ENVIRONMENTAL RESEARCH
LETTERS

LETTER • OPEN ACCESS

Food flows between counties in the United States from 2007 to 2017

To cite this article: Deniz Berfin Karakoc *et al* 2022 *Environ. Res. Lett.* **17** 034035

ENVIRONMENTAL RESEARCH
LETTERS

OPEN ACCESS

RECEIVED
11 June 2021

REVISED
28 January 2022

ACCEPTED FOR PUBLICATION
7 February 2022

PUBLISHED
28 February 2022

LETTER

Food flows between counties in the United States from 2007 to 2017

Deniz Berfin Karakoc, Junren Wang and Megan Konar*

Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, United States of America
* Author to whom any correspondence should be addressed.

E-mail: mkonar@illinois.edu

Keywords: food flows, high-resolution, data driven, time series, United States

Supplementary material for this article is available online

Abstract

Food supply chains are essential for distributing goods from production to consumption points. These complex supply chains are important for food security and availability. Recent research has developed novel methods to estimate food flows with high spatial resolution, but we do not currently understand how fine-grained food supply chains vary in time. In this study, we use an improved version of the Food Flow Model to estimate food flows (kg) between all county pairs across all food commodity groups for the years 2007, 2012, and 2017 (which requires estimating 206.3 million links). We then determine the core counties to the US food flow networks through time with a multi-criteria decision analysis technique. Our estimates of county-to-county food flows in time are freely available with this paper and could be useful for future research, policy, and decision-making.

1. Introduction

Food supply chains are complex systems that incorporate production, distribution, intermediate processing, and consumption of food commodities (Porkka *et al* 2013, MacDonald *et al* 2015). Recent studies have estimated food flows with high subnational spatial resolution (Lin *et al* 2019), but we do not currently understand how fine-grained food supply chains vary in time. Food supply chains propagate and attenuate shocks (Heslin *et al* 2020, Gomez *et al* 2021, Karakoc and Konar 2021), embody resources (Weber and Matthews 2008, Dang *et al* 2015, Robinson *et al* 2016, Metson *et al* 2020), and depend upon infrastructure (Attavanich *et al* 2013). Estimates of how high-resolution food flows vary in time would enable an assessment of spatiotemporal risks in food supply chains, critical infrastructure, and environmental footprints. The goal of this paper is to estimate food flows between counties in the United States for multiple years (e.g. 2007, 2012, and 2017) and identify the counties that are core to the network.

The United States is an important nation in the global food system. It is a major producer, consumer, and trade power in agri-food commodities (Xu *et al* 2011, Konar *et al* 2018). The US produces over 50% of the world's soybean and 30% of the world's corn (2020b). The US also accounts for significant fractions of the world's export market for sorghum and wheat (70% and 25%, respectively) (2020b). The US is also a key nation for global processed food trade. In fact, the US is the top exporter of processed food commodities with an average of 16.19% of market share between 1980 and 2012 (Baiardi *et al* 2015). It is also the top importer nation for processed foods with 13% of market share in 2018 (Suanin 2021). Climate change is likely to influence domestic grain transport in both volumes and modes within the US (Attavanich *et al* 2013), making it important to understand current patterns.

The US also has widely available data to enable data-driven studies of its food supply chain. Of particular importance, the commodity flow survey (CFS) and freight analysis framework (FAF) provide comprehensive data on freight movement among the 132 FAF zones within the US, which represent states and major metropolitan areas. Note that the CFS/FAF data is only available every 5 years (years ending with '2' and '7') and for coarse commodity categories, rather than individual items. The

availability of CFS/FAF data within the US enabled Lin *et al* (2019) to develop the Food Flow Model to estimate food flows between counties. The Food Flow Model is a data-driven approach to estimate food flows between counties in the US that integrates machine learning, network properties, production and consumption statistics, mass balance constraints, and linear programming (Lin *et al* 2019). The Food Flow Model was initially developed for a single year (e.g. 2012). In this paper, we extend the Food Flow Model in time with some improvements to the original model.

This study builds upon the previous literature of agri-food flow modeling within the United States. Smith *et al* (2017) estimated the movement of corn used as biofuel among the counties of the US with a transportation optimization model. Lin *et al* (2014) analyzed the network properties of the state-level US food flows, finding a power-law relationship between node betweenness centrality and node degree, indicating potential network vulnerability to the disturbance of key nodes. Konar *et al* (2018) studied how the statistical network properties of US food flows compares with global food trade and village scale food sharing. Konar *et al* (2018) concluded that nodal mass flux follows a Gamma distribution across the full range of spatial scales, which means that there is high heterogeneity in the distribution of food mass, such that the majority of nodes exchange small masses of food, while some outlier nodes exchange large quantities of food. This observation was a key insight in the development of the Food Flow Model in Lin *et al* (2019), which preserves this high mass flux heterogeneity through maintenance of the Gamma distribution of mass flux at the county spatial resolution (Lin *et al* 2019). Note that all of these studies examined spatial patterns and did not consider the time trends in domestic food flows. We build on this literature by continuing to focus on fine grained spatial fluxes of food, but with the additional consideration of time.

The goal of this study is to estimate food flows between all county pairs in the United States through time. To do this, we apply an improved version of the Food Flow Model (Lin *et al* 2019), which is a data-driven model, to quantify high spatial resolution agri-food commodity fluxes within the US for the years 2007, 2012, and 2017. We then use our estimates to determine the core counties to the US food flow networks over time. The research questions that guide this study of food flows in the United States are: (a) How do the food flow network properties change over time? (b) Which counties and links participate in food flows over time? (c) What are the core counties over the study period? We briefly present our data, overview the Food Flow Model, and our improvements in section 2. Our findings are presented in section 3. We discuss our results in section 4 and conclude in section 5.

2. Methods

We extend the Food Flow Model developed by Lin *et al* (2019) to estimate food flows between US counties for the years 2007, 2012, and 2017. The Food Flow Model is a data-driven approach that incorporates logistic regression, gamma regression (with a gravity model structure), mass balance, and linear programming. A schematic of The Food Flow Model is provided in figure 1 (see the supporting information for a more detailed description of the original Food Flow Model). We introduce three key improvements: (a) systematic handling of estimator selection, (b) smoothing distance data, and (c) a quantitative approach to select the core nodes. We provide a brief overview of the Food Flow Model below; the interested reader is referred to Lin *et al* (2019) for a full description. We describe our model improvements in more detail below.

2.1. Input data

The Food Flow Model relies on empirical data to estimate county-level food flows. Input data include multiple factors such as the geography of production, transportation, input-output requirements, and consumption that combine to determine food transport (Lin *et al* 2019). In the model, consumption is not restricted to the purchase of final goods by households, but instead also accounts for the intermediate stages in supply chain production and processing. Hence, the transformation of commodities from raw to more refined items is also considered to be consumption. For example, live animals that are sent to a slaughterhouse are transformed into meat, so the counties containing the slaughterhouses consume the live animals and produce meat.

The FAF data is a key input to the Food Flow Model (2020b). FAF information is available for the years 2007, 2012, and 2017, as FAF3, FAF4, and FAF5, respectively. FAF provides information on the transfer of commodities within the 132 FAF zones of the US. The Food Flow Model relies on the FAF database which reports commodity flows by the standard classification of transported goods (SCTG). Thus, the Food Flow Model inherits the SCTG commodity categories and definitions. The agricultural and food commodities included in this study are listed in table 1.

Statistical information on the production, population, and personal income for each county and study year are used. For unprocessed agricultural commodities, production data is obtained from the US Department of Agriculture (2020a). For processed food items, per industry revenue in thousand US dollars is computed for each county by combining the SCTG-NAICS crosswalk table (Lin *et al* 2019), NAICS codes, and I-O accounts table (2020b). Since processed foods (SCTG 05-07) require industry inputs, we considered the corresponding SCTG 05-07

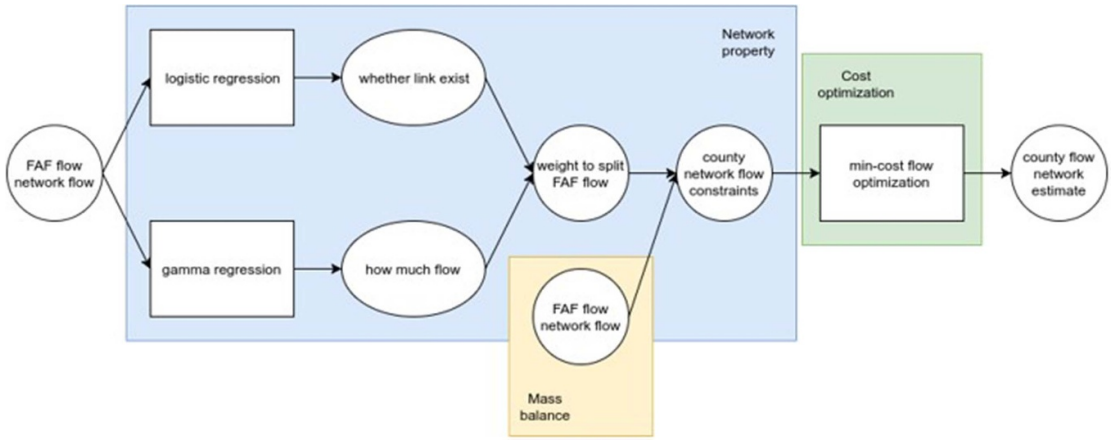

Figure 1. Schematic of the food flow model reproduced from Lin *et al* (2019). © The Author(s). Published by IOP Publishing Ltd CC BY 3.0. The food flow model is a data-driven framework that incorporates a variety of statistical methods.

Table 1. List of SCTG food commodity groups in this study.

SCTG code	Food commodity
01	Live animal and fish
02	Cereal grains
03	Agricultural products (except for animal feed, cereal grains, and forage products)
04	Animal feed, eggs, honey, and other products of animal origin
05	Meat, poultry, fish, seafood, and their preparations
06	Milled grain products and preparations, and bakery products
07	Other prepared foodstuffs, fats, and oils

production industry revenues as supply level indicators of counties.

The last set of required data is the geodesic distance between counties obtained from Oak Ridge National Laboratory (2020a) and port trade data from the US Bureau of Transportation Statistics (2020). The geodesic distance is a simplification of real-world transportation path lengths which is a commonly used distance measure for the gravity model of trade (De Benedictis and Tajoli 2011, Shepherd 2013). Port data is used to boost the food flows to/from the counties that contain the ports as they are assumed to be the main transit hubs for import and export. All input data along with a brief description of how they are utilized in the Food Flow Model are listed in table 2. Additional details regarding input data are provided in the SI (available online at stacks.iop.org/ERL/17/034035/mmedia).

2.2. Improvements to the food flow model

We introduce three key improvements to the original Food Flow Model:

(a) Systematic handling of estimator selection. The logistic and gamma regression components of the Food Flow Model are the only part of the methods that are sensitive to parameter values. A high number of variables might cause overfitting and introduce noise to the regression, while a low number of variables could decrease the model accuracy. To balance overfitting with accuracy and improve model reproducibility, we implement a grid-based search approach in the improved version of the Food Flow Model. This allows us to achieve area under the curve (AUC) measures in the 0.78–0.94 range across our study time period and commodity categories with the number of variables between 7 and 10 out of 130 possibilities.

(b) Smoothing distance data. The value of trade is negatively correlated with distance in the gravity model (Shepherd 2013). We winsorize the inter-county distance values (Ghosh and Vogt 2012) to avoid high flow estimates between small counties. Winsorization is a common technique in economics to modify the values of outliers to bring them closer to the other sample values (Hwang *et al* 2011, Orth 2013). Additionally, large self-loop flows were driven by counties with small areas in the original Food Flow Model. To avoid this issue, self-loop area is now set equal to the mean of all self-loop distances to remove the effect of extremely small counties. Flow assignment is now mainly driven by the other variables, such as production and income (see SI for a more detailed explanation).

(c) Quantitative approach to select the core nodes. To determine the core nodes over time we adopt TOPSIS, a multi-criteria decision analysis technique. TOPSIS is a commonly used approach to determine the importance of network components (Du *et al* 2014, Hu *et al* 2016, Karakoc *et al* 2020) as it ranks the components based on their pre-determined criteria performances (Hwang and Yoon 1981). We use node betweenness centrality and degree (Ercsey-Ravasz *et al* 2012,

Table 2. A list of the data that is used in this study.

Name	References	Description	Purpose
Freight Analysis Framework (FAF) Version 3, 4, 5	Oak Ridge National Laboratory (2020b)	Freight movement between FAF zones with information on origin destination pair, commodity type, transportation-mode and value both in US dollars and thousand tonnes.	Commodity flow in FAF data is used for (i) developing the regression models to estimate the food flows, and (ii) introducing the mass balance constraint for county food flows.
Population	United States Census Bureau (2020a)	Population data per county.	Population data is included in the regression models to estimate the county demand levels.
Personal Income	US Bureau of Economic Analysis (2020a)	Personal income data per county.	Personal income data is included in the regression models combined with Input–Output Accounts table to determine the final county demand levels per SCTG.
Unprocessed Food and Livestock Production	United States Department of Agriculture (2020a)	Agricultural production data for each crop type and inventory counts for livestock per county.	The US Department of Agriculture census is included in the gamma regression models to estimate the food flow amounts between counties.
Input–Output Accounts Data	US Bureau of Economic Analysis (2020b)	This supply chain data represents one industry's requirement degree for another industry's output. The required input is to produce unit dollar output.	The requirement coefficients of SCTG 05-07 are multiplied by production data of SCTG 01-04 to determine the need of each industry per commodity. The sum of industry and end consumer input needs per commodity represents county's total consumption. This is used in gamma mixture hurdle model for link prediction and flow amount estimation.
Port Trade	US Bureau of Transportation Statistics (2020)	Data for commodity trade from sea, air, and land ports in the United States	The counties where these ports are located in are considered as the transit hubs for import and export. Port trade values of SCTG commodities (in US dollars) are used in gamma regression models to boost up the flows of food to/from these certain counties.
North American Industry Classification System (NAICS)	United States Census Bureau (2020c)	This production-oriented NAICS data groups industries according to similarity in their production processes.	NAICS data, I-O Accounts table, and SCTG-NAICS crosswalk table are used together with agricultural production data to compute industry specific consumption of each SCTG food commodity. By using the consumption levels as input requirements per one dollar output, county revenues are calculated for each NAICS code.
SCTG-NAICS Crosswalk table	Food flows between US counties (Lin *et al* 2019)	This table matches NAICS codes with SCTG codes based on the industry input requirement of commodities.	This table is used with I-O Accounts table to compute NAICS code specific county revenue. The computed industry revenue data is included in the gamma regression models to estimate the flow amounts between counties.
County-to-County Distance Matrix	Oak Ridge National Laboratory (2020a)	The geodesic distance between county centroids based on their latitude and longitude information.	The regression models consider distances between all county-pairs to estimate the existence and strengths of the links. Also, linear programming algorithm uses distance matrix to assign the food flows to the shortest paths.
Land Area Coverage Per County	United States Census Bureau (2020b)	The land area coverage in square miles per US census area.	In the regression models, the land area in per county is used as the distance measures for self-loops. The square are in miles is converted into meters.

Table 3. List of the original and improved components of the food flow model.

Original model	
Model component	Functionality
Logistic regression	Binary logistic regression is used to estimate the existence of flow links between FAF zones based on the available independent variables. It estimates two possible outcomes and then the outcome values are coded as either '1' or '0' indicating the 'existence' or 'non-existence' of a link between any two FAF zone, respectively. Once the desired level of model accuracy is achieved in the FAF scale, it is implemented on the county scale.
Gamma regression	Since food mass fluxes follow the gamma distribution (Konar *et al* 2018), gamma regression is used to estimate the flow capacities (i.e. weights) of the links between counties based on the available independent variables. The shape of the gamma distribution could be interpreted as the effective units of food commodity that is actually delivered from origin to destination apart from the wasted amount during transit (Lin *et al* 2019). Through the gamma regression model, between 3% and 10% of link weights across all SCTG commodities are underestimated. To satisfy the total mass balance of FAF data, a separate gamma regression model is developed for underestimated links to boost their weight and achieve mass balance. To boost the weight estimates, available port data is introduced to the second gamma regression as a separate variable. Similar to logistic regression, the two gamma regressions are first implemented on FAF scale and then on the county scale.
Linear programming	Linear programming component aims to minimize the total transportation cost (i.e. travel distance per unit commodity) in the county food flow network. It is a common approach in supply chain transportation studies (Klein 1967). As it is based on the gravity-model, l Linear programming also assigns food flows between counties to the link with largest capacity. It is another common approach in supply chain studies to deliver the goods among origin and destination pairs in the most efficient way (Chen *et al* 1999, Schrijver 2002).
Mass balance	Mass balance is introduced as a constraint in the linear programming component of the Food Flow Model. Sum of the total outflow/inflow of counties that are located in a single FAF zone is matched with total outflow/inflow of that corresponding FAF zone. Additionally, mass balance is included in the gamma regression. Sum of the estimated link weights is compared with the empirical total flows per FAF zone. Once the underestimated link weights are identified, a second gamma regression is implemented on them to match the total FAF mass in the estimations.
Gravity-model	The gravity model of international trade proposes that the trade flows are inversely correlated with the distances between origin and destination pairs (Disdier and Head 2008). Both the regressions and linear programming components of Food Flow Model is based on the gravity-model of trade. Hence, shorter distances between counties are assigned with higher flow capacities.
Model improvements	
Model component	Functionality
Grid-based search	A grid-based search is introduced to the logistic and gamma regression components of the Food Flow Model. By using grid-based search, the trade-off between overfitting due to high number of estimators and underfitting due to low number of estimators is balanced. Accuracy measure (AUC) in range 0.78–0.94 is achieved with 7–10 estimators for each commodity for each study year.
Winsorizing	To avoid extremely high flow estimates between small counties, winsorizing technique is introduced to the Food Flow Model. By winsorizing technique, value of extremely small outliers, 1.5% of the inter-county distances, are brought up the first quantile of the distribution. Hence, effect of extremely small distances on flow estimation is overpowered by other estimators i.e. production, population, etc.
TOPSIS	A multi-criteria decision analysis technique, TOPSIS, is introduced to the Food Flow Model in order to methodologically identify core nodes. Counties are ranked based on their score for two criteria, node degree and betweenness centrality. For each commodity network in each study year, counties with highest aggregated score for degree and betweenness are determined as core.

Gaur *et al* 2020) (see section 2.3) to assess the core counties in each study year (see SI for more details).

Table 3 lists the original and improved components of the Food Flow Model. Additional details are provided in the SI.

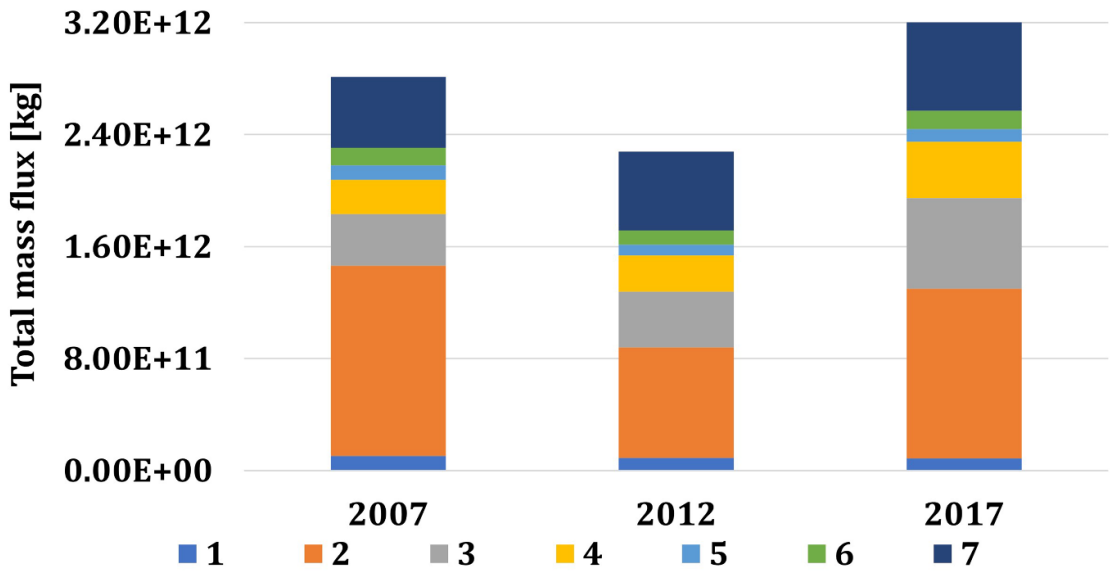

Figure 2. Total mass flux of food flow networks over time by SCTG commodity. Note the mass flux increases from 2007 to 2017, but is lowest in 2012, which was a drought year in the Corn Belt and Central Valley. The total mass flux of county and FAF networks are perfectly matched due to the mass balance constraint of the Food Flow Model.

2.3. Food flow networks

We construct food flow networks for FAF data and our county-scale estimates. Nodes (N) are the spatial locations that serve as the origin and destination of food flows (e.g. FAF zones, counties). Links (L) indicate connections between origin (o) and destination (d) nodes. Link weight is the mass flux between nodes. Density (d) is the ratio of existing links over the potential number of links, including self-loops: $d = \frac{L}{N*N}$. The core nodes are defined to be those with both high total degree and betweenness centrality. Total degree is $c^o_{total} = \sum_d^N l_{od} + \sum_d^N l_{do}$ (Barabási 2016). Node betweenness centrality is the portion of network shortest paths, σ, that pass through that node over all potential shortest paths in the network and is given by $b^o = \sum_{s \neq o \neq t} \frac{\sigma^o_{st}}{\sigma_{st}}$. Nodes with higher betweenness centrality are located on more shortest paths in the network and are more central to the national network structure (White and Borgatti 1994). We explain our multi-criteria decision approach to core node selection in the SI.

3. Results

In this section, we answer the research questions listed in section 1. We compare county estimates with FAF data to address each question.

3.1. How do the food flow network properties change over time?

There are 123 nodes (i.e. FAF zones) in the FAF3 data for 2007; the number of nodes increases to 132 in 2012 and 2017. This means that there are 15 129 potential links at the FAF zone spatial scale in 2007 and 17 424 potential links in the other study years. There are 3134 counties in each year of the study, so there are 9 821 956 potential links (including the self-loops) at the county spatial scale (see tables 7 and 8 in the SI for FAF and county level network statistics through time, and a list of eight counties excluded due to missing data).

The FAF network is densest in 2007, because the increase in the number of nodes offsets the gains in the number of links with time. The total mass increased from 2007 to 2017 (see figure 7), with the exception of 2012, which was a drought year in the Corn Belt and Central Valley (Dall'Erba *et al* 2021). These observations taken together indicate that FAF networks have become more spatially concentrated over time, with fewer links carrying larger quantities of food. The mass flux trend is preserved in the county flow networks due to our methodological constraint of mass balance between county and FAF flows. County networks are sparser than FAF networks. However, county network density mostly increases with time, with some variation. The density of SCTG 01, 03, 04, 06, and 07 increases from 2007 to 2017, while the density of SCTG 02, 05, and all commodities summed up decreases over the same period. The density is the most variable in the drought year 2012. We explore this drought year in more detail in section 4.2.

3.2. Which counties and links participate in food flows over time?

Figure 3 illustrates our estimates for 2017 county-level food flow networks broken down by each SCTG commodity code. Table 4 lists the locations with the largest net mass flow (= total inflows—total outflows) (see the SI for the break down by SCTG).

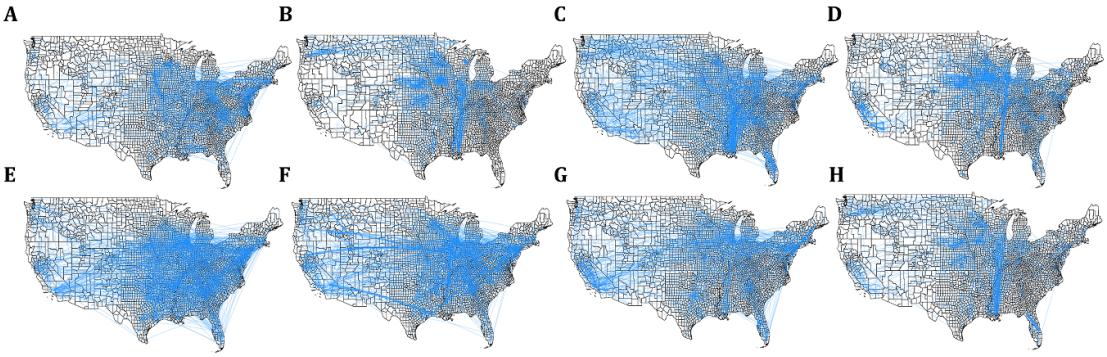

Figure 3. Estimated county-scale food flow maps in 2017 for (A) live animals and fish (SCTG 01), (B) cereal grains (SCTG 02), (C) agricultural products (except for animal feed, cereal grains, and forage products) (SCTG 03), (D) animal feed, eggs, honey, and other products of animal origin (SCTG 04), (E) meat, poultry, fish, seafood, and their preparations (SCTG 05), (F) milled grain products and preparations, and bakery products (SCTG 06), (G) other prepared foodstuffs, fats and oils (SCTG 07), and (H) all agri-food commodities summed together. Top 2000 links are plotted, and link weights are scaled by the maximum mass flux per SCTG commodity in each map.

Table 4. Top 10 net flow FAF zones and counties for all agri-food commodities by year. Net flow is the difference between inflows and outflows. Higher inflow leads to higher net flow per FAF zone/county.

2007			
FAF	Mass (kg)	County	Mass (kg)
New Orleans, LA CFS Area	5.94×10^{10}	Orleans County, LA	3.56×10^{10}
Minneapolis-St. Paul MN-WI CFS Area (MN Part)	1.80×10^{10}	Black Hawk County, IA	1.03×10^{10}
Remainder of Texas	1.78×10^{10}	Harris County, TX	9.48×10^{9}
Seattle, WA CFS Area	1.51×10^{10}	Sedgwick County, KS	9.02×10^{9}
Houston, TX CFS Area	1.50×10^{10}	Olmsted County, MN	8.30×10^{9}
Remainder of Pennsylvania	1.47×10^{10}	Jefferson County, LA	8.21×10^{9}
Denver, CO CFS Area	1.15×10^{10}	Kern County, CA	7.95×10^{9}
Remainder of Alabama	9.56×10^{9}	Randolph County, IL	6.63×10^{9}
Chicago IL-IN-WI CFS Area (IL Part)	9.09×10^{9}	Maricopa County, AZ	6.55×10^{9}
Indianapolis, IN CFS Area	8.66×10^{9}	King County, WA	5.17×10^{9}
2012			
FAF	Mass (kg)	County	Mass (kg)
New Orleans-Metairie-Hammond, LA CFS Area	3.07×10^{10}	Winnebago County, IL	1.17×10^{10}
Remainder of Texas	1.43×10^{10}	Nobles County, MN	1.08×10^{10}
Remainder of Mississippi	1.24×10^{10}	Platte County, NE	1.04×10^{10}
Omaha-Council Bluffs-Fremont, NE CFS Area	1.23×10^{10}	Webster County, IA	1.01×10^{10}
Corpus Christi-Kingsville-Alice, TX CFS Area	1.22×10^{10}	Kandiyohi County, MN	9.42×10^{9}
Portland-Vancouver-Salem, WA CFS Area	1.09×10^{10}	Buena Vista County, IA	8.62×10^{9}
Chicago-Naperville, IL CFS Area	1.06×10^{10}	Dodge County, NE	8.27×10^{9}
Los Angeles-Long Beach, CA CFS Area	1.01×10^{10}	Washington County, LA	7.92×10^{9}
Houston-The Woodlands, TX CFS Area	9.64×10^{9}	Waukesha County, WI	7.76×10^{9}
Remainder of Georgia	8.72×10^{9}	Dawson County, NE	6.66×10^{9}
2017			
FAF	Mass (kg)	County	Mass (kg)
New Orleans-Metairie-Hammond, LA CFS Area	7.94×10^{10}	Polk County, IA	8.84×10^{10}
Baton Rouge, LA CFS Area	2.09×10^{10}	Orleans County, LA	5.50×10^{10}
Portland-Vancouver-Salem, WA CFS Area	1.85×10^{10}	Riley County, KS	2.11×10^{10}
Chicago-Naperville, IL CFS Area	1.74×10^{10}	Clark County, WA	1.66×10^{10}
Atlanta-Athens-Clarke County-Sandy Springs, GA CFS Area	1.70×10^{10}	Platte County, NE	1.57×10^{10}
Remainder of Florida	1.54×10^{10}	Sedgwick County, KS	1.41×10^{10}
Remainder of Texas	1.35×10^{10}	Buffalo County, NE	1.34×10^{10}
Seattle-Tacoma, WA CFS Area	1.15×10^{10}	McLean County, IL	1.26×10^{10}
Los Angeles-Long Beach, CA CFS Area	1.14×10^{10}	East Baton Rouge County, LA	1.12×10^{10}
St. louise-St. Charles-Farmington, IL CFS Area	1.01×10^{10}	Dawson County, NE	1.11×10^{10}

Some FAF zones are common over time, including: New Orleans, LA CFS Area; Seattle, WA CFS Area; and Chicago, IL CFS Area. These FAF zones are consistently important producers and transport hubs for agri-food. The majority of the largest net flow counties are within large net flow FAF zones. Some of the common top counties through time are: Orleans County, LA; Sedgwick County, KS; and Platte County, NE. Orleans County, LA contains the New Orleans port which is a major port for the trade of agri-food commodities via the Mississippi River (NOL 2020). Sedgwick, KS has a large agricultural economy that is built on food processing (KDA 2020). Platte County, NE is in the list of top 5 Nebraska counties for agricultural sales (NDA 2020).

Table 5 provides the top 10 links in county and FAF flow networks for all SCTG commodities summed up (see the SI for a break down by SCTG commodity). Many of the top 10 links are self-loops, which is in line with previous results presented in Lin *et al* (2019). Many of the largest county-level links are located in the FAF zones with the largest food flows. However, some county-level links exceed their respective FAF level ranking. For example, Los Angeles County, California to Los Angeles County, California is consistently in the top 10, although Los Angeles-Long Beach CFS Area is not always in the top FAF links. This is sensible because counties in California are bigger in area than counties in the Midwest and Eastern parts of the US, which mechanically means that allocating to them will lead to a larger mass.

Figure 4 presents a heatmap of FAF and county-scale net flows over time. Changes in county-scale net flows between study years are also illustrated. FAF zones and counties that import more have higher net flows whereas the ones that export more have lower net flows. Southern and Eastern FAF zones bring in more food (high net flows) and this spatial pattern is constant through time. In contrast, more rural FAF zones in the Midwest and Northwest send out more food (low net flows). Similar trends are observable in the county-scale maps. Wealthy population centers such as Los Angeles, CA, Chicago, IL, and New York City, NY have higher net flows through time. (Refer to the SI for heatmaps of net flow by SCTG, as well as separate heatmaps for inflows and outflows through time.)

Figure 5 maps the change in county food flows over time. By design, the county-scale networks capture the spatial patterns of the FAF data (e.g. compare figure 5(A) with (E); figure 5(B) with (F)). Link-level changes in mass flux between counties align well with the FAF-level data (e.g. compare figure 5(C) with (G); figure 4(D) with (H)). For example, the Mississippi River-band experiences some of the largest fluctuations in food flows across spatial scales. There is also a relatively high increase in food flows within the counties of Florida and inter-county flows of Washington. These maps align with the heatmaps of county inflow and outflow changes. Food flow network maps broken-down by SCTG commodity are provided in the SI.

3.3. What are the core counties over the study period?

Table 6 lists the FAF zones and counties that are core to the national network through time. Core locations tend to remain so through time, which indicates persistence in their importance. The consistently core FAF zones are: Los Angeles, CA CFS Area, Chicago-Naperville CFS Area (IL Part), and Remainder of Texas. The consistently core counties are: Los Angeles County, CA, Cook County, IL, Maricopa County, AZ, Shelby County, TN, Riverside County, CA, Bexar County, TX and Harris County, TX. Core FAF zones and counties—which are defined in terms of their topological importance—are also major movers of food mass (see table 16 in SI). This means that these FAF zones and counties are important in terms of both their contribution to network structure and mass flux. Refer to the SI for the core FAF zones and counties by SCTG commodity.

Figure 6 illustrates that a power-law relationship exists between node degree and betweenness centrality. The power-law relationship is stronger in dense networks and weaker in sparse networks, such as SCTG 01 and 02. This means that the identification of core counties is less clear for SCTG 01 and SCTG 02. Refer to the SI for the power-law fit by SCTG commodity. Importantly, the power-law relationship—which has been observed in empirical food flow data (Lin *et al* 2014, Konar *et al* 2018)—was not predetermined by our modeling approach. This means that a power-law relationship between node degree and betweenness centrality naturally arises through our algorithm. The self-arising power-law indicates that our model captures the critical attributes of food flow networks and gives us more confidence in our approach.

4. Discussion

Here, we discuss the advantages and limitations of our model, estimates during the 2012 Corn Belt drought, and future research directions.

4.1. Advantages and limitations of the food flow model

The Food Flow Model is a data-driven approach to estimate food flows at a fine spatial resolution through time. The data-driven nature of our approach is both a limitation and an advantage. It is a limitation because we do not explicitly include mechanisms that would enable us to address *why* questions about food flows. Rather, we describe the *who*, *what*, and *where* of food flows with time. Yet, the empirical patterns that the Food Flow Model incorporates

Table 5. Top 10 links in FAF and county networks for all SCTG commodities summed up for 2007, 2012, and 2017. Top 10 links are listed in the descending order of total mass (kg).

2007			
FAF link	Mass (kg)	County link	Mass (kg)
Iowa → Iowa	1.64×10^{11}	Los Angeles County, CA → Los Angeles County, CA	1.89×10^{10}
Remainder of Illinois → Remainder of Illinois	1.03×10^{11}	Maricopa County, AZ → Maricopa County, AZ	1.07×10^{10}
Nebraska → Nebraska	9.70×10^{10}	Palm Beach County, FL → Palm Beach County, FL	9.12×10^{9}
Remainder of Kansas → Remainder of Kansas	8.46×10^{10}	O'Brien County, IA → Black Hawk County, IA	8.46×10^{9}
Remainder of Minnesota → Remainder of Minnesota	7.64×10^{10}	Canyon County, ID → Canyon County, ID	7.60×10^{9}
Remainder of Texas → Remainder of Texas	7.00×10^{10}	Stainlaus County, CA → Stainlaus County, CA	7.60×10^{9}
Remainder of Wisconsin → Remainder of Wisconsin	6.04×10^{10}	Cook County, IL → Cook County, IL	6.91×10^{9}
Remainder of California → Remainder of California	5.42×10^{10}	Pope County, MN → Pope County, MN	6.86×10^{9}
North Dakota → North Dakota	4.77×10^{10}	Sangamon County, IL → Randolph County, IL	6.82×10^{9}
South Dakota → South Dakota	4.62×10^{10}	Kings County, CA → Kern County, CA	6.52×10^{9}
2012			
FAF link	Mass (kg)	County link	Mass (kg)
Remainder of Iowa → Remainder of Iowa	1.32×10^{11}	Burke County, ND→Burke County, ND	1.76×10^{10}
Remainder of Nebraska → Remainder of Nebraska	8.43×10^{10}	Red Lake County, MN → Red Lake County, MN	1.75×10^{10}
Remainder of Minnesota → Remainder of Minnesota	7.71×10^{10}	Mills County, IA → Mills County, IA	1.73×10^{10}
Remainder of Illinois → Remainder of Illinois	7.10×10^{10}	Elmore County, ID → Elmore County, ID	1.30×10^{10}
Remainder of North Dakota → Remainder of North Dakota	4.72×10^{10}	Gallatin County, IL→Gallatin County, IL	1.25×10^{10}
Remainder of Kansas → Remainder of Kansas	4.31×10^{10}	Los Angeles County, CA→Los Angeles County, CA	1.21×10^{10}
Remainder of South Dakota → Remainder of South Dakota	4.13×10^{10}	Pierce County, NE→Pierce County, NE	5.98×10^{9}
Remainder of Texas → Remainder of Texas	3.80×10^{10}	Lewis County, MO →Lewis County, MO	5.16×10^{9}
Remainder of Idaho → Remainder of Idaho	3.13×10^{10}	Red River County, LA→Red River County, LA	5.11×10^{9}
Los Angeles-Long Beach CFS Area → Los Angeles-Long Beach CFS Area	3.09×10^{10}	Fresno County, CA→Fresno County, CA	4.39×10^{9}
2017			
FAF Link	Mass (kg)	County Link	Mass (kg)
Remainder of Iowa → Remainder of Iowa	2.15×10^{11}	Linn County, IA → Linn County, IA	2.37×10^{10}
Remainder of Nebraska → Remainder of Nebraska	1.35×10^{11}	Gallatin County, IL → Gallatin County, IL	2.36×10^{10}
Remainder of Illinois → Remainder of Illinois	1.06×10^{11}	Blaine County, ID → Blaine County, ID	1.42×10^{10}
Remainder of Minnesota → Remainder of Minnesota	1.04×10^{11}	Davison County, SD → Davison County, SD	1.18×10^{10}
Remainder of Texas → Remainder of Texas	7.84×10^{10}	Armador County, CA → Armador County, CA	1.05×10^{10}
Remainder of Kansas → Remainder of Kansas	7.23×10^{10}	Los Angeles County, CA → Los Angeles County, CA	8.55×10^{9}
Remainder of South Dakota → Remainder of South Dakota	6.11×10^{10}	Lafayette County, FL →Lafayette County, FL	6.31×10^{9}
Remainder of North Dakota → Remainder of North Dakota	5.82×10^{10}	New York County, NY → New York County, NY	6.00×10^{9}
Remainder of Wisconsin → Remainder of Wisconsin	5.28×10^{10}	Riverside County, CA → Riverside County, CA	5.91×10^{9}
Remainder of California → Remainder of California	4.96×10^{10}	San Bernardino County, CA → San Bernardino County, CA	4.69×10^{9}

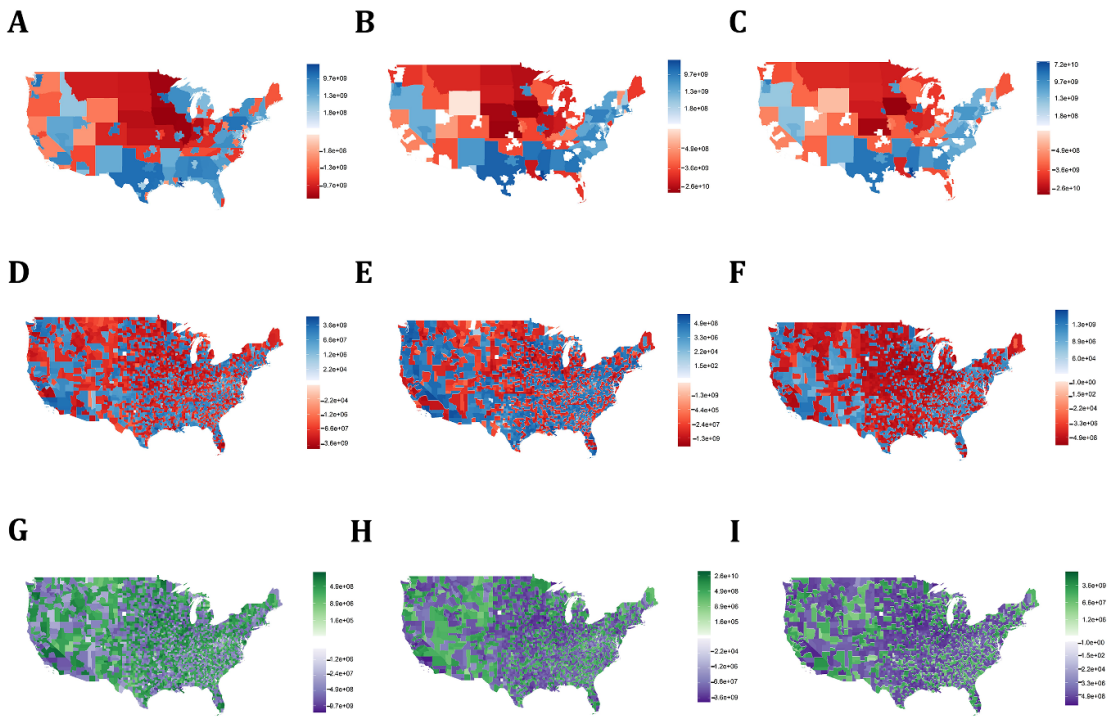

Figure 4. Net food flows (kg) for FAF Zones and counties over time for all SCTGs summed up. (A) 2007 net flows in FAF, (B) 2012 net flows in FAF, (C) 2017 net flows in FAF, (D) 2007 net flows in county, (E) 2012 net flows in county, (F) 2017 net flows in county. Net flows are defined as the difference between inflows and outflows. Positive net flows are blue colored, negative net flows are red colored. (G) Change in net flows from 2007 to 2012, (H) change in net flows from 2012 to 2017, (I) change in net flows from 2007 to 2017. Increases in net flow are 'green' colored (e.g. more inflows) and decreases in net flow are 'purple' colored (e.g. more outflows).

have mechanistic explanations in certain cases, such as the gravity model of trade (Anderson 1979), which is incorporated in our methodology (see section 2). The data-driven nature of our model also means that input data availability represents a limitation to its scope. This is the reason that our study time series is restricted to the years 2007, 2012, and 2017, for example.

The main shortcoming of our study is the absence of ground truth data to validate the county food flow estimates. Yet, our data-driven approach employs a variety of measures to ensure that our county-level estimates are bounded by reality (e.g. mass balance requirement). Additionally, we introduced model improvements to limit human error and enhance realism. Grid-based search was used to automate variable selection and TOPSIS was used for core node selection (see table 3). This reduces the potential for human bias and error with a scientific method for these model components. The additional improvements that we made in the handling of outliers (i.e. winsorizing distances and setting self-loop distances constant) further enhances the realism of model outputs. Flows of relatively large counties (in terms of income, population, production, and other distance-unrelated characteristics) are now estimated to be higher than relatively small counties. This represents an improvement to the original Food Flow Model, in which the flow values of some counties was likely overestimated due to their very small size and the previous way that self-loop distance was handled. Additionally, the self-arising power-law relationship between degree and betweenness centrality (see section 3.3) provides additional confidence in model performance.

The main advantage of our model is the provision of county-level estimates of food flows. We provide an example of the output of our model in figure 7 for Los Angeles County, CA and Hillsborough County, FL, which are both in the list of core nodes (see table 6). Figure 7 illustrates that we are able to map the food mass inflows and outflows for each county in the US per commodity per study year. Similar maps could be generated for each of the 3134 counties by SCTG-year. Researchers and policy makers could evaluate a specific location of interest over time. However, our estimates are best suited for national-level analyses and local-level decision-makers may want to augment our estimates with additional site-specific information.

The total mass in/outflow of counties are constrained to sum to the mass in/outflow of their corresponding FAF zone. However, the mass in/outflow is heterogeneously allocated to links between counties according to our regression models. Regression models are fit to each SCTG-year and provided in the 'Regression Models and Network Statistics.xlsx' spreadsheet in the SI. For example,

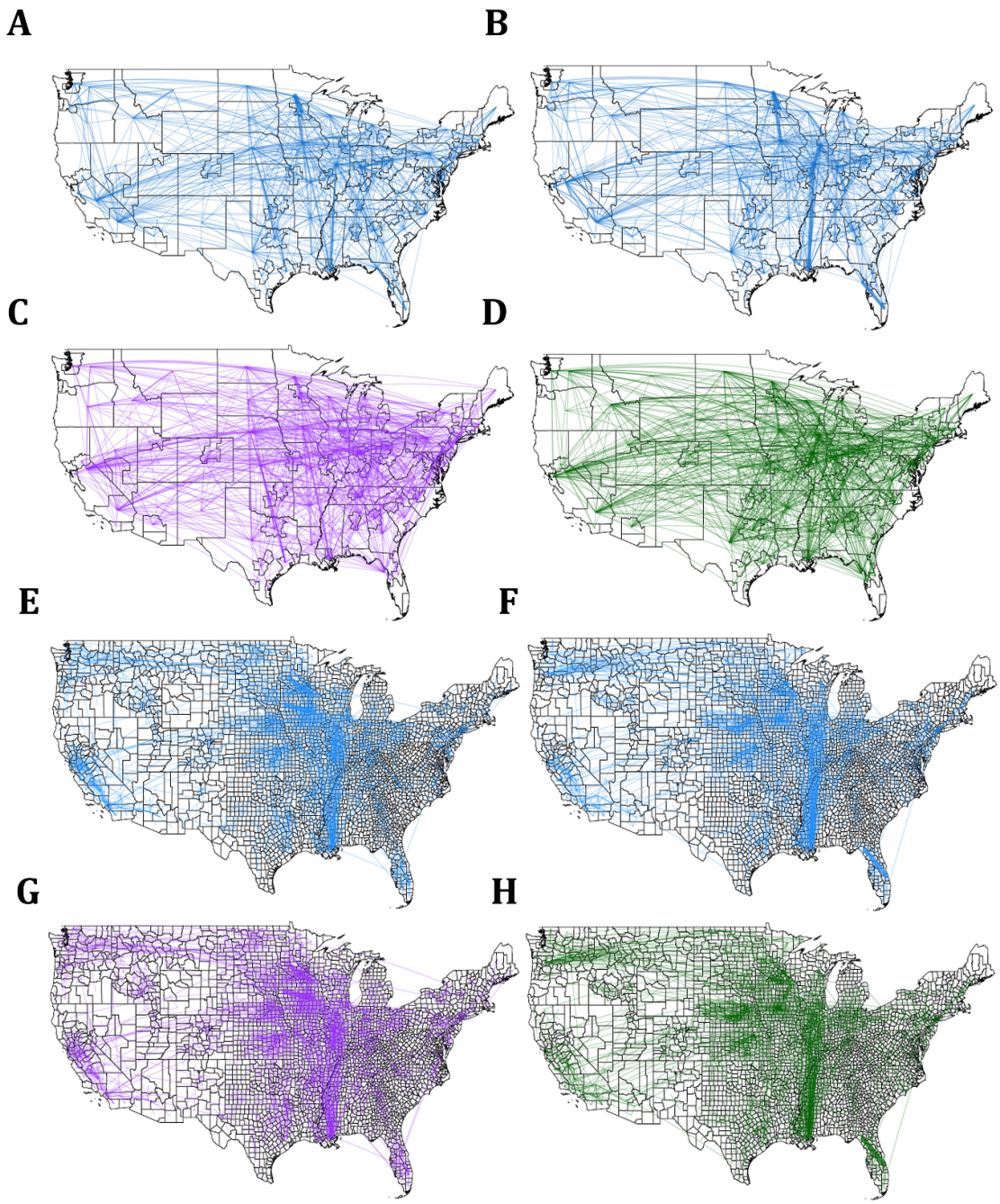

Figure 5. Food flows of all SCTG commodities summed up for 2007, 2017 and change in mass between 2007 and 2017. In FAF-level networks top 1000, in county-level networks top 2000 links are drawn. (A) 2007 FAF flows, (B) 2017 FAF flows, (C) FAF flows mass decrease, (D) FAF flows mass increase, (E) 2007 county flows, (F) 2017 county flows, (G) county flows mass decrease, and (H) county flows mass increase.

the regression equation for SCTG 02 for grains is: $\ln(E(F2)) = -0.74 \log(D) - 0.18 \log(GDP_o) + 0.41 \log(P_o) + 0.13 \log(A1_d) + 0.45 \log(D3_d) - 0.15 \log(S1_d) + 0.14 \log(C1_d) - 0.06 \log(T3_d) + 0.24 \log(LIVE_d)$. This means that grain flows will be allocated to inter-county links according to county-level regressor values for the distance between counties (D), personal income of the origin county (GDP_o), grain production of the origin county (P_o), revenue for accommodation in the destination county ($A1_d$), revenue for drinking places (alcoholic beverages) in the destination county ($D3_d$), revenue for scientific research and development services industries in the destination county ($S1_d$), cattle population in the destination county ($C1_d$), turkey population in the destination county ($T3_d$), and total livestock population in the destination county ($LIVE_d$). The heterogeneity in the spatial and temporal distribution of regressor variables explains the differences in food flows between counties and FAF zones with time. Maps of all regressor variables are provided in the SI.

Table 6. The core FAF zones and counties in the US food flow network. These FAF zones and counties are structurally important to the national total agri-food flows in each study year.

2007					
Rank	FAF Name	CFS Code	Rank	County Name	FIPS
1	Chicago IL-IN-WI CFS Area (IL Part)	17-176	1	Cook County, IL	17 031
2	Los Angeles, CA CFS Area	06-348	2	Maricopa County, AZ	4013
			3	Los Angeles County, CA	6037
			4	Shelby County, TN	47 157
			5	Harris County, TX	48 201
			6	Orleans County, LA	22 071
			7	Johnson County, KS	20 091
			8	Bexar County, TX	48 029
			9	Clark County, NV	32 003
			10	Dallas County, TX	48 113
			11	Hennepin County, MN	27 053
			12	Jackson County, MO	29 095
			13	Riverside County, LA	6065
2012					
Rank	FAF Name	CFS Code	Rank	County Name	FIPS
1	Los Angeles-Long Beach, CA CFS Area	06-348	1	Los Angeles County, CA	6037
2	Remainder of Texas	48-99999	2	Shelby County, TN	47 157
3	Houston-Woodlands, TX CFS Area	48-288	3	San Bernardino County, CA	6071
4	Chicago-Naperville CFS Area (IL Part)	17-176	4	Dallas County, TX	48 113
5	San Jose-San Francisco-Oakland, CA CFS Area	06-488	5	Harris County, TX	48 201
6	New York-Newark CFS Area (NY Part)	36-408	6	Maricopa County, AZ	4013
			7	Cook County, IL	17 031
			8	Riverside County, CA	6065
			9	Bexar County, TX	48 029
2017					
Rank	FAF Name	CFS Code	Rank	County Name	FIPS
1	Remainder of Texas	48-99999	1	Riverside County, CA	6065
2	Los Angeles-Long Beach, CA CFS Area	06-348	2	Maricopa County, AZ	4013
			3	San Bernardino County, CA	6071
			4	Los Angeles County, CA	6037
			5	San Diego County, CA	6073
			6	Shelby County, TN	47 157
			7	Cook County, IL	17 031
			8	Hillsborough County, FL	12 057
			9	Harris County, TX	48 201
			10	Bexar County, TX	48 029
			11	Dallas County, TX	48 113

4.2. Impact of the 2012 Corn-Belt drought on grain flows

In 2012 a severe drought hit the US Corn-Belt, which is a highly productive region for grain (Boyer *et al* 2013). The 2012 drought led to a 55% variation in corn yield across the region (Wan *et al* 2015). Illinois, Iowa, Indiana, Minnesota, and Nebraska were the main states impacted by the drought (Wu *et al* 2015, Prokopy *et al* 2017). Here, we examine our estimates of grain (SCTG 02) flow changes during the drought for the Corn-Belt.

The drought effect is captured in our model since the total mass flux of grain in 2012 is the lowest among all study years (see figure 2). The Corn-Belt remains a top region for grain outflow in 2012, despite the drought (see table 18 in SI). However, the mass of grain outflows are lower in 2012 due to decreased grain production during the drought. Grain outflows from the Corn-Belt decreased by 3.29×10^{11} kg (53.18%) from 2007 to 2012, while inflows also decreased (2.84×10^{11} kg; 53.67%). Across FAF zones, the mass of grain outflow is 15.28%–68.05% lower in 2012 than in 2007. As in figure 40 in the SI, grain inflows also decrease from 2007 to 2012. Corn-Belt grain outflows and inflows rebound following the drought in 2017: outflows increase by 11.67%–129.65% and inflows increase by 30.72%–114.22% across FAF zones from 2012 to 2017, primarily due to an increase in self-loops (see table 25 in SI). This indicates the importance of grain production and processing within the regional agricultural economy.

Link-level changes in grain flows are mapped in figure 8. The largest mass decreases in grain links are concentrated in the Corn-Belt FAF zones (see figure 8(A), which indicates more internal movement in a non-drought year. From 2007 to 2012, mass increases in cereal grain flows to/from the Corn-Belt are connected with the rest of the nation, such as California, Texas, and Mississippi FAF zones (figure 8(B)).

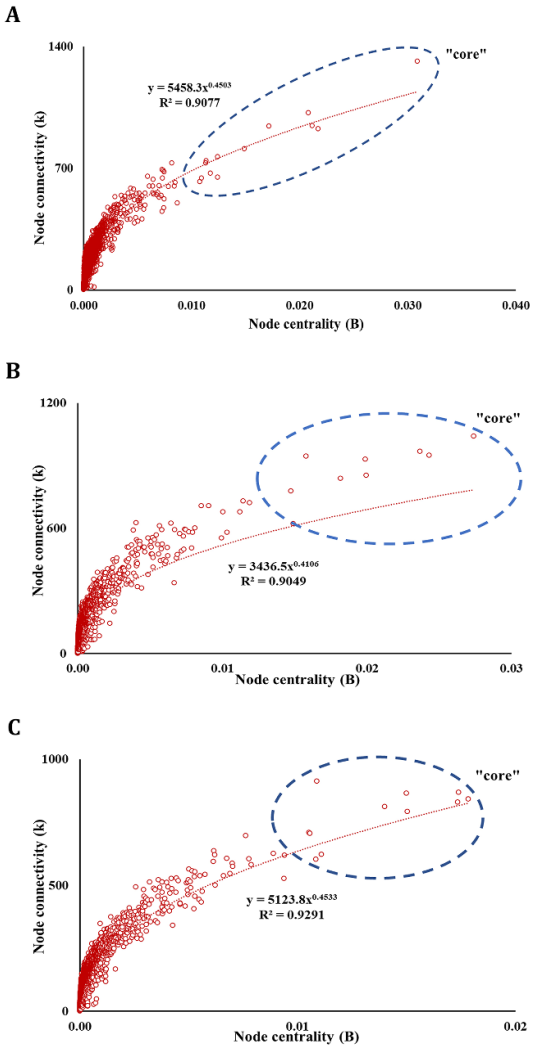

Figure 6. The relationship between node degree (k) and betweenness centrality (B) follows a power-law through time. These plots are for all SCTGs summed up for (A) 2007, (B) 2012, and (C) 2017. The blue set indicates the core nodes of the network.

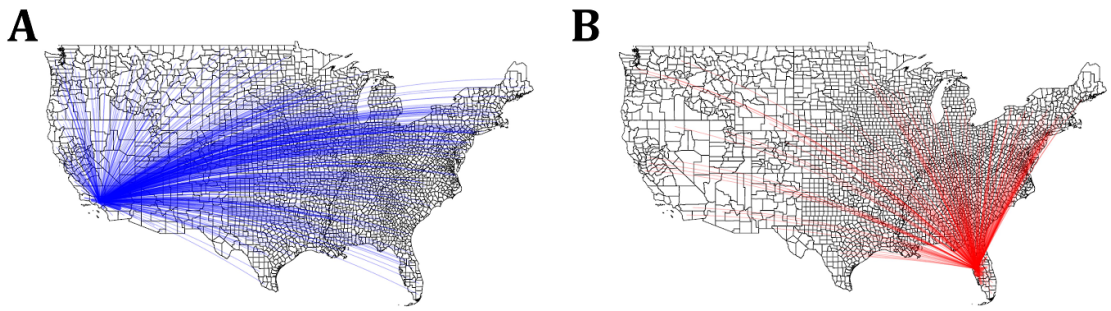

Figure 7. Example of the estimates provided by the food flow model. (A) Food inflows are shown for Los Angeles County, CA in 2017. (B) Food outflows are shown for Hillsborough County, FL in 2017.

Similar to FAF-level networks, decreases in mass flux from 2007 to 2012 are mainly concentrated in the Corn-Belt, while the increases are heterogeneously distributed around the nation in the county-level networks (see figures 8(C) and (D)). These findings illustrate that the 2012 drought is captured in our flow estimates.

4.3. Future research

There are many potential future research applications of this study. The spatially detailed maps of food flows that we have generated in this study (and make freely available with the paper) could be paired with footprint estimates to quantify embodied resource (e.g. water, carbon, etc) in future work.

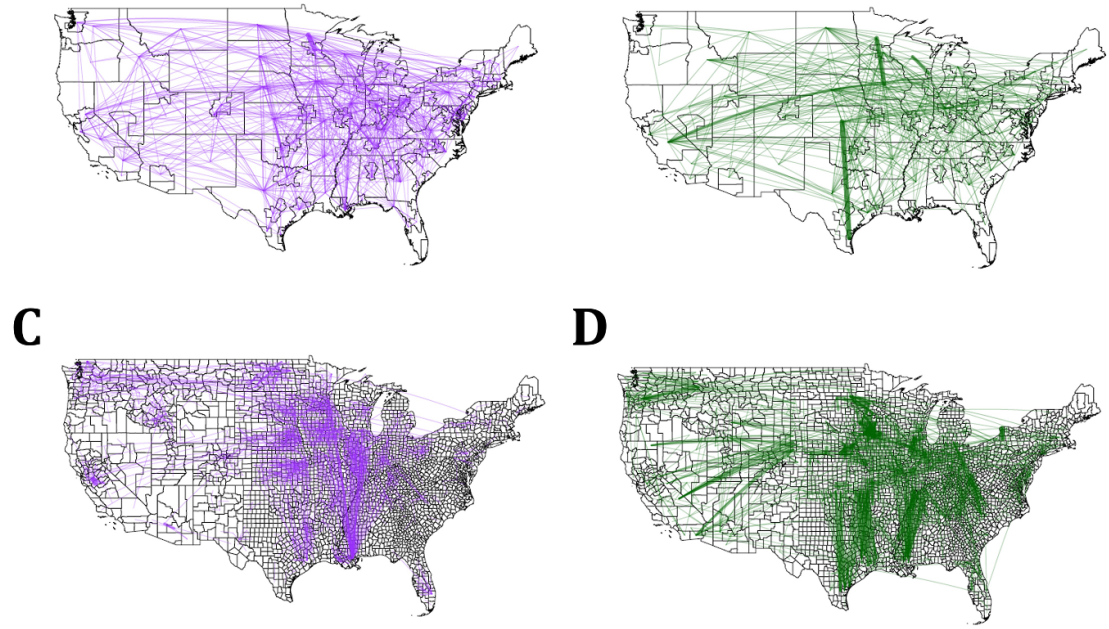

Figure 8. Grain flows (kg) during the 2012 drought. The top 2000 links by mass are shown in county-level. The top 1000 links by mass are shown in FAF-level. (A) Mass reductions from 2007 to 2012 in FAF-level. (B) Mass increases from 2007 to 2012 in FAF-level. (C) Mass reductions from 2007 to 2012 in county-level. (D) Mass increases from 2007 to 2012 in county-level.

Our work could also be used to develop understanding of county-level food security, dietary preferences, and agricultural sustainability. Additionally, more realistic transportation distance measures could be considered to develop mode-specific estimates of food flows. For example, the Food Flow Model could incorporate mode-specific travel times (Weiss *et al* 2018, Nelson *et al* 2019). This would move us closer to assessing the critical infrastructure that undergirds the production, processing, transport, and storage of agricultural and food flows within the US. Another important avenue of future research is to determine the vulnerabilities and resiliencies that exist in the national food supply network. These findings may be useful for researchers and decision makers interested in food systems security within the United States.

The Food Flow Model is a framework that could be applied to other locations. However, since it is a data-driven model, the necessary input data would need to be available in other locations. A major constraint that we foresee as being a likely impediment to implementing the Food Flow Model in other locations is the lack of sub-national commodity flow information. The US government collects information every 5-years in the US Census and uses that information to build the CFS (and subsequent FAF database), which is a key data requirement of the Food Flow Model described here. So, other countries or world regions with comparable coarse-scale commodity flow information to downscale from, could implement the Food Flow Model to estimate fine-grained food flows.

5. Conclusion

We estimated food flows between US counties through time with an improved version of the Food Flow Model. We provide 206.3 million data points with this paper (9 821 956 links including zeros for each of the 7 SCTG agri-food commodities and 3 study years). Our estimates present good general agreements with FAF data (by design), and capture a self-arising power-law relationship among node degree and betweenness centrality. The 2012 Corn-Belt drought is also evident in our estimates and our core counties are mainly consistent through time. The core counties represent some of the major transit hubs, such as Houston, TX, Chicago, IL, and Los Angeles, CA. Thus, our time-series estimates of food flows between US counties contribute to a more comprehensive picture of our national food system for researchers and policymakers.

Data availability statement

The data that support the findings of this study will be openly available at the following URL/DOI: https://doi.org/10.13012/B2IDB-9585947_V1. Data will be available from 7 February 2022.

Acknowledgments

This material is based upon work supported by the National Science Foundation Grant No. ACI-1639529 ('INFEWS/T1: Mesoscale Data Fusion to Map and Model the U.S. Food, Energy, and Water (FEW) System'), CBET-1844773 ('CAREER: A National Strategy for a Resilient Food Supply Chain'), DEB-1924309 ('CNH2-L: Feedbacks between Urban Food Security and Rural Agricultural Systems'), and CBET-2115405 ('SRS RN: Multiscale RECIPES (Resilient, Equitable, and Circular Innovations with Partnership and Education Synergies) for Sustainable Food Systems'). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

ORCID iD

Megan Konar https://orcid.org/0000-0003-0540-8438

References

Anderson J E 1979 A theoretical foundation for the gravity equation *Am. Econ. Rev.* **69** 106–16

Attavanich W, McCarl B A, Ahmedov Z, Fuller S W and Vedenov D V 2013 Effects of climate change on US grain transport *Nat. Clim. Change* **3** 638–43

Baiardi D, Bianchi C and Lorenzini E 2015 Food competition in world markets: some evidence from a panel data analysis of top exporting countries *J. Agric. Econ.* **66** 358–91

Barabási A-L 2016 *Network Science* (Cambridge: Cambridge University Press)

BEA 2020a *US Bureau of Economic Analysis, Personal Income* (available at: www.bea.gov/data/gdp/gdp-county-metro-and-other-areas)

BEA 2020b *US Bureau of Economic Analysis, Input–Output Accounts Table* (available at: www.bea.gov/industry/input-output-accounts-data)

Boyer J *et al* 2013 The US drought of 2012 in perspective: a call to action *Glob. Food Secur.* **2** 139–43

Chen A, Yang H, Lo H K and Tang W H 1999 A capacity related reliability for transportation networks *J. Adv. Transp.* **33** 183–200

Dall'Erba S, Chen Z and Nava N J 2021 U.S. interstate trade will mitigate the negative impact of climate change on crop profit *Am. J. Agric. Econ.* **103** 1720–41

Dang Q, Lin X and Konar M 2015 Agricultural virtual water flows within the United States *Water Resour. Res.* **51** 973–86

De Benedictis L and Tajoli L 2011 The world trade network *World Econ.* **34** 1417–54

Disdier A-C and Head K 2008 The puzzling persistence of the distance effect on bilateral trade *Rev. Econ. Stat.* **90** 37–48

Du Y, Gao C, Hu Y, Mahadevan S and Deng Y 2014 A new method of identifying influential nodes in complex networks based on TOPSIS *Physica* A **399** 57–69

Ercsey-Ravasz M, Toroczkai Z, Lakner Z and Baranyi J 2012 Complexity of the international agro-food trade network and its impact on food safety *PLoS One* **7** e37810

Gaur V, Soni G, Yadav O P and Rathore A 2020 Comparison between centrality measures for a network based on cascading nature of nodes *Proc. 6th Int. Conf. Recent Trends in Computing: ICRTC 2020* (Springer Nature) p 181

Ghosh D and Vogt A 2012 Outliers: an evaluation of methodologies *Joint Statistical Meetings* (available at: www.asasrms.org/Proceedings/y2012/Files/304068_72402.pdf)

Gomez M, Mejia A, Ruddell B L and Rushforth R R 2021 Supply chain diversity buffers cities against food shocks *Nature* **595** 250–4

Heslin A *et al* 2020 Simulating the cascading effects of an extreme agricultural production shock: global implications of a contemporary us dust bowl event *Front. Sustain. Food Syst.* **4** 26

Hu J, Du Y, Mo H, Wei D and Deng Y 2016 A modified weighted topsis to identify influential nodes in complex networks *Physica* A **444** 73–85

Hwang C-L and Yoon K 1981 Methods for multiple attribute decision making *Multiple Attribute Decision Making* (Berlin: Springer) pp 58–191

Hwang R-C, Siao J-S, Chung H and Chu C 2011 Assessing bankruptcy prediction models via information content of technical inefficiency *J. Product. Anal.* **36** 263–73

Karakoc D B, Barker K, Zobel C W and Almoghathawi Y 2020 Social vulnerability and equity perspectives on interdependent infrastructure network component importance *Sustain. Cities Soc.* **57** 102072

Karakoc D B and Konar M 2021 A complex network framework for the efficiency and resilience trade-off in global food trade *Environ. Res. Lett.* **16** 105003

KDA 2020 *Kansas Department of Agriculture* (available at: https://agriculture.ks.gov/economic-development-statistics/county-ag-statistics-(j-z)/sedgwick-county)

Klein M 1967 A primal method for minimal cost flows with applications to the assignment and transportation problems *Manage. Sci.* **14** 205–20

Konar M, Lin X, Ruddell B and Sivapalan M 2018 Scaling properties of food flow networks *PLoS One* **13** e0199498

Lin X, Dang Q and Konar M 2014 A network analysis of food flows within the United States of America *Environ. Sci. Technol.* **48** 5439–47

Lin X, Ruess P J, Marston L and Konar M 2019 Food flows between counties in the United States *Environ. Res. Lett.* **14** 084011

MacDonald G K, Brauman K A, Sun S, Carlson K M, Cassidy E S, Gerber J S and West P C 2015 Rethinking agricultural trade relationships in an era of globalization *BioScience* **65** 275–89

Metson G S, MacDonald G K, Leach A M, Compton J E, Harrison J A and Galloway J N 2020 The US consumer phosphorus footprint: where do nitrogen and phosphorus diverge? *Environ. Res. Lett.* **15** 105022

NDA 2020 *Nebraska Department of Agriculture* (available at: https://nda.nebraska.gov/)

Nelson A, Weiss D J, van Etten J, Cattaneo A, McMenomy T S and Koo J 2019 A suite of global accessibility indicators *Sci. Data* **6** 1–9

NOL 2020 *The Port of New Orleans* (available at: www.portnola.com/business/cargo)

Oak 2020a *Oak Ridge National Laboratory County-to-County Distance Matrix* (available at: https://tedb.ornl.gov/)

Oak 2020b *Oak Ridge National Laboratory Freight Analysis Framework* (available at: https://faf.ornl.gov/faf5/)

Orth W 2013 Multi-period credit default prediction with time-varying covariates *J. Empir. Finance* **21** 214–22

POR 2020 *US Bureau of Transportation Statistics Port Trade Data* (available at: www.bts.gov/transborder)

Porkka M, Kummu M, Siebert S and Varis O 2013 From food insufficiency towards trade dependency: a historical analysis of global food availability *PLoS One* **8** e82714

Prokopy L S, Carlton J S, Haigh T, Lemos M C, Mase A S and Widhalm M 2017 Useful to usable: developing usable climate science for agriculture *Clim. Risk Manage.* **15** 1–7

Robinson C, Shirazi A, Liu M and Dilkina B 2016 Network optimization of food flows in the US *2016 IEEE Int. Conf. Big Data (Big Data)* (IEEE) pp 2190–8

Schrijver A 2002 On the history of the transportation and maximum flow problems *Math. Program.* **91** 437–45

Shepherd B 2013 *Gravity Model of International Trade: A User Guide* (ARTNeT, United Nations ESCAP) (https://hdl.handle.net/20.500.12870/128)

Smith T M, Goodkind A L, Kim T, Pelton R E, Suh K and Schmitt J 2017 Subnational mobility and consumption-based environmental accounting of US corn in animal protein and ethanol supply chains *Proc. Natl Acad. Sci.* **114** E7891–9

Suanin W 2021 Demand elasticity of processed food exports from developing countries: a panel analysis of US imports *J. Agric. Econ.* **72** 413–29

USC 2020a *United States Census Bureau, Population* (available at: www.census.gov/data/datasets/time-series/demo/popest/2010s-counties-total.html)

USC 2020b *United States Census Bureau Land Area of Counties* (available at: www.census.gov/library/publications/2011/compendia/usa-counties-2011.html)

USC 2020c *United States Census Bureau, North American Industry Classification System* (available at: www.census.gov/naics/)

USD 2020a *United States Department of Agriculture* (available at: https://quickstats.nass.usda.gov/)

USD 2020b *United States Department of Agriculture Economic Research Service* (available at: www.ers.usda.gov/data-products.aspx)

Wan J, Qu M, Hao X, Motha R and Qu J J 2015 Assessing the impact of year 2012 drought in corn yield in the US corn belt using precipitation data *J. Earth Sci. Eng.* **5** 333–7

Weber C L and Matthews H S 2008 Food-miles and the relative climate impacts of food choices in the United States *Environ. Sci. Technol.* **42** 3508–13

Weiss D J *et al* 2018 A global map of travel time to cities to assess inequalities in accessibility in 2015 *Nature* **553** 333–6

White D R and Borgatti S P 1994 Betweenness centrality measures for directed graphs *Social networks* **16.4** 335–46

Wu D, Qu J J and Hao X 2015 Agricultural drought monitoring using modis-based drought indices over the USA corn belt *Int. J. Remote Sens.* **36** 5403–25

Xu M, Allenby B R and Crittenden J C 2011 Interconnectedness and resilience of the US economy *Adv. Complex Syst.* **14** 649–72