

Generating crop calendars with Web search data

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2012 Environ. Res. Lett. 7 024022

(<http://iopscience.iop.org/1748-9326/7/2/024022>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 193.137.168.194

This content was downloaded on 28/09/2016 at 10:12

Please note that [terms and conditions apply](#).

You may also be interested in:

[Global crop exposure to critical high temperatures in the reproductive period: historical trends and future projections](#)

Sharon M Gourdji, Adam M Sibley and David B Lobell

[Untangling relative contributions of recent climate and CO2 trends to national cereal production in China](#)

Wei Xiong, Ian Holman, Erda Lin et al.

[Public interest in climate change over the past decade and the effects of the 'climategate' media event](#)

William R L Anderegg and Gregory R Goldsmith

[Subfield profitability analysis reveals an economic case for cropland diversification](#)

E Brandes, G S McNunn, L A Schulte et al.

[Effects of land cover change on moisture availability and potential crop yield in the world's breadbaskets](#)

Justin E Bagley, Ankur R Desai, Paul A Dirmeyer et al.

[Response of double cropping suitability to climate change in the United States](#)

Christopher A Seifert and David B Lobell

Generating crop calendars with Web search data

Marijn van der Velde¹, Linda See¹, Steffen Fritz¹, Frank G A Verheijen², Nikolay Khabarov¹ and Michael Obersteiner¹

¹ International Institute of Applied Systems Analysis (IIASA), Ecosystem Services and Management Program, Schlossplatz 1, Laxenburg, A-2361, Austria

² Department of Environment and Planning, Center for Environmental and Marine Studies (CESAM), University of Aveiro, 3810-193 Aveiro, Portugal

E-mail: velde@iiasa.ac.at

Received 13 March 2012

Accepted for publication 24 May 2012


Published 22 June 2012

Online at stacks.iop.org/ERL/7/024022

Abstract

This paper demonstrates the potential of using Web search volumes for generating crop specific planting and harvesting dates in the USA integrating climatic, social and technological factors affecting crop calendars. Using Google Insights for Search, clear peaks in volume occur at times of planting and harvest at the national level, which were used to derive corn specific planting and harvesting dates at a weekly resolution. Disaggregated to state level, search volumes for corn planting generally are in agreement with planting dates from a global crop calendar dataset. However, harvest dates were less discriminatory at the state level, indicating that peaks in search volume may be blurred by broader searches on harvest as a time of cultural events. The timing of other agricultural activities such as purchase of seed and response to weed and pest infestation was also investigated. These results highlight the future potential of using Web search data to derive planting dates in countries where the data are sparse or unreliable, once sufficient search volumes are realized, as well as the potential for monitoring in real time the response of farmers to climate change over the coming decades. Other potential applications of search volume data of relevance to agronomy are also discussed.

Keywords: corn, maize, crop planting, internet, web search data, crowd-sourcing, crop model, harvest, web tools, Google Insights for Search

 Online supplementary data available from stacks.iop.org/ERL/7/024022/mmedia

1. Introduction

Access to the Internet has fundamentally impacted the lives of many people and become increasingly interwoven into society. Global Web penetration has reached 30% and almost 80% in North America alone (Miniwatts Marketing Group 2011). One measure of Web activity is the volume of Web searches through the major search engines such as Google and Yahoo. The search volume in October 2005 was estimated at more than 5 billion based on the top 5 search engines (Nielsen/NetRatings 2005), and by the end of 2009, had increased to more than 130 billion searches in the month of

December (comScore 2010). To analyze this increasing search activity both spatially and temporally, Google has produced three tools: Google Trends, Google Insights for Search and Google Correlate (Mohebbi *et al* 2011). The potential for tools like these in the area of medical research and public health has been highlighted by Noll-Hussong and Lahmann (2011), and several examples have recently appeared relating search volume to incidence data, e.g. influenza (Ginsberg *et al* 2009, Mohebbi *et al* 2011), dengue fever (Althouse *et al* 2011), kidney stones (Breyer *et al* 2011), stroke prevalence (Walcott *et al* 2011), deaths by suicide (Yang *et al* 2011) and depression (Yang *et al* 2010). Other areas of application include the

forecasting of private consumption (Vosen and Schmidt 2011), unemployment (Tefft 2011) and interest in the public understanding of science (Baram-Tsabari and Segev 2011). The advantages of using Web search data for understanding trends and for forecasting include: contemporaneity and timeliness of the data, which are available from 2004 on a real-time basis; spatial coverage, which is sub-national (e.g. state and county in the USA) when sufficient search volumes are available; and open access to a rich database that could potentially replace costly field data (Mohebbi *et al* 2011).

Thus far, purely environmental applications of Web search query data are rare. A recent paper (Sherman-Morris *et al* 2011) examined the potential of searches for hurricane information to inform the communication of weather messages. However, opportunities clearly exist in relation to cyclical phenomena (see e.g. Mathias *et al* 2009). For example, analysis of the search term 'soil erosion' reveals an increased interest in spring and autumn, and a low interest during the summer when crops and vegetation cover are well established. These types of analysis hold potential for improved timing or targeting of specific educational or extension campaigns (cf Baram-Tsabari and Segev 2011). Another seasonal activity is cropland management, which follows a regular pattern of events. In addition to activities such as purchasing seeds and fertilizer, farmers must make decisions about when to plant and harvest their crops, which are based on factors such as rainfall forecasts, temperature, soil moisture, and increasingly technological and socio-economic factors (Sacks *et al* 2010). For example, Kucharik (2006) found that changes in technology, such as corn hybrids that are tolerant to colder temperatures, improved planting equipment and conservation tillage, may be the main contributing factors to the earlier corn planting dates in the USA relative to the early 1980s. Planting dates might also be chosen in relation to the harvesting date of a previous crop or to ensure favorable conditions during a critical stage of growth (Sacks *et al* 2010).

Planting dates are necessary inputs for crop models, whether at the plot level or globally. Most global crop models use climate to predict planting dates, in addition two global products are now available (Sacks *et al* 2010, Portmann *et al* 2010). However, the information, which has been collected from sources such as FAO and the USDA, generally refers to the 1990s and early 2000s. Therefore, Web search data provides a potential source of more up-to-date information that effectively integrates different climatic, social and technological factors that affect these dates, where timeliness of the data and ease in data collection and analysis may be one of the main advantages of search data for this and many other applications, as highlighted by Goel *et al* (2010). Improved spatially explicit knowledge about timing of cropland management practices could be used to calibrate and/or evaluate crop model response, improve assessments of food security and inform climate change adaptation potential. The present study aims to provide a proof of concept, using Google Insights for Search and Google Correlate, to investigate whether search activity can be used to determine

the crop calendar of corn in the USA. Wheat and soybean were also attempted but with less success, and are included to highlight the current limitations of the methodology. The USA provides a good case study since Web usage is widespread among farmers (USDA 2011a) and the search language is English. The results were then compared with crop calendar data from Sacks *et al* (2010).

2. Data and methods

Google Correlate was first used to establish whether the search audience is likely to be farmers. This tool provides search terms that exhibit the highest correlation with the normalized search activity (σ ; standard deviations away from the mean) of a search term of interest (i.e. whose search frequency follows a similar pattern), which was used at the state level to examine the types of terms correlated with 'corn planting', 'corn harvest', etc (see Google Correlate Tutorial online). A similar analysis was undertaken for wheat and soybean.

Google Insights for Search was then given the terms 'corn planting', 'corn harvest', etc (see supplementary table 1a available at stacks.iop.org/ERL/7/024022/mmedia) for the period 2004 onwards for the USA in all search categories. Similar searches were undertaken for wheat and soybean, all of which are widely cultivated in the USA. The original idea was to illustrate the method on all three crops. However, search volumes were too low for wheat and soybean beyond a national level result. This is a function of the way search volumes are processed. Google Insights for Search normalizes and scales search volume data between 0 and 100 by dividing the search volumes by the highest search volume and multiplying by 100 (see Google Insights for Search (2011) online help). When the search volumes are too low relative to other search terms, which reflects popularity of the term rather than absolute search volumes, the results cannot be displayed below a national level and the data are only available at a monthly resolution. Thus, for wheat and corn, the peaks of Web search activity corresponding to planting and harvesting could only be determined as the month in which they occurred while a more precise planting and harvesting date and length of the growing season was possible for corn. The spatial distribution of corn by US state based on search volumes was then compared to the spatial extent of the cultivation of corn. For the reasons stated above, this was not possible for wheat and soybean. Finally, state level planting and harvest dates for corn were derived where they were available at a weekly temporal resolution. The dates were compared with those reported in Sacks *et al* (2010) in terms of statistical similarity (see supplementary text available at stacks.iop.org/ERL/7/024022/mmedia) as well as the NASS Quick Stats Progress Reports for 2011 (USDA-NASS 2011e).

3. Results

3.1. The search population

Results from Google Correlate showed that the search term with the highest correlation to 'corn planting' is 'corn growth' with a Pearson correlation coefficient of 0.975, and that the

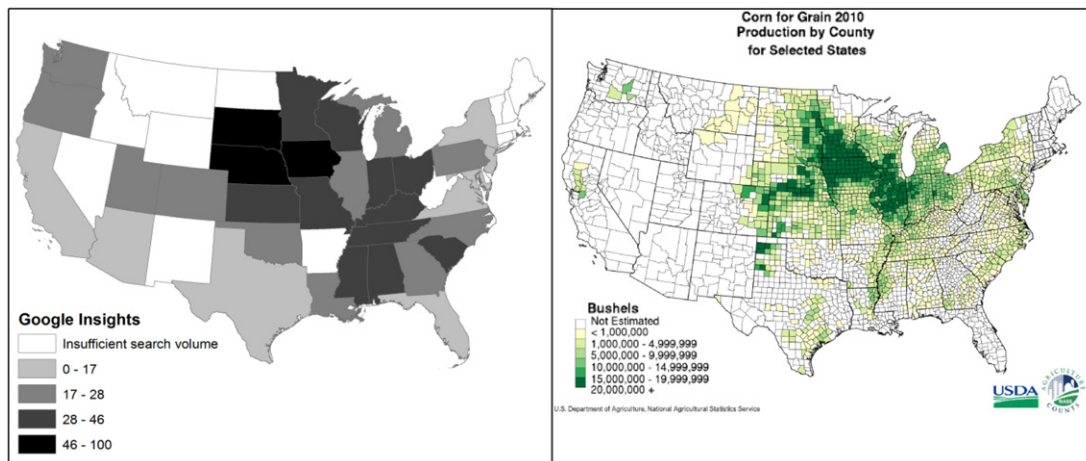


Figure 1. Comparison of normalized Google Insights Web search volume for ‘planting corn’ (left panel) with the production of corn by state (right panel; one metric ton = 39.37 bushels of corn; source: USDA-NASS 2010).

search activity for these terms is highest in the state of Iowa, which is the top corn producer (see supplementary figure 1 available at stacks.iop.org/ERL/7/024022/mmedia). Other terms found to be highly correlated with ‘corn planting’ include ‘bushel of corn’, ‘corn yield’, ‘corn diseases’ and ‘pedal tractor’. A list of the top 25 correlated search terms is provided in supplementary table 2 (available at stacks.iop.org/ERL/7/024022/mmedia), which shows that most are related to agriculture. However, there are also other terms on the list that are unrelated to agriculture such as food related terms (snack mix recipes, snack mixes) and toys (ertl toys, john deere farm toys). Moreover, similar searches with wheat and soybean yielded very little relation to agriculture, e.g. many graduation related words for soybean planting and a range of unrelated words for wheat. Although 67% of farmers are connected to the Web (USDA 2011b), we cannot definitively establish the composition of the search population using this tool. However, we can observe that for corn, the community searching appears to have a larger agricultural nature than for wheat or soybean. This may also be reflected in the higher search volumes available at state level and at a weekly temporal resolution.

3.2. National level

The results of searches for corn, wheat and soybean in combination with terms such as seed, planting, and harvest at national level reveal a clear seasonal periodicity in search volume from 2004 to 2011 (supplementary figure 2 available at stacks.iop.org/ERL/7/024022/mmedia). Corn shows a peak for planting in the end of May, and a peak for harvest in the middle of October. In the case of wheat, a peak is observed in June for harvest, while planting is characterized by two peaks differentiating spring and winter wheat planting. Soybean revealed peaks in planting around April/May and harvest in September/October.

Figure 1 shows the spatial pattern of Web search volume for ‘corn planting’ at the state level compared to corn production by county (USDA-NASS 2010), which show similar patterns. A scatter plot between peak search volume

by state and corn production is provided in supplementary figure 3 (available at stacks.iop.org/ERL/7/024022/mmedia) ($r = 0.68, p < 0.0001$). Iowa and Nebraska are two of the states with the highest Web search volume, and also are among the highest corn producing states with annual corn production in 2010 of 2.2 and 1.5 billion bushels (55.9 and 38.1 million tons) respectively (USDA 2011b) or 17.3% and 11.8% of total US production. Illinois is the second corn producing state with 1.9 billion bushels (48.3 million tons) in 2010 with 15.6% of total US production but shows a lower search volume. This could be a function of the normalization algorithm of Google Insights for Search, which reflects popularity rather than trends in absolute search volumes. This state has a broader demographic, which will have a wider search interest and will be influenced by the presence of Chicago, so this result may simply reflect the lower popularity of farm-related search terms in this state. On the other hand, states such as Arizona, Oregon and Utah have search volumes yet do not appear to produce any corn based on the USDA map. However, the USDA (2011b) reports that these states do produce a very small amount of corn.

The spatial pattern of ‘wheat planting’ resulted in the highest search activity in Kansas, which is also the top wheat producing state (USDA-NASS 2011c). Searches for ‘winter wheat planting’ and ‘spring wheat planting’ did not produce sufficient search volumes for spatial disaggregation. A search on the term ‘soybean planting’ showed search volumes in the state of Iowa, which is the top soybean producing state (USDA-NASS 2011d) and Missouri, another prominent soybean planting state. However, search volume in other states was low, which indicates that applicability to both wheat and soybean is limited at present.

Google Insights for Search was then used to examine the search volumes for ‘corn’ in combination with ‘seed’, ‘planting’, ‘herbicide’, ‘fungicide’ and ‘harvest’ at the national level. The annual normalized data were averaged for 2007–11 and are shown in figure 2. Data prior to 2007 were not included to produce a more recent set of dates. Vertical lines were drawn at the peaks to indicate the average corn planting and harvesting dates. The thinner vertical lines for

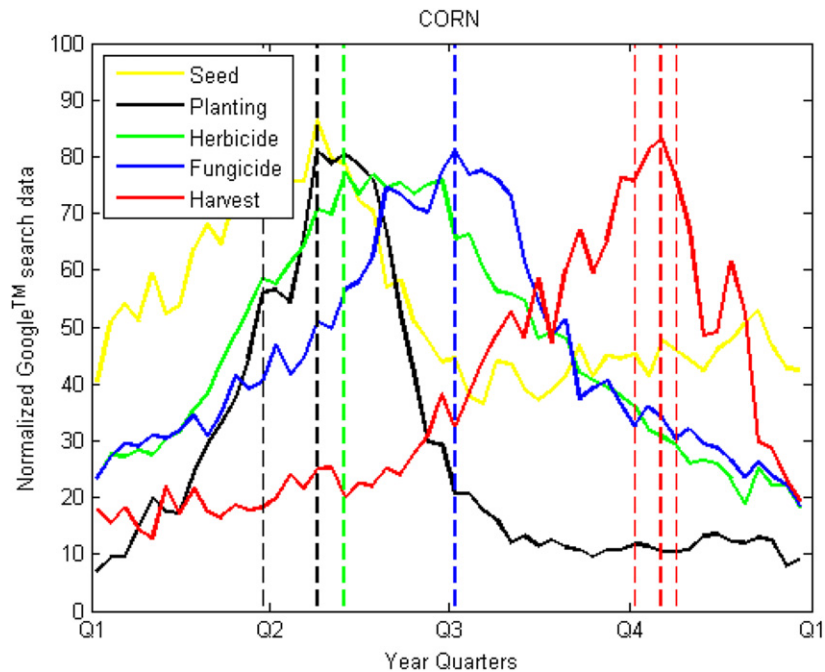


Figure 2. Annual average (2007–2010) US national normalized Google Web search data performed on 21/10/2011 for ‘corn seed’, ‘corn planting’, ‘corn herbicide’ and ‘herbicide’, ‘corn fungicide’ and ‘fungicide’, and ‘corn harvest’. The peaks of ‘herbicide’, ‘fungicide’ and respectively ‘corn herbicide’ and ‘corn fungicide’ line up but the latter are only available at a monthly and not a weekly resolution; therefore the generic and corn specific search terms were combined. Vertical lines indicate the peak search data for the corresponding term averaged from 2007 to 2010. The earliest and latest peak in both corn planting and corn harvest from 2007 to 2010 is indicated with thinner vertical lines with corresponding colors. Note that the vertical line for the latest planting date is obscured by herbicide, while the line for seed is obscured by the earliest planting date.

planting and harvest indicate earlier and later peaks around the main peaks, to indicate the range due to variability in climate from 2007 to 2011. The data suggest a national average corn planting date of 25 April and a harvest date of 17 October with a growing season length of 175 days. Although the corn seed peak corresponds to the corn planting peak, searches for corn seed start much earlier compared to planting. Farmers will first focus on minimizing weeds with herbicides after which the crops will become increasingly susceptible to fungus infestation, which is then treated with fungicides. It should be acknowledged, however, that a single planting and harvesting date for corn for a country the size of the USA is difficult to interpret due to the large spatial variability in planting and harvesting times nationally. For this reason, an analysis at state level has also been carried out (section 3.3).

Deriving planting and harvesting dates for wheat is more complex than for corn because of the growing of both spring and winter wheat. At the national level, searches for ‘spring wheat planting’, and ‘winter wheat planting’ result in planting dates in April, and in September and October respectively, which corresponds to the planting of this crop in the USA. ‘Wheat harvest’ is dominated by a single peak in June for winter wheat but the spring wheat harvest is not clearly identifiable, which occurs later in the season. Searches related to soybeans show clear peaks in activity during the month of May, which coincides with general planting of this crop, and during September/October, which corresponds to harvesting. Both wheat and soybean have insufficient search volumes at this time to provide data at a weekly resolution

so only the main month of planting and harvest are currently available. Thus, no further spatial disaggregation to state level is possible.

3.3. State level

The corn planting dates derived from Google Insights for Search at state level were based on a combination of corn specific terms (e.g. corn planting + planting corn + plant corn) for 2011 for those US states (see supplementary tables 1a and 1b available at stacks.iop.org/ERL/7/024022/mmedia) where the search volume was sufficient to produce output with a weekly temporal resolution, i.e. Iowa, Mississippi, Missouri, Minnesota, Texas and Wisconsin. The date of the peak search activity was extracted and plotted as a square with a black outline (figure 3(a)). For the remaining states with insufficient search volume, corn planting dates were based on a combination of ‘planting’ and ‘corn planting’ (figure 3(a); the bars indicate the location of the peak in search volume by state). Although we acknowledge that ‘planting’ is a generic term, search volumes for ‘corn planting’ and ‘planting’ are highly correlated ($r = 0.94$; $p < 0.00001$) between 2008 and 2011 (see supplementary figure 4 available at stacks.iop.org/ERL/7/024022/mmedia). Thus, we have used this combination of search terms to provide additional weekly dates at the state level that would otherwise not have been possible due to low search volumes as another way of illustrating the potential of utilizing Web search volumes in this manner. Superimposed on the bars are the planting

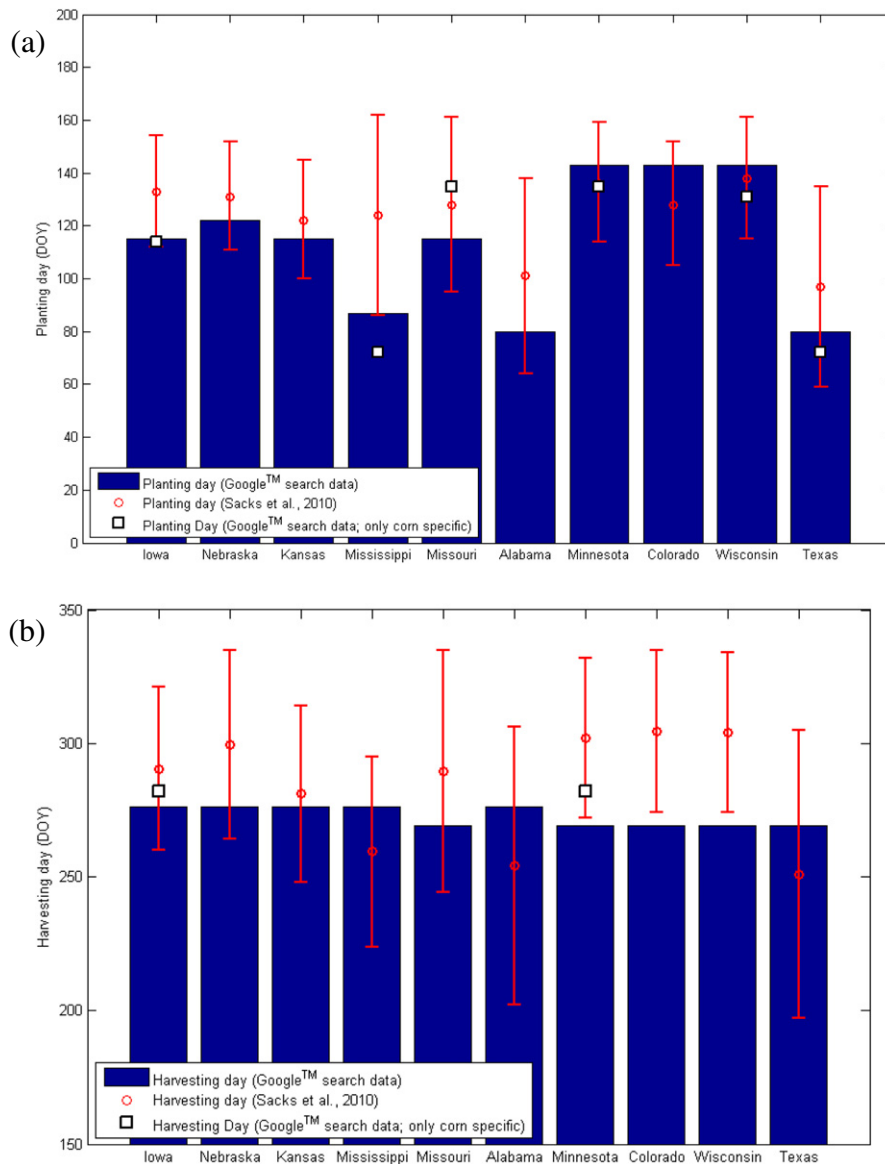


Figure 3. (a) State level corn planting dates and (b) state level harvesting dates derived from Google Web search data (bars) using both corn specific and generic planting/harvesting to boost search volumes compared to planting dates (error bars indicate average, begin and end) as reported by Sacks *et al* (2010). The white squares indicate corn specific planting dates derived from Google Web search data for 2011 (only indicated where ‘corn planting’ or ‘corn harvest’ yielded data that were reported at weekly temporal resolution). The dates of the searches are provided in supplementary table 1 (available at stacks.iop.org/ERL/7/024022/mmedia). Supplementary figure 4 (available at stacks.iop.org/ERL/7/024022/mmedia) presents planting windows at state level derived from search volume data.

dates reported by Sacks *et al* (2010), providing the typical average, start and end dates for each state in red. Overall the results show reasonable correspondence between the two datasets and statistical comparison reveals similarities (see supplementary text available at stacks.iop.org/ERL/7/024022/mmedia).

For Iowa, Missouri, Minnesota and Texas the peaks coincide well with the dates of Sacks *et al* (2010). However, the Web search data suggests an earlier corn planting date for Mississippi when compared to the global data set, which was originally taken from USDA-NASS (1997) for the USA. High weekly ‘corn planting’ Web search volumes for Mississippi for 2011 started on 27 February (DOY 58) and ended on 26 March (DOY 85). Reports on planting of corn in Mississippi

in 2011 confirmed an early planting with the majority of corn planted during the last 10 days of March (Coblentz 2011). Unfortunately no data were available from the NASS Quick Stats Progress Reports for 2011 for the state of Mississippi, which could have provided additional evidence for this finding. However, data were available from USDA-NASS (2011e) for Iowa, Missouri, Minnesota, Wisconsin and Texas, i.e. the other five states where corn specific terms were used to generate the planting dates. NASS provides weekly data on the cumulative percentage of fields planted. From this cumulative curve it is possible to determine the planting peaks, which could then be compared to the Web search peaks. They suggest similar planting peaks for Iowa, Missouri, Wisconsin or with user generated information (of geowiki.

org, Fritz *et al* 2012) and Minnesota (i.e. within 10 days) but a disagreement of ~40 days for Texas (see supplementary table 3 available at stacks.iop.org/ERL/7/024022/mmedia) where the search peak occurred much earlier than the actual planting peak.

Planting windows by state can also be established from search volume data using a threshold in search volume (see an example for 2011 in supplementary figure 5 (available at stacks.iop.org/ERL/7/024022/mmedia) using a peak threshold of 50 to establish the window and indicating the date at which the maximum peak occurred). The peak of planting from USDA-NASS (2011e) was then superimposed on supplementary figure 5 (available at stacks.iop.org/ERL/7/024022/mmedia) for all states for which data were available, including those states where dates were derived from search phrases that included the more generic search term 'planting'. The pattern generally shows that peaks in search volume occur before the peak in planting, suggesting that they are indicative of precursors to planting or coincide with the start of the planting process.

Figure 3(b) provides corn harvest dates, derived by the same method used for planting dates. It was possible to determine harvesting dates with corn specific search terms (corn harvesting + corn harvest) for Iowa and Minnesota only. In both cases, USDA-NASS (2011e) indicated DOY 296 while the Web search dates were earlier at 282, which again suggests that searching coincides with the start of the actual event or could be a precursor. The data of Sacks *et al* (2010) show much more variation by state for these dates than the Web search data would indicate where the statistical comparison shows only similarities between the average harvest dates of Sacks *et al* (2010) and the search volume derived dates (see supplementary text available at stacks.iop.org/ERL/7/024022/mmedia). This may be due to harvest being an activity which—compared to planting—is much more closely related to the festivals and cultural activities of a society (e.g. Thanksgiving). Thus, the 'crowd' being sampled for 'harvest' may be much broader than the farming community compared to searches for 'planting'. This is also reflected in the search volumes for 'corn harvest' by state which were only large enough to report for Iowa and Minnesota.

4. Discussion

The searches undertaken considered only the word planting in the search phrase. Planting dates could be further refined using search data on meteorological conditions and pest risk used by farmers. For example, before planting spring wheat, farmers wait until the greatest frost risk has passed (Sacks *et al* 2010). A Web search for 'frost risk' indicates peak volumes between March and May. Similarly, farmers generally wait to plant winter wheat until the risk of Hessian fly infestation has subsided (Sacks *et al* 2010). The search term 'Hessian fly' resulted in peaks in September. Furthermore, online purchases of farm equipment may provide additional useful information. The question remains open as to how far search terms reflect preparatory behavior and will be a precursor to the actual

implementation of a practice, or will reflect a response in real-time. This will also be differentiated by agricultural practice. For instance, searching for crop planting dates might inform the decision when to plant, while searches for fungus treatment might reflect observations of fungus infestation in the field. How many days there might be between searching for planting dates and the actual planting will depend on behavioral characteristics of the individual farmer that will need to be elucidated. This will also aid in an improved definition of planting and harvesting windows from Web search data.

Moreover, the crop calendar of Sacks *et al* (2010) is based on older data, where crop calendars may have shifted as a result of changes in climate. Adjusting planting dates may be one of the adaptation interventions that farmers take to maintain or increase yields (Lauer *et al* 1999). For instance, farmers can adapt planting dates in response to expected shifts in seasonal water deficits. Monitoring changes in real-time Web search activity over the coming decades for key terms such as those presented here might be used to aid in predicting how farmers are responding to climate change.

Other agronomical uses could include combining real-time weed and pest infestation searches with remotely sensed satellite observations of phenological development (Wardlow *et al* 2007) or with user generated information (cf geowiki.org, Fritz *et al* 2012). For instance, in 2009, the searches for 'weeds' and 'blight' (including 'potato blight') both peaked in comparison to other years, which corresponds to online reports of blight impacting both potatoes and tomatoes (Martin 2010). Monitoring Web search query data during extreme events such as droughts or heat waves may yield in improved estimates of irrigation water use, or monitor crop loss due to heavy precipitation (see e.g. van der Velde *et al* 2010, 2012), and similar to detecting influenza epidemics (Ginsberg *et al* 2009), Web search query data may be used to detect and track contagious diseases and pests in livestock and arable farming.

However, we also recognize that there a number of limitations with this method. Adequate search volume is the main limiting factors at present. Firstly, we acknowledge that the potential for expanding the method to other countries, particularly at lower latitudes where Sacks *et al* (2010) found that planting dates were more difficult to predict using climate variables, relies on the continued expansion of the Internet to the developing world and improvements in literacy, where at present search volumes are low and access to the Internet is not ubiquitous. It may take another decade before this type of analysis will become valuable for data-poor countries in Africa where 11.4% of the population is currently connected compared to a global average of 30% or approximately 2 billion Web users (Miniwatts Marketing Group 2011). By 2020, this number is projected to increase to 5 billion users (Fox 2011) and will substantially increase penetration figures in developing countries. The equivalent Portuguese search terms for planting in combination with different crops such as corn, soybean, wheat and sugar cane reveal similar seasonal patterns in Brazil. Supplementary figure 6 (available at stacks.iop.org/ERL/7/024022/mmedia) shows the result of the search

phrase ‘planting wheat’ translated into Portuguese (‘plantio trigo’) which shows distinctive peaks in the month of June, which corresponds with the planting of winter wheat. The highest search volumes are found in the states of Parana and Rio Grande do Sul, which have the highest production of wheat in Brazil (George *et al* 2009). Not all wheat growing states show appreciable search volumes, similar to what occurred with maize in the USA. Thus similar weaknesses in the methodology exist with this example. However, Brazil’s Internet access is also set to increase rapidly over the next decade. In 2011, Brazil had almost 76 million people online or 37.4% of the population (Miniwatts Marketing Group 2012). This is predicted to increase to almost 100 million by 2015 and 1 billion mobile broadband connections by 2022, which is predicted to substantially overtake fixed broadband connections (Beach 2011). Although this example was undertaken in Portuguese, we recognize that language will be an issue, particularly in those countries which are highly agricultural and have multiple languages so the task of extracting search volumes becomes more complex.

Another limitation concerns the search population. Google Correlate did not provide sufficient evidence of the composition of the search population. Thus, searches may have been carried out by individuals other than farmers and therefore for different reasons or in different contexts. Finally, it is clear that corn provided the best result to date but that search volumes on other US crops resulted in insufficient search volumes. Moreover, there are further limitations with sufficient search volumes spatially and temporally such that at present, disaggregation would be necessary involving other datasets such as climate. Therefore, until search volumes increase further, the method is currently limited in developing crop calendars at present. However, as growth in the Internet has been exponential over the last decade and is foreseen to continue, this potential may yet be realized in the future.

5. Conclusions

This paper examined the potential of using Web search volumes from the Google Insights for Search tool to create crop calendars. The results showed that the potential does exist, as planting dates for maize in the USA were shown to have statistical similarity to those of Sacks *et al* (2010). The peak in search volume also occurred before the planting peak based on data from USDA-NASS (2011a, 2011b, 2011c, 2011d, 2011e) suggesting that they occur near the start of the planting period or are precursors to these events. They also have the added advantage of being recent and easily updated. However, inadequate search volumes were the main limiting factor behind the ability to create state level crop calendars for other crops or for maize in some states. In order to be able to use this tool for this purpose, there would need to be two main changes. The first is an absolute increase in the amount of Internet searches globally, especially if this method is to be utilized in countries where data are currently sparse or in lower latitudes where crop calendars are not as directly related to climatic variables as those in higher latitudes. This will most certainly be realized in the short to medium

term as Internet and mobile connections are set to increase exponentially over the next 10 years (Beach 2011, Fox 2011). The second concerns improvements to the data provided by Google Insights for Search. At present the data are normalized relative to other search terms so indicate popularity rather than true trends in a specific search term. There is a real need to access absolute search volumes that can be normalized by users for a given purpose. The real trends over time will then become much more apparent. This problem has already been recognized as a limitation by Taylor (2011). We therefore call upon Google to reconsider the way in which they present their data to allow for better analysis of cyclical events.

Acknowledgments

The research leading to these results has received funding from the European Community’s Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 262937 (ISAC: Information Service on Agricultural Change). We also appreciate the input of two anonymous reviewers whose suggestions have improved the quality of the article significantly.

References

- Althouse B M, Ng Y Y and Cummings D A T 2011 Prediction of dengue incidence using search query surveillance *PLoS Negl. Trop. Dis.* **5** e1258
- Baram-Tsabari A and Segev E 2011 Exploring new web-based tools to identify public interest in science *Public Understand. Sci.* **20** 130–43
- Beach J 2011 Billion mobile broadband connections in Brazil by 2022 says regulator (www.telecoms.com/33856/billion-mobile-broadband-connections-in-brazil-by-2022-says-regulator/?utm_source=rss&utm_medium=rss&utm_campaign=billion-mobile-broadband-connections-in-brazil-by-2022-says-regulator)
- Breyer B N, Sen S, Aaronson D S, Stoller M L, Erickson B A and Eisenberg M L 2011 Use of Google Insights for Search to track seasonal and geographic kidney stone incidence in the United States *Urology* **78** 267–71
- Coblentz B 2011 Mississippi corn planted early, looking good (<http://deltafarmpress.com/corn/mississippi-corn-planted-early-looking-good>)
- comScore 2010 comScore reports global search market growth of 46 percent in 2009 (www.comscore.com/Press_Events/Press_Releases/2010/1/Global_Search_Market_Grows_46_Percent_in_2009)
- Fox W 2011 Computers and the web (www.futuretimeline.net/subject/computers-web.htm)
- Fritz S, McCallum I, Schill C, Perger C, See L, Schepaschenko D, van der Velde M, Kraxner F and Obersteiner M 2012 Geo-Wiki: an online platform for improving global land cover *Environ. Model. Soft.* **31** 110–23
- George H, Verbeke I and Corsi S 2009 Agro-MAPS. Mapping sub national land use statistics on a global scale (Rome: FAO) (www.fao.org/fileadmin/templates/ess/pages/rural/wye_city_group/2009/presentations/slides_2.1_VerbekeEtAl_Agro-MAPS_WYE2009_Rome04_nonotes.ppt)
- Ginsberg J, Mohebbi M H, Patel R S, Brammer L, Smolinski M S and Brilliant L 2009 Detecting influenza epidemics using search engine query data *Nature* **257** 1012–4
- Goel S, Hofman J M, Lahaie S, Pennock D M and Watts D J 2010 Predicting consumer behavior with web search *Proc. Natl Acad. Sci.* **107** 17486–90

- Google Insights for Search 2011 Google Insights for Search, help section (www.google.com/support/insights/bin/answer.py?hl=en-US&answer=87285, accessed: 14 November 2011)
- Kucharik C J 2006 A multidecadal trend of earlier corn planting in the central USA *Agron. J.* **98** 1544–50
- Lauer J G, Carter P R, Wood T M, Diezel G, Wiersma D W, Rand R E and Mlynarek M J 1999 Corn hybrid response to planting date in the northern corn belt *Agron. J.* **91** 834–9
- Martin J 2010 Spuds Unearthed! Exhibit digs up history of the potato. USDA Blog (<http://blogs.usda.gov/2010/10/04/%e2%80%9cspuds-unearthed%e2%80%9d-exhibit-digs-up-history-of-the-potato/#more-28987>)
- Mathias Y, Efron N and Shimshoni Y 2009 On the predictability of search trends (<http://googleresearch.blogspot.com/2009/08/on-predictability-of-search-trends.html#!/2009/08/on-predictability-of-search-trends.html>)
- Miniwatts Marketing Group 2011 Web world stats: usage and population statistics (www.webworldstats.com/stats.htm)
- Miniwatts Marketing Group 2012 Latin American Internet usage statistics (www.Internetworldstats.com/stats10.htm)
- Mohebbi M, Vanderkarn D, Kodysh J, Schonberger R, Choi H and Kumar S 2011 Google Correlate Whitepaper (www.google.com/trends/correlate/whitepaper.pdf)
- Nielsen//NetRatings 2010 Online searches grow 55 percent year-over-year to nearly 5.1 billion searches in December 2005 (www.nielsen-online.com/pr/pr_060209.pdf)
- Noll-Hussong M and Lahmann C 2011 Whiplash and werther effect—the potential of Google Insights for Search for Medical Research and Public Health *Fortschr. Neurol. Psychiatr.* **79** 340–4
- Portmann F T, Siebert S and Döll P 2010 MIRCA2000—global monthly irrigated and rainfed crop areas around the year 2000: a new high-resolution data set for agricultural and hydrological modeling *Glob. Biogeochem. Cycles* **24** GB1011
- Sacks W J, Deryng D, Foley J A and Ramankutty N 2010 Crop planting dates: an analysis of global patterns *Glob. Ecol. Biogeogr.* **19** 607–20
- Sherman-Morris K, Senkbeil J and Carver R 2011 Who's Googling what? what Internet searches reveal about hurricane information seeking *Bull. Am. Meteorol. Soc.* **92** 975–85
- Taylor S J 2011 Using Google Insights search volume *Decomposition* (<http://seanjtaylor.com/2011/04/04/using-google-insights-search-volume/>)
- Tefft N 2011 Insights on unemployment, unemployment insurance and mental health *J. Health Econom.* **20** 258–64
- USDA-NASS 1997 Usual planting and harvesting dates for US field crops *Agricultural Handbook No. 628* (Washington, DC: USDA-NASS) (<http://usda01.library.cornell.edu/usda/nass/planting/1990s/1997/planting-12-05-1997.pdf>)
- USDA-NASS 2010 Corn: production acreage by county (www.nass.usda.gov/Charts_and_Maps/Crops_County/cr-pr.asp)
- USDA-NASS 2011a Farm computer usage and ownership (<http://usda01.library.cornell.edu/usda/current/FarmComp/FarmComp-08-12-2011.pdf>)
- USDA-NASS 2011b Crop production (www.usda.gov/nass/PUBS/TODAYRPT/crop1111.pdf)
- USDA-NASS 2011c Kansas wheat history (www.nass.usda.gov/Statistics_by_State/Kansas/Publications/Crops/whthist.pdf)
- USDA-NASS 2011d Iowa's rank in United States agriculture (www.nass.usda.gov/Statistics_by_State/Iowa/Publications/Rankings/2011Rankings.pdf)
- USDA-NASS 2011e Quick Stats (www.nass.usda.gov/Quick_Stats/)
- van der Velde M, Tubiello F N, Vrieling A and Bouraoui F 2012 Impacts of extreme weather on wheat and maize in France: evaluating regional crop simulations against observed data *Clim. Change* at press (doi:10.1007/s10584-011-0368-2)
- van der Velde M, Wriedt G and Bouraoui F 2010 Estimating irrigation use and effects on maize yield during the 2003 heat wave in France *Agric. Ecosyst. Environ.* **135** 90–7
- Vosen S and Schmidt T 2011 Forecasting private consumption: survey-based indicators versus Google trends *J. Forecast.* **30** 565–78
- Walcott B P, Nahed B V, Kahle K T, Redjal N and Coumans J-V 2011 Determination of geographic variance in stroke prevalence using web search engine analytics *Neurosurg. Focus* **30** 1–4
- Wardlow B D, Egbert S L and Kastens J H 2007 Analysis of time-series MODIS 250 m vegetation index data for crop classification in the US Central Great Plains *Remote Sens. Environ.* **108** 290–310
- Yang A C, Huang N E, Peng C-K and Tsai S-J 2010 Do seasons have an influence on the incidence of depression? The use of an web search engine query data as a proxy of human affect *PLoS One* **5** e13728
- Yang A C, Tsai S-J, Huang N E and Peng C-K 2011 Association of web search trends with suicide death in Taipei City, Taiwan, 2004–2009 *J. Affect. Disord.* **132** 179–84