDA (2012) PREDICT Fact Sheet. U.S. Food and Drug Administration.
9. Nyambok E (2010) Import Security: US And Global Approaches. Food and Agriculture Security: An Historical, Multidisciplinary Approach: An Historical, Multidisciplinary Approach: 83.
10. GPO (2003) Imports and Exports. Title 21 U.S. Code, Sec. 381.
11. Mirzabeiki V, Ringsberg H, Lumsden K. Effects of Using Smart Goods on Traceability Information and Carried out Activities in Supply Chains of Fresh Food–A Cross Case Analysi; 2010.
12. Opara LU (2003) Traceability in agriculture and food supply chain: a review of basic concepts, technological implications, and future prospects. Journal of Food Agriculture and Environment 1: 101-106.
13. Benza M, Bersani C, D'Inca M, Roncoli C, Sacile R, et al. Intelligent Transport Systems (ITS) applications on dangerous good transport on road in Italy; 2012 2012. IEEE. pp. 223-228.
14. Nganje WE, Skilton P (2011) Food Risks and Type I & II Errors. Supporters and Partners 14: 109.
15. Brewster RM, LeVert R (2005) Identifying Vulnerabilities and Security Management Practices in Agricultural & Food Commodity Transportation: American Transportation Research Institute.
16. DHS (2010) Surface Transportation Security Priority Assessment. Washington, D. C.: Department Of Homeland Security.
17. Tsamboulas D, Moraiti P (2008) Identification of potential target locations and attractiveness assessment due to terrorism in the freight transport. Journal of Transportation Security 1: 189-207.
18. HSIP (2013) Homeland Security Infrastructure Program Gold. [CD ROM]. 2013. Department of Homeland Security. Available: https://www.hifldwg.org/hsip-guest.

19. Andrienko N, Andrienko G (2012) Visual analytics of movement: An overview of methods, tools and procedures. Information Visualization 12: 3-24.
20. Andrienko G, Andrienko N (2008) A Visual Analytics Approach to Exploration of Large Amounts of Movement Data. In: Sebillo M, Vitiello G, Schaefer G, editors. Visual Information Systems Web-Based Visual Information Search and Management: Springer Berlin Heidelberg. pp. 1-4.
21. MacEachren A, Xiping D, Hardisty F, Diansheng G, Lengerich G. Exploring high-D spaces with multiform matrices and small multiples; 2003 2003. IEEE. pp. 31-38.
22. Elwood S (2011) Geographic Information Science: Visualization, visual methods, and the geoweb. Progress in Human Geography 35: 401-408.
23. Fabrikant SI, Lobben A (2009) Introduction: Cognitive Issues in Geographic Information Visualization. Cartographica 44: 139-143.
24. Schneider RG, Schneider KR, Webb CD, Hubbard M, Archer DL (2011) Agroterrorism in the US: An Overview.
25. Mansmann F, Fischer F, Keim DA, North SC. Visual Support for Analyzing Network Traffic and Intrusion detection of events using TreeMap and Graph Representations.; 2009; Baltimore, Maryland, USA.
26. Slingsby A, Dykes J, Wood J (2008) Using treemaps for variable selection in spatio-temporal visualisation. . Information Visualization 7: 210-224.
27. Atkinson DM, Deadman P, Dudycha D, Traynor S (2005) Multi-criteria evaluation and least cost path analysis for an arctic all-weather road. Applied Geography 25: 287-307.
28. Chandio IA, Matori ANB, WanYusof KB, Talpur MAH, Khahro SH, et al. (2012) Computer Application in Routing of Road using Least-Cost Path Analysis in Hillside Development. Research Journal of Environmental and Earth Sciences 4: 907-911.
29. Gonzales EK, Gergel SE (2007) Testing assumptions of cost surface analysis—a tool for invasive species management. Landscape Ecology 22: 1155-1168.
30. Bagli S, Geneletti D, Orsi F (2011) Routeing of power lines through least-cost path analysis and multicriteria evaluation to minimise environmental impacts. Environmental Impact Assessment Review 31: 234-239.
31. Rees WG (2004) Least-cost paths in mountainous terrain. Computers & Geosciences 30: 203-209.
32. Devlin GJ, McDonnell K, Ward S (2008) Timber haulage routing in Ireland: an analysis using GIS and GPS. Journal of Transport Geography 16: 63-72.
33. Flisberg P, Forsberg M, Rönnqvist M (2007) Optimization based planning tools for routing of forwarders at harvest areas. Canadian Journal of Forest Research 37: 2153-2163.
34. Osvald A, Stirn LZ (2008) A vehicle routing algorithm for the distribution of fresh vegetables and similar perishable food. Journal of Food Engineering 85: 285-295.
35. ESRI (2013) ArcGIS Desktop (Version 10.2) [Computer software]. Redlands, CA: Environmental Systems Research Institute.
36. Blanford JI, Kumar S, Luo W, MacEachren AM (2012) It's a long, long walk: accessibility to hospitals, maternity and integrated health centers in Niger. Int J Health Geogr 11: 24.
37. Kwiatkowski K.
38. Rodrigue J-Pea (2013) The Geography of Transport Systems, . Hofstra University: Hofstra University, Department of Global Studies & Geography.
39. Fender KJ, Pierce DA. An Analysis of the Operational Costs of Trucking: 2012 Update; 2012.
40. Burges C (2012) Cargo theft, loss prevention, and supply chain security. : Butterworth-Heinemann.
41. Cattan N (2011) Mexico drug war's latest victim: the lime. Christian Science Monitor.
42. Tarr RW, McGurk V, Jones C (2005) Intermodal Transportation Safety and Security Issues: Training against Terrorism. Journal of Public Transportation 8: 87.

43. NAFTA (1993) North American Free Trade Agreement Implementation Act, Public Law No. 103-182, 107 Stat. 2057.
44. USDepTransport (2004) Federal Size Regulations for Commercial Motor Vehicles. U.S. Department of Transportation.
45. Patil GR, Holguín-Veras J (2005) Observed trip chain behavior of commercial vehicles. Transp Res Rec 1906: 74-80.
46. Patil GR, Holguín-Veras J (2008) A multicommodity integrated freight origin-destination synthesis model. Networks and Spatial Economics 8: 309-326.
47. Southworth F, Peterson BE (2000) Intermodal and international freight network modeling. Transportation Research Part C 8: 147-166.
48. Pisano P, Goodwin LC (2002) Surface transportation weather applications. . Federal Highway Administration in concert with Mitretek Systems, Institute of Transportation Engineers Annual Meeting.