Title: Estimating County-Level Farm-Based Agricultural Commodity Movements

Fei Xie
R&D Staff
865-946-1306
Center for Transportation Analysis
Oak Ridge National Laboratory
Knoxville, TN 37931
Email: xief@ornl.gov

Ho-Ling Hwang (corresponding author)
R&D Staff
865-946-1224
Center for Transportation Analysis
Oak Ridge National Laboratory
Knoxville, TN 37931
Email: hwanghl@ornl.gov

Shih-Miao Chin
R&D Staff
865-946-1254
Center for Transportation Analysis
Oak Ridge National Laboratory
Knoxville, TN 37931
Email: chins@ornl.gov

Submitted to Agriculture and Food Transportation Committee (AT030) for presentation and publication at the 98th Annual Meeting of the Transportation Research Board

Word counts: 3,208 (text) + 750 (3 tables) = 3,958 words

Submitted August 01, 2018

This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

ABSTRACT

The Freight Analysis Framework (FAF), a nation-wide commodity-based freight flow database, is the only publicly available data source that provides a comprehensive resource of national freight movement data. The FAF estimates origin-destination (O-D) commodity flows between states, sub-state regions, and major international gateways. Recently, there are emerging needs of the FAF data at a more detailed geographical level than is presently available. Taking the agricultural commodities as an example, this study presents a disaggregation approach that aims at disaggregating the existing zone-level commodity flows to county-level commodity flows. With limited agricultural commodity data and statistics, the disaggregation process includes three major steps: (1) estimation of county-level agricultural commodity production at origins, (2) estimation of county-level agricultural commodity attraction to destinations, and (3) estimation of county-level origin-destination agricultural commodity flows. The disaggregation process is applied to major farm-based agricultural products considered in the FAF. This paper discusses the process and illustrates the resulting geographic distributions of county-level agricultural commodity production and attraction. A few examples of county-level flows between origins and destinations are also presented.

INTRODUCTION

The Federal Highway Administration (FHWA), with partnership from Bureau of Transportation Statistics (BTS), has developed and maintained the Freight Analysis Framework (FAF) *(1)*, a nation-wide commodity-based freight flow database and tool. Designed as a policy tool for the U.S. Department of Transportation (USDOT), FAF is the only publicly available data source that provides a comprehensive resource of national freight movement data across all modes of transportation. FAF estimates origin-destination (O-D) commodity flows between states, sub-state regions, and major international gateways. These regions are called FAF zones which are equivalent to Commodity Flow Survey (CFS) areas defined in the CFS database *(2)*. The latest available FAF data is the FAF version 4 (FAF4), which is 2012 CFS based. Over the years, FHWA often receives requests from State Departments of Transportation (DOTs) and metropolitan planning organizations (MPOs) wanting to use FAF data at a more detailed geographic level than is presently available. Specifically, there is an emerging need of developing nationwide county-level freight flow database.

To explore the potential of disaggregating FAF4 data further into more detailed geography, thus to meet regional planning needs, this study takes the farm-based agricultural commodities as an example to demonstrate the disaggregation process. Specifically, this study focuses on the farm-base agricultural shipment that is out-of-scope of the CFS *(3)*. This research aims to identify relevant data sources and methodologies to disaggregate the existing FAF4 zone-level farm-based agricultural movements into county-level flows. Estimation processes include the steps to identify: (1) county origins (O) where agricultural commodities are produced, (2) county destinations (D) where agricultural commodities are sent, and (3) farm-based agricultural commodity flows between origins and destinations (O-D).

The methodologies and results presented in this study are expected to bring values to future freight database development and application in several ways. First, the disaggregation methodologies developed in this study could bring insights on potential disaggregation processes applicable to other commodities under the FAF's scope. Second, the county-level disaggregation results for the farm-based agricultural commodities could be used by various government agencies at the national, state, and regional levels, as well as private transportation agencies to have a better understanding of agricultural freight flow. Finally, this county-level flow data will enable FHWA to support more enhanced national highway network modeling methods.

In the rest of this paper, the farm-based agricultural commodity shipment disaggregation process and the data sources are discussed first under the Methods and Data section. Major disaggregation results generated from this study are then presented in the Results section, and followed by the Conclusion section.

METHODS AND DATA

Under the FAF4, there are 132 domestic FAF zones, which are considered as both origins and destinations geographies for the shipments of farm-based agricultural commodities *(1, 3)*. As shown in Figure 1, FAF4 estimates the zone-level production at origins, zone-level distribution at destinations, and zone-level O-D assignment of these agricultural commodity flows. The shipment quantities are measured in two metrics, weight (tons) and value ($). Multiple agricultural commodity categories are captured in FAF, and they are represented with the Standard Classification of Transported Goods (SCTG) commodity codes at the 2-digit level.

At origins, FAF models the production of five major farm-based agricultural commodities, including SCTG 01 (live animals and fish), 02 (grains), 03 (other agricultural products), 04 (animal feed, eggs, honey and other animal products), and 07 (other prepared foodstuffs (milk)). These agricultural commodities are then delivered to destinations where they are stored, processed, and distributed. In addition to the five commodity categories stated above, FAF models two additional agricultural commodities on the destination side, SCTG 05 (meat, poultry, etc.) and 09 (tobacco), which are produced based on SCTG 01 and SCTG 03 commodities, respectively. For detailed assumptions and theories on estimating farm-based agricultural products production, attraction, and O-D assignment at the FAF zone levels, readers are referred to the report *(3)*.

To disaggregate the agricultural commodity shipment from zone levels to county levels, this study considers three major steps as shown in Figure 1:

- Step 1: Estimation of county-level agricultural commodity production at origins
- Step 2: Estimation of county-level agricultural commodity attraction to destinations
- Step 3: Estimation of county-level O-D agricultural commodity flows

More details on these steps are described in the following.

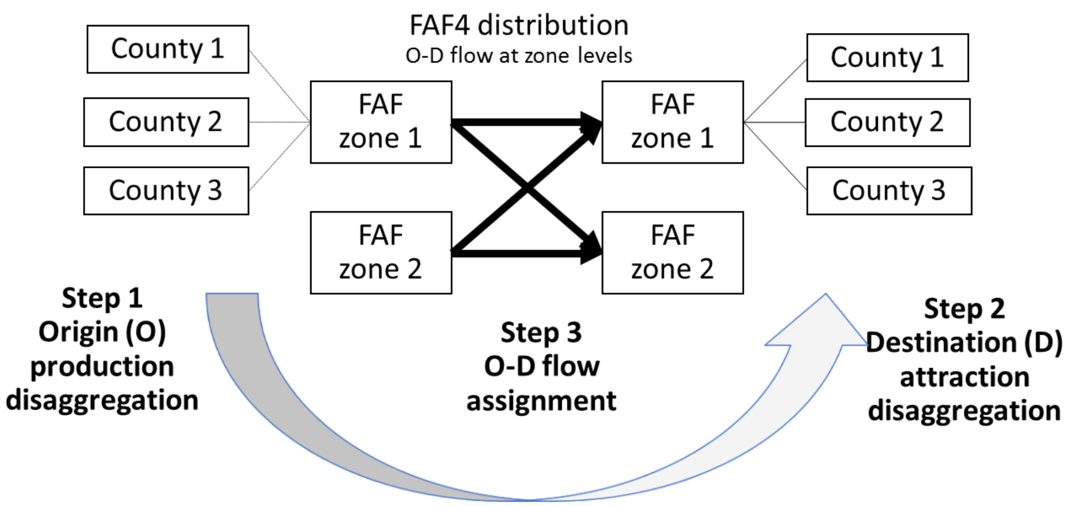


FIGURE 1 Three major steps in disaggregating farm-based agricultural shipments.

Estimation of County-Level Agricultural Commodity Production at Origins

Farm assets, such as available farm land in acres, are the major factor affecting the production capacity of the corresponding commodity in one area. Given scarce information obtainable for the county-level farm production, we assume each county's farm production volume is in proportion to the quantities of related farm assets available in the given county. This is the basis used for the disaggregation of zone-level agricultural production into county-level production.

For each agricultural commodity, there are many related farm assets. For simplicity, we only consider major farm assets with available county-level quantitative data in existing data sources. Specifically, the United States Department of Agriculture (USDA) desktop data query tool *(4)* provides one or several key county-level farm asset statistics for each of the agricultural commodities considered under this study. Below is a list of farm assets by SCTG category used in this study:

- SCTG 01 – cattle, hogs, and poultry;
- SCTG 02 – grain;

- SCTG 03 – vegetables and soybean;
- SCTG 04 – hay;
- SCTG 07 – cattle.

For SCTG 02, 04, and 07, as each of these commodities is related to only one farm asset in this study, disaggregation can be simply made by proportionally allocating the total zone-level production into each county based on the corresponding asset quantity. To express this process mathematically, let i indexes counties belonging to zone j with $i \in N_j$, where N_j is the set of counties within the zone j. Given the farm asset quantity A_i (e.g., acres of available grain harvest farm-lands for SCTG 02) for each county i and the total farm production Y_j^O in the zone j, we could determine the corresponding farm production X_i^O in each county i with equation (1). The farm asset quantities in each county are extracted from the USDA desktop data query tool *(4)* and the zone-level total farm production is obtained from the FAF database *(1, 3)*.

$$X_i^O = \frac{A_i}{\sum_{i \in N_j} A_i} \times Y_j^O \quad (1)$$

However, for SCTG 01 and 03, each is related to multiple farm assets. Simple summation of asset quantity is not applicable because different farm assets could have different yields on agricultural commodities. For example, according to the Agriculture Statistics *(5)*, one cattle on average yields 0.510 tons of meat while one poultry on average yields only about 0.002 tons of meat (see Table 1). Therefore, a weighted sum method is used when estimating commodity yield by different farm assets. The weighting factors by asset type, adopted from the Agriculture Statistics *(5)*, are summarized in Table 1.

TABLE 1 Weighting Factor for Farm Assets[a]

SCTG Category	Farm Asset	Unit[b]	Weighting factor
SCTG 01	Cattle	tons/livestock	0.510
	Hogs	tons/livestock	0.128
	Poultry	tons/livestock	0.002
SCTG 03	Vegetables	tons/acre	14.260
	Soybean	tons/bushels	0.030

a. Source: Agriculture Statistics 2013 *(5)*
b. unit is based on quantity unit for each farm asset

For these two commodity types (SCTG 01 and 03), let $k \in K$ indicates the farm asset for each agricultural commodity. In each county i, given the weighting factor W_k^O and asset quantity A_{ik} for farm asset k, we could alternatively determine farm production X_i^O using equation (2).

$$X_i^O = \frac{\sum_{k \in K} W_k^O \times A_{ik}}{\sum_{i \in N_j} \sum_{k \in K} W_k^O \times A_{ik}} \times Y_j^O \quad (2)$$

Estimation of County-Level Agricultural Commodity Attraction to Destinations

The second step of the disaggregation process is to determine the volume of shipments on farm-based agricultural commodities are attracted to each destination county for further storage, process, and distribution. We assume that higher agricultural product-related economic development in a county will have more agricultural products attracted to the area. Therefore, the disaggregation at destinations is based on information obtained from the County Business Patterns (CBP), published annually by U. S. Census Bureau *(6)*, which contains a set of county-level economic data including number of establishments, payrolls, employment, etc.

The business patterns are classified by different industry sectors, and we need to determine which sectors are related to each agricultural commodity. Note that business sectors are indexed using the North American Industry Classification System (NAICS) code *(7)*. We identify the relationship between agricultural commodities and business sectors based on the shipment characteristics by NAICS by mode by commodity in the United States *(8)*. For simplicity, we only consider the most relevant business sectors at county destinations for each agricultural commodity, and they are summarized in Table 2. As shown in the table, one industry sector may serve multiple agricultural commodities, and the selected business sectors for each SCTG agricultural commodity represents at least 90% of the total shipment by weight.

TABLE 2 Selected Major Relevant Business Pattern for Each Farm Commodity

SCTG	Name	NAICS	Description	% of all shipment
1	Animals and fish (live)	4245	Farm product raw material merchant wholesalers	90.87%
2	Cereal grains	4245	Farm product raw material merchant wholesalers	98.20%
		4249	Miscellaneous nondurable goods merchant wholesalers	
3	Agricultural products	311x	Food manufacturing	97.40%
		4244	Grocery and related product merchant wholesalers	
		4245	Farm product raw material merchant wholesalers	
		4249	Miscellaneous nondurable goods merchant wholesalers	
4	Animal feed, etc.	311x	Food manufacturing	97.10%
		325x	Chemical manufacturing	
		4244	Grocery and related product merchant wholesalers	
		4245	Farm product raw material merchant wholesalers	
		4249	Miscellaneous nondurable goods merchant wholesalers	
5	Meat, etc.	311x	Food manufacturing	98.60%
		4244	Grocery and related product merchant wholesalers	
7	Other prepared foodstuffs, etc.	311x	Food manufacturing	97.60%
		312x	Beverage and tobacco product manufacturing	
		325x	Chemical manufacturing	
		4244	Grocery and related product merchant wholesalers	
9	Tobacco products	312x	Beverage and tobacco product manufacturing	97.30%
		4249	Miscellaneous nondurable goods merchant wholesalers	

Among multiple economic data presented in the CBP, annual payrolls are applied to estimate county-level attraction of farm-based agricultural products.

Though multiple business sectors could be related to one agricultural commodity, they may be weighted differently. For each FAF zone, we determine relative importance of each business sector by comparing the total payroll by sector as follows. For one agriculture commodity (e.g., SCTG 03), let $m \in M$ indexes related business sectors (e.g., NAICS 311x, 4244, 4245, and 4249 for SCTG 03) shown in Table 2. At one zone j, we could determine the zone-level total payroll W_{jm}^{D} for each sector m using the CBP database *(6)*. W_{jm}^{D} is then the weighting factor for each sector m in zone j. Given the county-level payroll P_{im} for county i for sector m and the total commodity quantity Y_j^D shipped to zone j, we could determine the county-level commodity attraction X_i^D with equation (3).

$$X_i^D = \frac{\sum_{m \in M} W_{jm}^{D} \times P_{im}}{\sum_{i \in N_j} \sum_{m \in M} W_{jm}^{D} \times P_{im}} \times Y_j^D \tag{3}$$

Estimation of County-Level O-D Agricultural Commodity Flows

The two steps demonstrated above could estimate the county-level commodity production X_i^O at origins and commodity attraction X_i^D at destinations. The third step is to estimate the county-level agricultural flow $f_{ii'}$ between counties i and i'.

There are two major challenges in estimating the county-level agriculture flow. First, the county-to-county matrix could be extremely large (up to 3,220 × 3,220 O-D pairs), and the direct estimation of the county-level flow using conventional methods (e.g., gravity model) is computationally challenging. Second, the flow assignment is a complicated process, and the direct estimation of the county-level flow using limited agriculture related data may not yield reliable results. Note that the FAF already presents relatively reliable estimation of zone level-commodity flow $F_{jj'}$ between zones j and j'. Therefore, we will retain the existing zone-level commodity flow assumption with the constraint in (4).

$$F_{jj'} = \sum_{i \in N_j} \sum_{i' \in N_{j'}} f_{ii'} \qquad j, j' \in J \tag{4}$$

With constraint (4), the large-scale flow assignment problem for all U.S. counties could be decomposed to a set of subproblems that can be solved separately. For each subproblem, we solve the county-level flow assignment problem between one pair of FAF zone-level origin j and destination j'.

The gravity model *(9)*, a widely used method for transportation distribution estimation, is used in this study for estimating the flow assignments. Based on Newton's law of gravitation, the inherent assumption of the gravity model is that the estimated flow always increases as the impedance factor (measured by distance or travel time) decreases. Note that there are also other more advanced freight distribution models available in the literature, such as the gravity model using spatial correlation of time series data used for the disaggregation of energy products *(10)*. However, the method *(10)* requires time series freight data which are not readily available in the agricultural case. Therefore, we only consider the gravity model in this study, and an extension with advanced flow assignment methods is considered as a future work when related data are available.

The gravity model formulation for the county-level agricultural commodity flows between zones j and j' can be described in (5) as follows.

$$f_{ii'} = K_i K_{i'} X_i^O X_{i'}^D \gamma_{ii'} \qquad i \in N_j, i' \in N_{j'} \quad (5)$$

Where,

$$K_i = \frac{1}{\sum_{i' \in N_{j'}} K_{i'} X_{i'}^D \gamma_{ii'}} \qquad i \in N_j \quad (5.a)$$

$$K_{i'} = \frac{1}{\sum_{i \in N_j} K_i X_i^O \gamma_{ii'}} \qquad i' \in N_{j'} \quad (5.b)$$

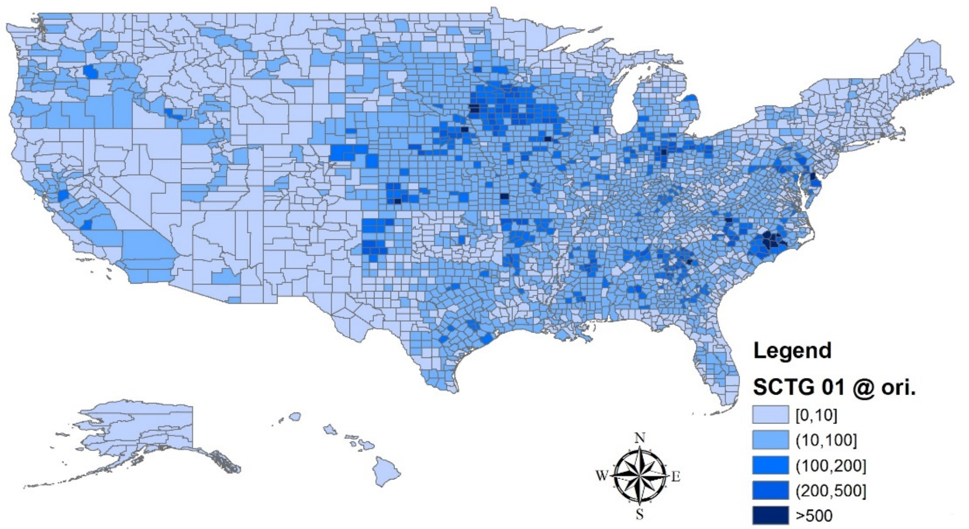
In the gravity model in (5), the county-level commodity production X_i^O and attraction $X_{i'}^D$ are determined in the first two steps as mentioned above. $\gamma_{ii'}$ is the impedance factor between origin i and destination i'. In this study, $\gamma_{ii'}$ is set to be inverse to the shipping distance $d_{ii'}$, and the county-to-county shipping distance data are adopted from the transportation network database *(11)*. Defined in (5.a) and (5.b), K_i and $K_{i'}$ are the balancing factors solved iteratively with the Iterative Proportional Fitting (IPF). For more detailed on the IPF process, readers are referred to the study *(9)*.

RESULTS

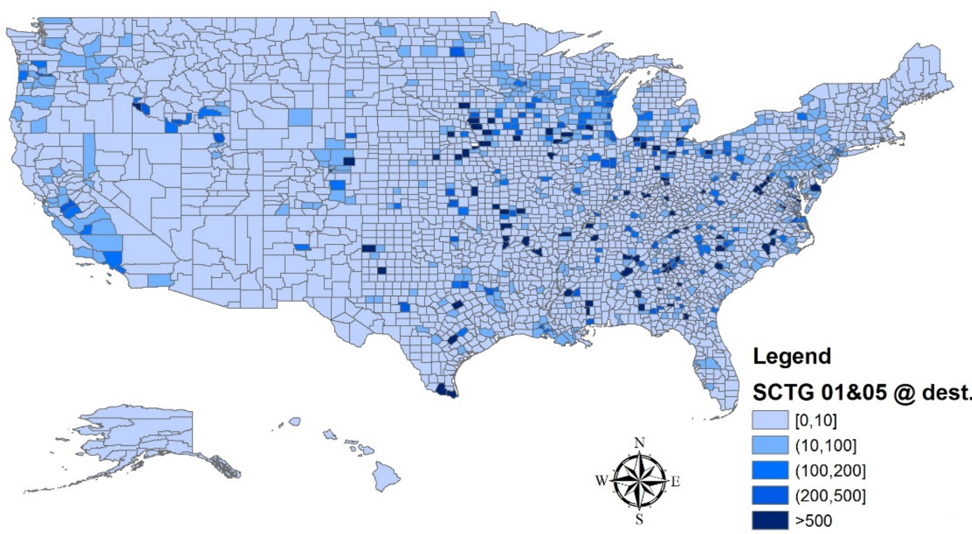
Based on results generated from this study, Figures 2 to 4 show the maps of county-level production and attraction of agricultural commodities SCTG 01, SCTG 02, and SCTG 03, respectively. These maps used five intervals for the display of annual production or attraction densities, measured by tons per square mile. As shown in these figures, the five density ranges are symbolized by varying shades of blue, where the darker the color the higher the production, or attraction, quantity per square mile is.

These figures illustrate that major agricultural production counties are geographically specific. According to Figure 2(a), the production of SCTG 01 (live animal and fish) is mainly clustered in the Midwest region, Texas, and other southern states. Most of counties have relatively small quantity of SCTG 02 production, and there are only 16 counties that have a production level of more than 500 tons/mi^2 (see “>500” in Figure 2(a)). Compared to the SCTG 01 commodity, the SCTG 02 commodity (grain) has more nation-wide total production quantity by weight (452 million tons of SCTG 02 compared to 90 million tons of SCTG 01). Therefore, as shown in Figure 3(a), there are 131 counties that have production of more than 500 tons/mi^2 and 486 counties that have production between 200 and 500 tons/mi^2. Most of these higher production-density counties are located in the Midwest region. Similar to SCTG 02, major production counties of the SCTG 03 commodity (other agricultural products) are also clustered in the Midwest region. In addition to the Midwest region, counties in the Pacific region and Florida also have large quantity of SCTG 03 production (see Figure 4(a)).

Major destinations of agricultural commodities are also geographically specific. According to these figures, geographic distributions of major destinations have similar patterns with those major production counties. For example, as shown in Figure 3(b), most of the SCTG 02 commodities are also attracted to counties in the Midwest regions. Note that raw agricultural commodities (i.e., farm-based) are generally shipped to storage, processing, and distribution centers close to the production locations, an economical solution for agricultural businesses. Shipments begin from these receiving locations (storage, processing & distribution centers) are in-scope for the CFS.

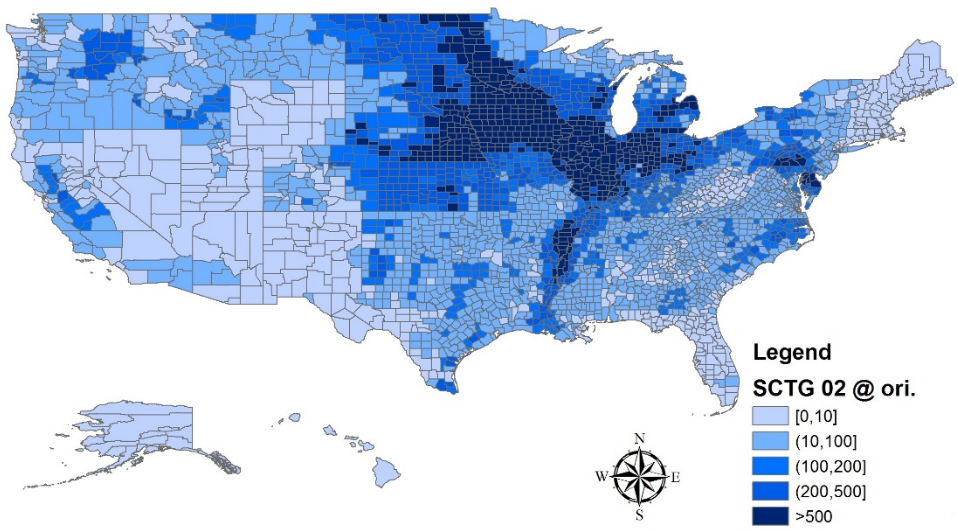


(a) Production of SCTG 01

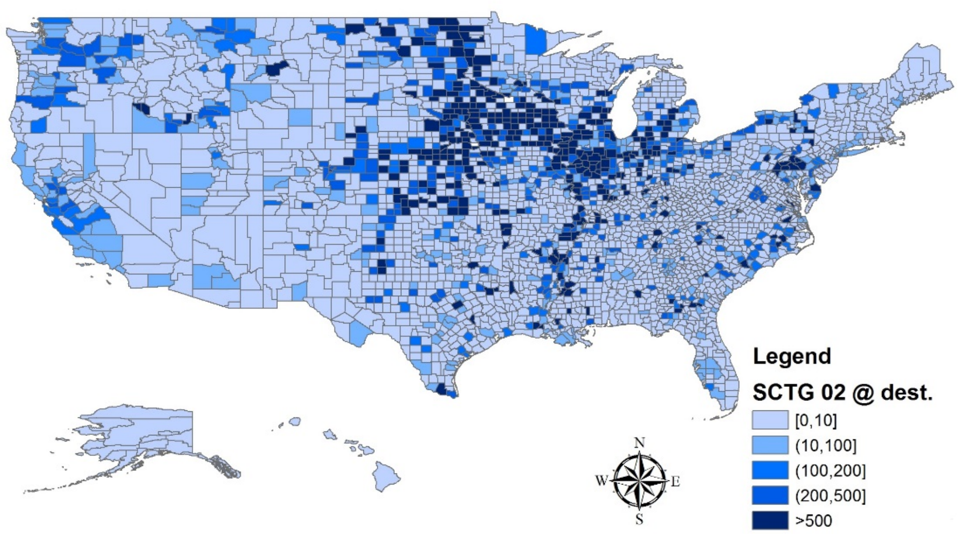


(b) Attraction of SCTG 01

FIGURE 2 Production and attraction (tons/square mile) of SCTG 01 (live animals and fish) at county level.

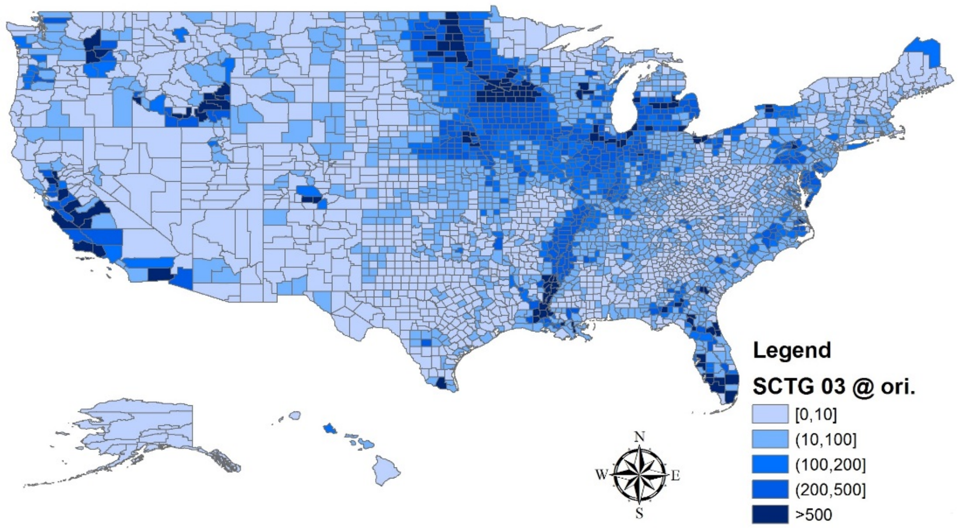


(a) Production of SCTG 02

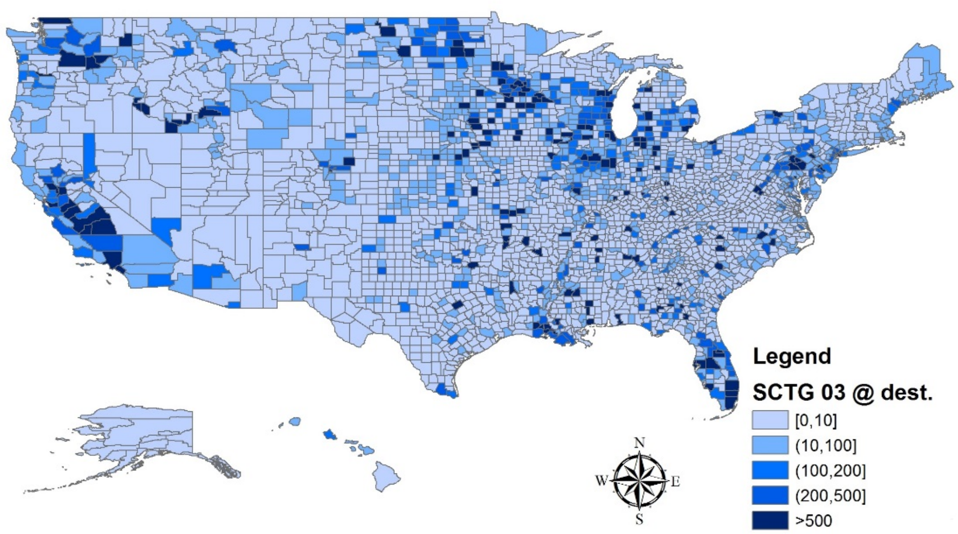


(b) Attraction of SCTG 02

FIGURE 3 Production and attraction (tons/square mile) of SCTG 02 (grains) at county level.



(a) Production of SCTG 03



(b) Attraction of SCTG 03

FIGURE 4 Production and attraction (tons/square mile) of SCTG 03 (other agricultural products) at county level.

TABLE 3 Example of Flow Assignment between Counties

Origin		Destination		SCTG01[a]		SCTG02		SCTG03		SCTG04		SCTG07		SCTG09	
FAF zone	County FIPS	FAF zone	County FIPS	live animals and fish		grains		other agricultural products		animal feed, eggs, honey and other animal products		other prepared foodstuffs		tobacco	
				K tons	$(M)	K tons	$(M)	K tons	$(M)	K tons	$(M)	K tons	$(M)	K tons	$(M)
101	10001	101	10001	0.0	0.0	9.2	1.9	21.9	11.0	0.0	0.0	0.0	0.0	0.0	0.0
101	10001	101	10003	0.0	0.0	23.9	4.9	74.4	37.5	0.0	0.0	0.0	0.0	0.0	0.0
101	10003	101	10001	0.0	0.0	0.3	0.1	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0
101	10003	101	10003	0.0	0.0	12.2	2.5	12.3	6.2	0.0	0.0	0.0	0.0	0.0	0.0
101	10001	109	10005	306.7	488.6	330.1	75.0	53.7	21.8	4.8	0.1	17.8	6.3	0.0	0.0
101	10003	109	10005	17.4	27.7	124.6	28.3	7.0	2.8	3.4	0.1	2.7	1.0	0.0	0.0
109	10005	101	10001	0.0	0.0	7.6	1.6	1.7	0.8	0.0	0.0	0.0	0.0	0.0	0.0
109	10005	101	10003	0.0	0.0	28.8	6.1	6.7	3.0	0.0	0.0	0.0	0.0	0.0	0.0
109	10005	109	10005	127.6	116.1	363.3	84.9	4.7	1.7	4.8	0.1	12.0	3.0	0.0	0.0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

a. At destination, SCTG 01 commodities are divided into two groups, SCTG 01 and SCTG 05.

With the production and attraction of farm-based agricultural commodities disaggregated at counties, shipment flow assignments are estimated using the IPF method. Table 3 shows an example of the shipment flows between counties, indexed by Federal Information Processing Standard (FIPS), and associated FAF zone codes, in terms of weight (thousand tons) and value ($ million) by commodity. For instance, there are 23.9 thousand tons or 4.9 million dollars of SCTG 02 shipped from the origin at county "10001" to the destination at county "10003". In Table 3, zero-value shipment quantity indicates that there is no shipment of the given agricultural commodity for the county pair. Note that all 10 pairs of counties in this example do not have any commodity flow for the SCTG 09 (tobacco).

CONCLUSION

In this study, we developed a process for disaggregating zone-level freight shipment to county-level shipment for farm-based agricultural commodities considered in the FAF4. The disaggregation process includes three major steps: (1) estimation of county-level agricultural commodity production, (2) estimation of county-level agricultural commodities shipped to destination, and (3) assignments of county-level O-D flows of agricultural commodities. Based on this research, we illustrated the resulting patterns of county-level farm-based agricultural commodity production and attraction are fairly reasonable.

The disaggregation method presented in this study is developed based on limited agricultural freight shipment data and statistics available in the literature. When additional data are available at a more detailed geographic resolution, the disaggregation method could be further improved. For example, we could adopt the gravity model using spatial correlation of time series data to better capture the inter-relationships between counties in shipment. Another immediate extension of this study is to apply the disaggregation method to other major commodities .

AUTHOR CONTRIBUTION STATEMENT

The authors confirm contribution to the paper as follows: study conception and design: F. Xie, H. Hwang, S. Chin; data collection: F. Xie; analysis and interpretation of results: F. Xie, H. Hwang, S. Chin; draft manuscript preparation: F. Xie. All authors reviewed the results and approved the final version of the manuscript.

ACKNOWLEDGMENT

The research presented in this paper was conducted as a part of the Freight Analysis Framework Project supported by the Federal Highway Administration under the guidance of Mr. Birat Pandey. The authors thank the support of the sponsors and remain solely responsible for the content and opinions expressed.

REFERENCE

1. Oak Ridge National Laboratory. *Freight Analysis Framework Version 4.0 - User's Guide for Release 4.0.* https://faf.ornl.gov/fafweb/data/FAF4%20User%20Guide.pdf. 2015.
2. Bureau of Transportation Statistics. *2012 Commodity Flow Survey.* 2015.
3. Oak Ridge National Laboratory. *The Freight Analysis Framework Version 4 (FAF4) - Building the FAF4 Regional Database: Data Sources and Estimation Methodologies*. 2016.

4. USDA. *USDA Desktop Data Query Tool.* *http://www.agcensus.usda.gov/Publications/2012/Online_Resources/Desktop_Application/*. 2013.

5. NASS. *Agricultural Statistics 2013.* *https://www.nass.usda.gov/Publications/Ag_Statistics/2013/index.php*. 2013.

6. U.S. Census Bureau. *County Business Patterns 2012.* *https://www.census.gov/data/datasets/2012/econ/cbp/2012-cbp.html*. 2012.

7. Executive Office of the President Office of Management and Budget. *North American Industry Classification System.* *https://www.census.gov/eos/www/naics/*. 2017.

8. U.S. Census Bureau. *CF1200A29 - Geographic Area Series: Shipment Characteristics by NAICS by Mode by Commodity for the United States: 2012*. 2014.

9. Deming, W. E. and F. F. Stephan. On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known. *Ann. Math. Statist.*, Vol. 11, No. 4, 1940, pp. 427-444.

10. Lim, H., S. M. Chin, H.-L. Hwang and L. D. Han. A Gravity Model Using Spatial Correlation of Time Series Data for Freight Distribution Estimation: A Case Study of the County-Level Coal Distribution Estimation in United States. *97th Transportation Research Board Annual Meeting*, 2018,

11. Oak Ridge National Laboratory. *CTA Transportation Networks.* *https://cta.ornl.gov/transnet/index.html*. 2009.