Text mining the food security literature reveals substantial spatial bias and thematic broadening over time

Matthew W. Cooper  
*University of Maryland*

Molly E. Brown  
*University of Maryland*

Meredith T. Niles  
*University of Vermont*

Moataz M. ElQadi  
*Monash University*

Follow this and additional works at: [https://scholarworks.uvm.edu/calsfac](https://scholarworks.uvm.edu/calsfac)

Part of the Human Ecology Commons, and the Medicine and Health Commons

**Recommended Citation**

Cooper MW, Brown ME, Niles MT, ElQadi MM. Text mining the food security literature reveals substantial spatial bias and thematic broadening over time. Global Food Security. 2020 Sep 1;26:100392.

This Article is brought to you for free and open access by the College of Agriculture and Life Sciences at ScholarWorks @ UVM. It has been accepted for inclusion in College of Agriculture and Life Sciences Faculty Publications by an authorized administrator of ScholarWorks @ UVM. For more information, please contact donna.omalley@uvm.edu.
Text mining the food security literature reveals substantial spatial bias and thematic broadening over time

Matthew W. Cooper\textsuperscript{a,b,}\textsuperscript{*}, Molly E. Brown\textsuperscript{a}, Meredith T. Niles\textsuperscript{c}, Moataz M. ElQadi\textsuperscript{d}

\textsuperscript{a} Department of Geographical Sciences, University of Maryland, College Park, MD, 20742, USA
\textsuperscript{b} T.H. Chan School of Public Health, Harvard University, Boston, MA, 02115, USA
\textsuperscript{c} Department of Nutrition and Food Sciences \& Food Systems Program, University of Vermont, Burlington, VT, 05405, USA
\textsuperscript{d} Faculty of Information Technology, Monash University, Clayton, 3800, Australia

\begin{abstract}
We conducted text mining analyses on nearly the entirety of academic literature related to food security. Assessing the literature’s spatial scope, we found a truly global body of research conducted across 187 different countries, but with significant spatial heterogeneities in where research is conducted. Comparing the spatial distribution of the literature to actual rates of food insecurity, we found only a slight association between where food security research is conducted and where food security needs are located. Using topic modeling to assess the thematic scope of the literature, we found that originally food security research focused on economic policy and global issues, and only later did the literature expand to encompass themes like livelihoods, health, and the environment. This analysis provides the first ever thematic scoping of the entire food security literature and the first assessment of spatial biases in where food security research is conducted.
\end{abstract}

1. Introduction

The literature related to food security is necessarily broad. The science of ensuring that “all people, at all times, have physical and economic access to sufficient safe and nutritious food” (FAO, 1996) is highly interdisciplinary and requires collaboration across fields as diverse as agriculture, land use, nutrition, economics, genetics, physiology, hydrology, sociology, public policy, and more. These disciplines are all represented in the food security literature, attesting to a massive multi-decadal collaboration across almost the entirety of academia to study how we meet the basic human need for food. However, the volume and diversity of the food security literature has meant that little work has been done to survey this body of research at comprehensive scales. This presents a challenge for identifying research gaps in such a wide-ranging, interdisciplinary literature, as well as assessing issues like regional heterogeneities in research focus, tracking changing meanings in keywords or identifying emergent themes over time. Thus, we used text-mining approaches to survey the geographic and thematic scope of the entire food security literature.

The term “food security” has had different meanings at different times and in different contexts (Pinstrup-Andersen, 2009). One of the most agreed-upon definitions comes from the 1996 World Food Summit, which defined food security as “when all people, at all times, have physical and economic access to sufficient safe and nutritious food to meet their dietary needs and food preferences for a healthy and active life” (FAO, 1996). This broad definition of food security has allowed a variety of academic disciplines, themes, issues and contexts to come together under the scope of the term.

Given the dynamism and complexity of the food security literature, higher-level overviews and syntheses are a valuable resource for researchers and policymakers working across food systems to stay aware of trends and issues in the literature (Rosegrant and Cline, 2003). While large-scale scientific collaborations and research agendas exist to provide such overviews (Ceres2030, 2020; Haddad et al., 2016), a text mining approach can explore large volumes of literature very efficiently and potentially complement these efforts. Thus, to provide a novel exploration of the food security literature, we assembled an exhaustive corpus of 16,152 abstracts of cited academic articles that contain the terms Food Security or Food Insecurity to assess both the geographic and thematic focus of the food security literature. Text mining approaches are common in other disciplines (Aranda et al., 2019; Movaghar et al., 2019; Simmons et al., 2016; Singhal et al., 2016), but have yet to be conducted on a large corpus of academic literature related to food security, although some studies have created novel datasets by summarizing text data from briefs and reports Pardey et al. (2013); Samimi et al. (2012); Surjandari et al. (2014). Similarly, assessing for statistical

\textsuperscript{*} Corresponding author. T.H. Chan School of Public Health, Harvard University, Boston, MA, 02115, USA.
E-mail address: mattcoop@terpmail.umd.edu (M.W. Cooper).

\url{https://doi.org/10.1016/j.gfs.2020.100392}

Received 7 January 2020; Received in revised form 29 April 2020; Accepted 26 May 2020
Available online 03 July 2020
2211-9124/ © 2020 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
biases in the spatial distribution of research is a common practice in other disciplines, especially where phenomena differ significantly across geographies, such as ecology (Roberts et al., 2016; Tittley et al., 2017; Trimble and van Aarde, 2012; Kim et al., 2016). However, to our knowledge, this is the first paper to test for statistical biases in the spatial distribution of the food security literature.

In this paper, we begin with an introduction to our methods and a description of our results. We then discuss the implications of our results, comparing the geographic distribution of food security studies to the distribution of actual food insecurity, as well as assessing how the themes represented in the literature compare to “official” definitions of food security and various food security research agendas that have been put forth. We conclude with a discussion of the implications of various findings from this study.

2. Methods

2.1. Collecting abstracts

We used Scopus, “the largest abstract and citation database of peer-reviewed literature,” (Elsevier, 2019) to collect abstracts for every publication that used the terms Food Security or Food Insecurity in the title, abstract or keywords. This approach of querying a database by a topic or phrase is a common approach in studies that use text mining to study a body of scientific literature (Liu and Liao, 2017; Cheng et al., 2018; Martí-Parreño et al., 2016; Guerreiro et al., 2016).

The initial search of the Scopus database yielded 26,085 abstracts. To ensure that we included only abstracts that were broadly comparable and representative of the literature, we removed those that had under 500 characters or that had no citations at the time the data was collected (Fall 2018). Furthermore, we kept only abstracts that were from documents that were either original research articles, opinion articles, or review articles, excluding other document types such as book chapters, errata, editorials, conference abstracts, or conference papers. Finally, in some cases, our data set had near-duplicate abstracts. This usually happened when both a working paper and a final abstract were in our data set. Thus, we removed one abstract from any pair of abstracts that were sufficiently similar according to the Jaccard similarity between the sets of abstract 3-g, a common metric of text similarity (Gomaa and Fahmy, 2013).

2.2. Geographic classification

To determine both where the food security literature tends to focus as well as how the topics that the literature focuses on vary across different geographies, we extracted and geolocated all place-names, or toponyms, in each abstract, as well as in the keywords and title associated with each abstract. To identify and geolocate toponyms in the abstracts, we use cloud-based text analysis services, which are increasingly utilized by academics (Drake, 2015). Specifically, we use two services provided by Application Program Interfaces (APIs) on the Google Cloud Platform: entity extraction with the Natural Language API and geolocation with the Maps API. The entity extraction service takes a body of text and labels words and phrases with categories such as person, event or location (Google Cloud Platform, 2020a). Within each abstract, for every location identified by the entity extraction service, we then used the geolocation service. The geolocation service takes a toponym and returns metadata such as a latitude and longitude, as well as an associated country and world region, where applicable (Google Cloud Platform, 2020b). We interacted with the APIs using the python package “google-cloud” (Google Cloud Platform, 2020c), and in between API calls to the google cloud platform, data was stored in json format. All of our scripts used to collect this data are publicly available at github.com/mcooper/mine-food-security. For a schematic figure of the workflow for assigning each abstract to a country and world region, see Fig. 1.

To determine the country associated with an abstract, we assessed the countries in which each toponym mentioned in the abstract was located. In cases where a majority of toponyms were from one country, the abstract was classified as being from that country (60.3% of abstracts). To conduct a regional analysis, we further classified each abstract as being in one of four broad continental categories: Africa, Asia-Pacific, Latin America and the Caribbean (LAC), and High-Income Countries (HICs) including Europe, the USA, Canada, Australia, and New Zealand. About 12.2% of abstracts were regional, where a majority of toponyms were not from one specific country but were from one world region. Abstracts were not included in spatial analyses in cases where abstracts had either no toponyms or no clear geographic focus (27.5% of abstracts).

After determining the country and region of focus for relevant abstracts, we validated our classification using a random sample of 300 abstracts from our data set. Because we only conducted our analyses on abstracts that were classified as being associated with a specific country or region and not on abstracts with no geographic focus, we similarly conducted our validation on those abstracts. Based on that subset of the manually classified data, 92.5% of the abstracts were associated with the correct world region and 93.2% of the abstracts were associated with the correct country.

Based on the number of abstracts associated with each country, we determined the number of food security abstracts published per million people in each country per year. We then assessed for statistical bias in the focus of the literature over time using a Moran’s I statistical test, a common test for spatial autocorrelation. Finally, we compared the number of abstracts per million people to each country’s score in the Proteus Index, developed by the World Food Programme (Caccavale and Giuffrida, 2020). This novel index measures national food security based on a variety of indicators, and takes into account uncertainty and sensitivity at all steps of index construction. We used this comparison to evaluate how the geographic focus of the food security literature compares with the global distribution of actual food security.

2.3. Topic modeling

Our primary method for exploring the thematic scope of the food security literature was to use a topic modeling algorithm to identify the various topics in the literature as well as their interrelationships. This involves algorithmically identifying clusters of words that co-occur together in documents and analyzing those clusters as “topics” (Blei, 2012). Specifically, we used a Correlated Topic Model, or CTM (Blei and Lafferty, 2007). CTMs are most appropriate for text datasets with substantial similarity and correlation between topics. For more details on CTMs as well as our method for selecting the number of topics, see the Supplementary Materials Section 6.1 and Table S1.

Based on the topics identified by the unsupervised classification performed by the CTM, the authors manually created a topic label and grouped the topics into broader themes in the food security literature using their expertise, as is common in food security assessments (Terpend, 2006; Krishnamurthy et al., 2014). For example, in Table 2, the columns “Top Three Words” and “Representative Article” were identified by the model, while the columns “Topic Label” and “Theme” were manually designated. Thus, in assessing the thematic scope of the food security literature, this paper has two units of analysis: the topics that were identified by the model in an unsupervised classification, and the broader themes that those topics were then manually grouped into.

2.4. Distribution of themes across world regions

Finally, having examined both the themes and geographies represented in each abstract, we examined how these themes vary by geography. Based on the broader themes from the model and the four world regions that abstracts were associated with, we tabulated the number of abstracts in each theme by world region. We then used a log-
linear analysis to determine whether different themes are more likely to occur in each of the four world regions. Because each abstract can contain multiple themes and a log-linear analysis requires count data, we summed the total number of abstracts in each theme by world region according to the fraction of each abstract representing each theme, and then rounded the count of abstracts to the nearest whole number. To determine whether the proportions of each theme vary independently across world regions, we made two log-linear models: one first-order model with no interaction between themes and world regions, and a 1-order model with an interaction term. We then used a chi-square test on the residual deviance to determine the goodness-of-fit of the models Agresti (2006). If the model without an interaction term is well fit, then themes vary independently of world regions. However, if the model without an interaction term is poorly fit and an interaction term is necessary to describe the tabulation, then certain themes are biased towards certain world regions.

3. Results

Our search of the literature yielded a final data set, or corpus, of 16,152 abstracts from 3297 journals from the years 1975–2018. The top ten journals represented in our corpus are given in Table 1.

3.1. Geographic results

We found that 60.3% of the abstracts in the corpus had the majority of their toponyms from one country, with 187 different countries having been the focus of a study related to food security. Of these countries, the most represented was the United States (n = 1470; 15.1% of abstracts associated with a country), followed by China (n = 1123; 11.5%) and India (n = 652; 6.7%). Countries with a high percent of abstracts per capita per year included a number of micro-states and island nations with low populations, such as Vanuatu, Greenland and the Solomon Islands. Larger countries with a high percent of abstracts per capita per year include Botswana, Swaziland, Belize, Timor-Leste, Malawi, and Canada. Finally, countries with relatively little per capita research include Iceland, Belarus, Armenia, Angola, and Suriname (see Fig. 2).

Using a Moran’s I test for spatial autocorrelation in publications per capita per year shows that there is a strong degree of spatial

![Fig. 1. Schematic representation of workflow for assigning abstracts to a country and world region, based on an example abstract highlighted in green. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)](image-url)
We further compared the number of food security publications per capita to the Proteus Index for the years 2013–2018 (See Fig. 3 and Fig. S2), and we found substantial variation in levels of research. High-income and food-secure parts of the world had many examples of both well-researched countries and relatively under-researched countries. Similarly, in Africa, where countries are generally less food secure, there were examples of well-researched countries, particularly in eastern, southern, and western Africa, while many central African countries had relatively less research per capita. In Asia, much of the former USSR was relatively less researched but also amply food secure, while in South Asia, countries were less food secure and less researched according to this per-capita metric due to high population levels. For this time period, we found a correlation of 0.111 between a country’s average food security status, as measured by the Proteus Index, and the number of publications per capita per year.

### 3.2. Topic model results

We used a Correlated Topic Model (Blei and Lafferty, 2007) to assess the thematic scope of the literature, and we manually grouped these topics into 9 broader themes present in the literature. Table 2 shows a summary of the topics identified by the model, including the label we assigned to each topic, the broader theme we grouped each of these topics into 9 broader themes present in the literature. Table 2 shows a summary of the topics identified by the model, including the label we assigned to each topic, the broader theme we grouped each topic into, the top three words associated with each topic, and the publication most associated with each topic that had at least 10 citations.

| %  | Theme                  | Topic Label | Top Three Words                        | Representative Article                      |
|----|------------------------|-------------|----------------------------------------|--------------------------------------------|
| 4.70 | Agronomy               | Yield Gap   | Agriculture Product Develop            | Lu et al. (2015)                           |
| 3.60 | Agronomy               | Crop Diseases | Plant Gene Resist         | Nicaise (2014)                             |
| 3.00 | Agronomy               | Crop Monitoring Data | Model Use Data        | Lambert et al. (2016)                      |
| 2.90 | Agronomy               | Crop Variety Research | Crop Yield Product | Grasini et al. (2011)                      |
| 2.40 | Agronomy               | Crop Genetics | Genet Crop Breed                  | Hirsch et al. (2014)                       |
| 1.90 | Agronomy               | Soil Fertility | Soil Fertilize Plant          | Singh et al. (2007)                        |
| 1.00 | Agronomy               | Fertilizer   | Oil Use Phosphorus             | Ledezma et al. (2015)                      |
| 0.70 | Agronomy               | Insect Pests | Pest Insect Loss               | Kedia et al. (2013)                        |
| 0.70 | Agronomy               | Rice         | Rice Paddy Rice-product        | Thomas and Ramai (2010)                    |
| 0.30 | Agronomy               | Wheat China  | Wheat China Winter             | Qin et al. (2015)                          |
| 0.20 | Agronomy               | Tuber Production | Cassava Potato Sweet       | Samatay et al. (2014)                      |
| 0.10 | Agronomy               | Ag Intensification | Arabi Land Arabi-land          | Wenwa (2012)                              |
| 0.10 | Agronomy               | Crop Production Technology | Ecolog Factor Region | McColl et al. (2017)                      |
| 0.10 | Agronomy               | Biofuel Research | Emiss Energi Biofuel     | Cornelissen et al. (2012)                 |
| 0.10 | Agronomy               | Precipitation Trends | Season Drought Rainfal | Awang et al. (2007)                       |
| 0.10 | Agronomy               | Climate Hazards | Risk Vulner Resili            | Hillbruner and Molney (2012)               |
| 0.10 | Agronomy               | Bio Energy | Bioenergi Energi Gas            | Acosta et al. (2013)                       |
| 0.30 | Economic Policy        | Food Prices  | Price Country Polici         | Dorward (2011)                             |
| 0.10 | Economic Policy        | Food Trade   | Trade Agricultur Product      | Götz et al. (2013)                         |
| 0.10 | Economic Policy        | Food Value Chain | Market Seed Farmer          | Louwaars and de Boel (2012)               |
| 0.10 | Economic Policy        | Food Waste   | Waste Chain Suppli            | Nahman et al. (2012)                       |
| 0.10 | Economic Policy        | Food Sovereignty | Polici Social Politi       | Jarroz (2014)                              |
| 0.10 | Economic Policy        | Agricultural Land Use | Land Cultiv Area             | Song and Pijanowski (2014)                |
| 0.10 | Economic Policy        | Food Security Measurement | Use Measur Indic         | Marques et al. (2014)                      |
| 0.10 | Economic Policy        | Food Access  | Urban Citi Garden             | Raja et al. (2008)                         |
| 0.10 | Economic Policy        | Food Security | Africa African South       | van Zyl and Coetzee (1990)                 |
| 0.10 | Economic Policy        | Food Aid     | Aid Bank Servic              | Tarzuk et al. (2014)                       |
| 0.10 | Health                 | Health Determinants And Equity | Health Social Care       | Hassan et al. (2013)                       |
| 0.10 | Health                 | Food Contaminants | Concentr Sampol Contamin | Amiard et al. (2008)                      |
| 0.10 | Health                 | HIV          | HIV Adher Intervent           | Cantrell et al. (2008)                     |
| 0.10 | Health                 | Community Health | Communiti Health Social | Power (2005)                              |
| 0.10 | Health                 | Health And Poverty | Food Develop Communiti | Siemieniuch et al. (2015)                 |
| 0.10 | Livelihoods             | Household Modeling | Household incom Rural | Owusu et al. (2011)                       |
| 0.10 | Livelihoods             | Income Diversification Farmer | Farmer Farm Use           | Ogundari (2014)                           |
| 0.10 | Livelihoods             | Traditional Knowledge | Communiti Interview Studi | Lori et al. (2013)                        |
| 0.10 | Livelihoods             | Food Stamps US | Program Particip Assist       | Mykersi and Mills (2010)                   |
| 0.10 | Livelihoods             | Gender       | Women Gender Men             | Buetel (2005)                              |
| 0.10 | Livelihoods             | Animal-Human Disease | Health Diseas Human       | Jenkins et al. (2015)                      |
| 0.10 | Livelihoods             | Livestock    | Product Livestock Anim        | Schader et al. (2015)                      |
| 0.10 | Livelihoods             | Poultry      | Product Poutrui Vaccin       | G’ueye (2002)                             |
| 0.10 | Livelihoods             | Livestock Disease | Coast West Popul           | Hernández-Pacheco et al.(2013)            |
| 0.10 | Livelihoods             | Under And Overweight | Assosi Household Among | Gullford et al. (2006)                    |
| 0.10 | Livelihoods             | Children And Infants | Children Child Nutrit | Trehan and Manary (2014)                  |
| 0.10 | Livelihoods             | Food Nutrition | Veget Fruit Consumpt         | Appiah et al. (2011)                      |
| 0.10 | Livelihoods             | Child Malnutrition | Nutri Nation Develop | Wuhrler and Wane (2011)                   |
| 0.10 | Livelihoods             | Dietary Diversity | Dietari Intak Diet          | Zizza et al. (2008)                       |
| 0.10 | Livelihoods             | School Meals  | School Meal Cost             | Gelli et al. (2009)                        |
| 0.10 | Livelihoods             | Micronutrients | Micronutri Knowledg Plant    | Fragoso et al. (2016)                      |
| 0.10 | Livelihoods             | Irrigation   | Water Irrig Resourc          | Mu et al. (2009)                           |
| 0.10 | Livelihoods             | Water         | Fish Fisheri Aquacultur      | Cisneros-Montemayor et al. (2013)          |
| 0.10 | Water                   | Food And Water | Asia South South-Asia        | Razul (2010)                              |
| 0.10 | Water                   | Dams And Displacement | Dam Displac Right       | Berhans and White (2000)                   |
One advantage of Correlated Topic Models over other topic modeling methods is that they allow the analyst to examine the relationships among topics and derive a graph of topic connectivity (Blei and Lafferty, 2007). Using this approach, we find that most of the larger topics identified in the data set are related to other topics, especially topics that are part of the same food security pillar, although there are some smaller topics that are not related to any other topic (See Fig. 4). The topics with the greatest degree, an indicator of connectedness, are “Food Aid” (14), “Fertilizer” (11), and “Crop Variety Research” (7), while the topics with the greatest weighted degree, which accounts for the strength of the connections between topics (Barrat et al., 2004), are “Biofuel Research” (3.2), “Food Aid” (3.15), and “Climate Adaptation” (3.05).

While the literature is broadly connected, some distinct topical clusters emerge, especially in light of the strength of the relationships between topics, as shown by the thickness of the edges in Fig. 4. These topical clusters correspond somewhat around the high-degree, central nodes. One cluster of tightly correlated topics is related to biophysical aspects of food production, including the topics “Soil Fertility,” “Irrigation,” “Agricultural Land Use,” “Natural Systems,” “Climate Adaptation,” “Biofuel Research,” and “Fertilizer,” most of which correspond to the pillars Availability and Stability. Another cluster is related to more human aspects of food security, including the topics “Food Aid,” “Gender,” “Child Malnutrition,” “Food Stamps US,” “Under and Overweight,” and “Traditional Knowledge,” most of which correspond to the pillars Access and Utilization.

3.3. Time series analysis

In Fig. 5, we show the share of the literature according to continent and according to topic theme over time. Compared to the present, the food security literature before the mid-1990s was much more focused on Africa and on themes related to economic and global issues. Before 1990, 69.3% of the literature was related to the themes Economic Policy or Global Food Security. Around 1990, livelihoods became a major share of the literature, and made up 19.5% of the food security literature in that decade. Many of the other topics in the current food security literature are a more recent focus. Currently, the largest share of the literature is focused on the Livelihoods theme (14.9%), followed by Global Food Security (14.6%) and Climate & Sustainability (13.1%).

![Fig. 2. Number of abstracts located in each country per million inhabitants per year, calculated over various time periods and over the entire study period. Values are indicated with a logarithmic scale.](image)

![Fig. 3. Comparison of food security abstracts per capita per year with each countries Proteus Index, for the years 2013–2018. Countries are color-coded by quantile for both variables. Pink colors indicate countries that are less food secure and less researched while green colors indicate countries that are more food secure and more researched. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)](image)
Fig. 4. Topic graph of identified topics. The size of the node for each topic corresponds to that topic’s relative share of the literature, while the size of the connections between topics corresponds to the strength of their connection, as estimated by the tuning parameter $\lambda_1$. Topics with no connections to other topics are shown in the bottom-right corner.

Fig. 5. A) Proportion of food security literature focused on different world regions over time, with a 5-year smoothing. The regions are Africa, Asia, High-Income Countries (HICs) and Latin America and the Caribbean (LAC). Abstracts not associated with a specific region are excluded. B) Proportion of food security literature related to different themes over time, with a 5-year smoothing. C) The number of publications each year. Note: there are fewer publications for recent years because those with no citations were filtered from the data set.
Table 3
Tabulation of abstracts by theme and world region.

| Theme                  | Africa | Asia | HICs | LAC | Total |
|------------------------|--------|------|------|-----|-------|
| Agronomy               | 865 (8%) | 865 (8%) | 543 (5%) | 193 (1.8%) | 2466 (22.7%) |
| Climate & Sustainability| 460 (4.2%) | 470 (4.3%) | 309 (2.8%) | 105 (1%) | 1344 (12.4%) |
| Economic Policy        | 283 (2.6%) | 236 (2.2%) | 174 (1.6%) | 61 (0.6%) | 754 (6.9%) |
| Global Food Security   | 600 (5.5%) | 537 (4.9%) | 425 (3.9%) | 136 (1.3%) | 1698 (15.6%) |
| Health                 | 267 (2.5%) | 229 (2.1%) | 228 (2.1%) | 72 (0.7%) | 797 (7.3%) |
| Livelihoods            | 492 (4.5%) | 374 (3.4%) | 394 (3.6%) | 110 (1%) | 1370 (12.6%) |
| Livestock              | 156 (1.4%) | 148 (1.4%) | 102 (0.9%) | 33 (0.3%) | 439 (4%) |
| Nutrition              | 476 (4.4%) | 391 (3.6%) | 466 (4.3%) | 137 (1.3%) | 1470 (13.5%) |
| Water                  | 171 (1.6%) | 181 (1.7%) | 115 (1.1%) | 59 (0.5%) | 517 (4.8%) |
| Total                  | 3770 (34.7%) | 3431 (31.6%) | 2757 (25.4%) | 897 (8.3%) | 10855 (100%) |

Geographically, since the early 21st century, focus has shifted to be more evenly distributed across the globe than in the early literature. Currently, Africa and high-income countries are over-represented in the literature, at 31.4% and 27.5% of the literature and 15.7% and 17.7% of the world’s population, respectively (UN, 2019); Asia is under-represented, at 34.1% of the literature and 58.1% of the world’s population; and Latin America and the Caribbean’s share of the literature, at 6.9%, is roughly proportional to its share of the world’s population, at 8.5%.

The volume of the food security literature has also increased dramatically in recent years, mirroring an exponential trend in publishing seen across all science (Bormann and Mutz, 2015). This means that there were relatively few publications before the 1990s, yet today there are thousands of publications a year.

3.4. Topics by world regions

After associating abstracts with their relevant world region and identifying the major themes present in the abstracts, we tested for patterns across world regions using a log-linear approach. Table 3 shows the tabulation of abstracts by theme and world region. The first-order model (See Table S3) had a residual deviance of 99.96 with 24 degrees of freedom, indicating a poor fit ($p < 0.0001$). This indicates that there is significant statistical bias among some themes towards certain geographies. A second-order model (See Table S4) indicates where this bias is found. The themes of Health and Nutrition are both significantly associated with High-Income Countries (HICs), Nutrition is significantly associated with Latin America and the Caribbean (LAC), while the themes of Nutrition and Livelihoods are significantly less likely to occur in Asia. However, while some themes are biased towards certain regions, many of the themes, including Climate & Sustainability, Economic Policy, Global Food Security, Livestock, and Water, were not significantly associated with any one world region.

4. Discussion

These results have several implications with respect to the geographic and thematic scope of the food security literature. Geographically, we find that, while the food security literature has global coverage, there is wide and significant variation in where research is conducted, with little concordance in recent years between where food security research takes places and where food insecure populations are actually located. Thematically, we found that topics cluster according to food security pillar, and most of the topics in the literature are related to other topics. Examining the literature over time, we find that originally, the term “food security” was used in the context of economic policy and global issues, with particular focus on Africa. Only later did the term expand to be applied to new themes and give more focus to other geographies. Finally, we find that there is a statistical association between some themes and geographies, although each theme in the literature is fairly well represented in all geographies.

4.1. Meaning of food security over time

Most researchers agree that the term food security was introduced in the 1974 World Food Conference and was mostly used in the context of concern around global food supplies and a booming global population (Maleta, 2014; Shaw, 2007). As the green revolution assuaged fears of global food shortages but famines continued to occur in many developing countries, the term came to signify national self-sufficiency in food production (Pinnstrup-Andersen, 2009). Nevertheless, most food security research was concerned with little more than national levels of crop production and trends in food prices (Upton et al., 2016; Brown et al., 2015). As it became increasingly apparent that hunger and famine can occur even when national crop production is high (Sen, 1983), the term came to cover many other aspects of food systems, including economic conditions from the global to the local level, the conditions of trade, the safety and nutrient content of the food available, disease and health of the individual, and sustainability, among many other factors (Niles and Brown, 2019; Erickson et al., 2009; Lang and Barling, 2012).

Our analysis largely confirms the accepted history of the term. Fig. 5 shows how, in the 1980s and before, most of the literature focused on economic policy and global issues of production and trade, with a particular geographic focus on Africa. While there was not a sudden shift or expansion in the usage of the term food security, gradually the meaning broadened to encompass new notions of health, nutrition and sustainability at national and local scales. Thus, the breadth and depth of food security research grew over time, potentially mirroring a shifting conception of the meaning of the term “food security” itself (7). Around 2005, the scope of the term stabilized, insofar as the topics and themes present in the literature then still define the scope of the literature today in the same proportion.

4.2. Spatial focus of food security research

Ensuring food security is a concern in every country on earth, and we found global body of literature with studies from 187 different countries. However, we found substantial statistical bias in the spatial distribution of research, as measured by the Moran’s I test. Anglophone countries in North America, East Africa, and Oceania receive a disproportionate amount of research, although some Anglophone countries, such as Ireland and Nigeria, have relatively little per-capita research. Other hotspots of research included West Africa and several landlocked countries in south Asia. These spatial biases in where research is conducted likely reflect the feasibility of conducting research. There was significantly more research in wealthy countries compared to countries with less conducive research environments due to low levels of development, high levels of corruption, or endemic conflict (Brown et al., 2020).

We would expect food security research to prioritize research in areas with higher rates of food insecurity. However, comparing where most research takes places to levels of food insecurity, as measured by...
the Proteus Index, reveals only a slight correlation between where food security research is conducted and where food security needs are. While all of the high-income countries are food secure, they vary greatly in their levels of food security research. Sub-Saharan African countries, on the other hand, are largely food insecure, but also have wide disparities in where research takes place.

The gaps that exist in the geographic scope of the literature are most meaningful when comparing between two similar countries at the regional level. This is because food security research funding and effort isn’t transferable between different countries and food systems: scientists and resources supporting food security research among Inuit communities in Canada, for example, cannot be re-diverted to Somalia. Moreover, our per-capita metric somewhat obscures decades of highly impactful food security work in heavily populated places like South Asia, and thus concluding that India is under-researched compared with sparsely-populated Greenland is incongruous.

Nevertheless, there are examples of stark gaps even among similar countries. Nicaragua and El Salvador both have 6.4 million people, yet there have been 6 times as many academic articles on food security focused on Nicaragua than El Salvador since 2013. Slovakia has received 4.5 times as much research as nearby Czechia in this same time period. Some of the gaps are most severe in Africa, where food security levels are most critical. In West Africa, the countries Mali, Niger, Burkina Faso, and Chad are all landlocked, arid, and francophone with a population between 14 and 20 million. Yet between 2013 and 2018, Mali and Niger have each received 20 times as much food security research per capita as Chad, and Burkina Faso has received nearly 30 times as much research.

While this paper shows a clear statistical bias in where food security research is conducted, whether or not this represents a problem for the research community is not something this analysis can address directly. This statistical bias is not attributable to any one researcher or group of researchers, and is the result of a confluence of factors such as where researchers come from, which locations they have access to, which areas are prioritized by funding sources, in which geographies data is available, as well as perceived food security need in different parts of the world. Moreover, some research in accessible countries with suitable research facilities will have spillover effects in similar countries. For example, research on livestock management in Nairobi, Kenya could improve food security in nearby under-researched countries like Somalia. Thus, the spatial bias in where research on food security is conducted does not necessarily mean that there is a bias in where the benefit from food security research is felt. However, this analysis does highlight some rather stark gaps, and we suggest that focusing research on under-studied contexts and food production systems should be a priority of the research community moving forward, especially in areas that are already quite food insecure.

4.3. Thematic scope and proposed research agendas

A number of research agencies and individual researchers have put forth various food security research agendas over the years. By identifying the various salient topics within the food security literature, we can proxy the extent to which these agendas have influenced the literature, as well as identify potential gaps. We find that, in many cases, research into topics that various agencies and individuals have called for are indeed represented in the current literature. Nevertheless, in some cases scholars have called for research that never came to make up a salient component of the literature.

A common theme in various food security agendas that have been put forth is the need for holistic, interdisciplinary and systems-level research (Horton et al., 2017; Sonnino et al., 2014). Many of these agendas emphasize applicability and policy relevance (Haddad et al., 2016), as well as the necessity to take into account environmental sustainability as well as economic externalities (Brinkley et al., 2013; Haddad et al., 2016). Most notably, as climate change worsens and the twin needs for carbon-neutral and climate-resilient food systems become more apparent, researchers are calling for more studies related to climate change (Steenwerth et al., 2014). To a large extent, the research community has risen to meet these stated research objectives. A variety of topics related to themes like systems, climate change, sustainability, and policy were identified by our model.

In some cases, researchers called for very specific research topics. In 2000, Pinstrup-Andersen interviewed a variety of experts about emerging food security issues in developing countries (Pinstrup-Andersen, 2000). Many of the topics identified by the surveyed experts show up quite clearly as topics identified by our model. For example, the experts called for research related to water and urban-rural issues, which are related to our “Water” theme as well as our “Food Access” and “Food Contaminants” topic. Similarly, Sonnino specifically called for more research into linkages between food chains and food waste (Sonnino et al., 2014), which is very much the type of research identified by our model’s “Food Waste” topic, which had the top three keywords of waste, chain, and supply.

While the model cannot demonstrate that a research domain is completely absent from the body of literature, if the model does not identify a topic related to a research theme that has been called for or would be expected, this is evidence that there is not a distinct set of keywords that correspond to this research domain. In only a few cases did researchers call for research into areas that were not really present in the topics identified by our model. For example, in 1998, Haddad el al. argued that the “human rights perspective” could shape the food and nutrition agenda (Haddad and Oshaug, 1998). While human rights are somewhat a part of the food security conversation, especially in the “Food Sovereignty” topic, human rights were not directly related to any topic identified by our model. Similarly, in the 2000 survey, Pinstrup-Andersen identified impacts of new technology and armed conflict as emerging areas in the food security literature (Pinstrup-Andersen, 2000). While “technology” and “conflict” were keywords for some topics, such as “Yield Gap”, “Income Diversification Farmer” and “Dams and Displacement”, these keywords were only a minor component of those topics, and the issues of technology and conflict were not identified by the model as stand-alone topics. Thus, theses issues that were once seen as an emerging part of the literature did not have a lasting enough impact to become a distinct sub-literature.

4.4. Connectivity of the literature

It is a common criticism that food security research and policy is “siloed” among the various subdomains (Fukuda-Parr and Orr, 2014; Obersteiner et al., 2016; Gallegos and Chilton, 2019; Candel, 2018). However, we find that, aside from niche technical topics, much of the food security literature is broadly unified and is not siloed into disconnected sub-disciplines. This connectivity is based on correlations between the words used in each topic and is largely facilitated through thematically “central” topics such as Food Aid, Fertilizer, Crop Variety Research, Biofuel Research and Climate Adaptation. While this analysis cannot show which sections of the literature are citing which other sections, it does show that, for every pair of connected topics in Fig. 4, those two topics are using similar words and are discussing the same issues, theories, methods, and frameworks that those words signify. Thus, for topics as unrelated as Soil Fertility and HIV, there exist a number of “connecting” topics based on shared vocabulary.

The few topics that were found to be uncorrelated with any other topic were often related to a very specific and technical aspect of food security, such as Livestock Disease or Micronutrients. This shows that the vocabulary used in those topics was unique to those topics and not frequently used in other domains of the food security research.

Finally, we labeled our topics according to the pillars of food security designated by the FAO: food availability, related to food production via farming and fishing; food access, related to the distribution of food via markets and government policies; food utilization, related to
the proper preparation and digestion of food; and food stability, related food security at all times, without seasonal gaps or vulnerability to shocks. We found that most topics map clearly onto a specific food security pillar, and that nearby topics were more likely to be a part of the same food security pillar. We further found two broad clusters in the derived graph, one related to more biophysical topics and the pillars availability and stability (left side of Fig. 4), and one related to more social science topics and the pillars access and utilization (right side of Fig. 4). This indicates that the FAO pillars are a valid framework for characterizing the food security literature.

4.5. Limitations

This data set was collected from the Scopus database, and thus literature related to food security that is missing from Scopus will be absent from our analysis. While Scopus is widely considered to be “the largest abstract and citation database of peer-reviewed literature” (Elsevier, 2017), it is possible that some important abstracts are missing from our analysis, especially in the early literature, and further possible that the text of an abstract is not representative of the entire text of a publication. Second, we only searched for the English words Food Security and Food Insecurity, meaning that relevant abstracts in other languages are not part of our analysis. This probably explains some of the spatial bias in the literature towards Anglophone countries, and may obscure thriving regional literatures in other languages in places like Latin America. Nevertheless, English is the most commonly used academic language and therefore our dataset of abstracts is reflective of the food security literature that is available to the global research community. Finally, our analysis is based on abstracts that mention food security, and our results must be interpreted in this light. Thus, research relevant to food production, markets and trade, or human nutrition that does not use the terms food security or food insecurity is absent from our analysis.

It should also be noted that not all of our results from the Correlated Topic Model are based on word frequencies and the distributions of word frequencies among the abstracts. Thus, our model cannot distinguish between words that have different meanings in different contexts, such as bank, in the context of food bank versus bank loan. However, using bigrams and including food bank and bank loan as standalone words, as we did, deals with this issue to some extent.

5. Conclusion

Given the breadth and proliferation of the food security literature in recent years, using computational methods can provide a novel overview of the literature. Thus, we used various text mining techniques to analyze the geographic and thematic scope of the literature, as well as how this scope has shifted over time.

We found that the food security literature is disproportionately likely to be conducted in certain countries and regions. We also found a trend of the literature before 2000 focusing heavily on Africa, with Asia and high-income countries representing an increasingly large share of the literature through the past two decades. Finally, by comparing where the food security literature focuses with actual food security status of each country, we found little correlation between the geographic focus of the literature and the spatial distribution of real food insecurity. Countries that are both under-researched and severely food insecure, such as Angola, the DRC, Chad, Sudan, Somalia, and Yemen, should be a priority for future food security research.

We used topic modeling to examine the themes in the literature, and we found that initially the food security literature mostly dealt with the themes of Economic Policy and Global Food Security, and only recently have other themes such as Livelihoods and Climate & Sustainability come to occupy a significant share of the literature. We further found that certain themes, especially Nutrition and Health are more likely to occur in high-income countries. Finally, we found that the literature is broadly unified, that all themes occur across all world regions, and that there are no major disjunctive thematic clusters in the research, suggesting that there are no siloed subdomains in the literature and that the concept of food security is a meaningful nexus for applied research from a wide range of academic disciplines.

Overall, this study represents the first high-level assessment of the spatial and thematic scope of the food security literature. These results provide a roadmap for future research to serve currently understudied regions and topics, to improve our comprehensive understanding of food security globally.

Declaration of competing interest

We have no conflicts of interest to declare.

Acknowledgments

Computation for this paper was run in the Microsoft Azure cloud using resources provided by an AI for Earth grant. The dataset of abstracts, data on the model results, and all code necessary to reproduce our results are available at github.com/mcooper/mine-food-security. We thank Chris Barrett for participating in early conversations related to this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gfs.2020.100392.

References

Acosta, L.A., Enano, N.H., Magacle-Macadang, D.B., Engay, K.G., Herrera, M.N.Q., Nicpior, O.B.S., Sumilang, M.I.V., Eugenis, J.M.A., Lucht, W., 2013. How sustainable is bioenergy production in the Philippines? A conjoint analysis of knowledge and opinions of people with different typologies. Appl. Energy 102, 241–253.
Agresti, A., 2006. An Introduction to Categorical Data Analysis, second ed. 
Amiard, J.-C., Amiard-Triquet, C., Charbonnier, L., Mesnil, A., Rainbow, P.S., Wang, W.-X., 2008. Bioaccessibility of essential and non-essential metals in commercial shellfish from Western Europe and Asia. Food Chem. Toxicol. 46 (6), 2010–2022.
Appiah, F., Oduro, I., Ellis, W.O., 2011. Proximate and mineral composition of artocarpus altilis pulp flour as affected by fermentation. Pakistan J. Nutr. 10 (7), 653–657.
Aranda, S., Alcaínez-Colet, A., Blanco, E., Borraís, E., Callott, S., Sadibè, E., Croce, L.D., 2019. Chromatopin capture links the metabolic enzyme ALCy to stem cell proliferation. Science Advances 5 (3). https://doi.org/10.1126/sciadv.aar2448.
Avanza, S., Sharma, P., Kumar, S., Nayan, R., Khanna, P.K., Zaidi, M.G.H., 2012. Goldnanoparticle induced enhancement in growth and seed yield of Brassica juncea. Plant Growth Regul. 66 (3), 303–310.
Awange, J.L., Aluoch, J., Ogállo, L.A., Omulo, M., Omondi, P., 2007. Frequency and severity of drought in the Lake Victoria region (Kenya) and its effects on food security. Clim. Res. 33, 135–142.
Barrat, A., Barthelemy, M., Pastor-Satorras, R., Vespignani, A., 2004. The architecture of complex weighted networks. In: Proceedings of the National Academy of Sciences of the United States of America.
Berhanu, B., White, M., 2000. War, famine, and female migration in Ethiopia, 1960-1989. Econ. Dev. Cult. Change 49 (1), 91–113.
Blei, D.M., 2012. Probabilistic topic models. Commun. ACM 55 (4), 77–84. https://doi.org/10.1145/2133896.2133897.
Blei, D.M., Lafferty, J.D., 2007. A correlated topic model of science. Ann. Appl. Stat. 1 (1), 17–35.
Bormann, L., Mutz, R., 2015. Growth rates of modern science: a bibliometric analysis based on the number of publications and cited references. J. Assoc. Inf. Sci. Technol. 66 (11), 2215–2222. https://doi.org/10.1002/asi.23329.
Brinkley, C., Birch, E., Keating, A., 2013. Feeding cities: charting a research and practice agenda toward food security. J. Agric. Food Syst. Community Dev. 5 (4). https://doi.org/10.1145/2133896.2133820.
Brown, M., Antle, J., Backlund, P., Carr, E., Easterling, W., Walsh, M., Ammann, C., Attavanich, W., Barrett, C., Bellemare, M., Danchez, V., Funk, C., Grace, K., Ingram, J., Jiang, H., Maletta, H., Mata, T., Murray, A., Ngugi, M., Ojima, D., O'Neil, B., Tebaldi, C., 2015. Climate Change, Global Food Security, and the U.S. Food System. Technical Report. USDA Technical Document, Washington, DC.
Brown, M.E., Backer, D., Billing, T., White, P., Grace, K., Doocy, S., Huth, P., 2020. Empirical Studies of Factors Associated with Child Malnutrition: Highlighting the Evidence about Climate and Conflict Shocks. Food Security.
Buttel, F.H., 2005. Ever since highlander: the politics of agricultural research activism in the molecular age. Agric. Hum. Val. 22 (3), 275–283.
Caccavale, O.M., Giaffreda, V., 2020. The Proteus Composite Index: towards a Better
