Using GPS to Detect and Prevent Data Falsification

RJ Marquette and Ryan Cecchi, United States Census Bureau

Decennial Statistical Studies Division, Quality Assurance Branch

We are researching using Global Positioning System (GPS) information as new avenue of detecting errors during field operations, and as a way to warn the lister before making an error, leading to improvements in the quality of the data collected, for minimal cost. During the 2010 Census address listing operation, our field staff used a handheld computer to collect the information about housing units. The device collected GPS coordinates while the lister indicated the relative location of the house on the map. Thus, we know where the device was actually located when the lister was completing that task, and our analysis found numerous examples of GPS coordinates clustered at one location, such as a coffee shop – a clear violation of procedures.

We will compare this GPS analysis with results from the Dependent Quality Control check we perform on the lister’s work, to see if it may be an effective tool for evaluating a lister’s work. We will also discuss our research into detecting this problem while the operation is ongoing, and we will describe some of the checks we have built into instruments to attempt to avoid the error in the first place.

# Background

The first operation of the 2010 Census was the Address Canvassing operation, during which the Census Bureau updates our master list of housing units (that is, places where people live or could live) in the United States. Census divides the country into collection blocks, which are then grouped into “assignment areas” (AAs) for operational convenience. A single Address Canvassing assignment area usually consisted of 50 to 150 housing units, but assignment areas of over 1,000 housing units existed. To accomplish this task, we issued a small, handheld computer the listers used to update addresses. When the lister confirmed each housing unit’s location on the automated map (known as collecting a manual map spot) by physically going to the location and tapping on the map in the area where the specific housing unit was located, the computer also collected Global Positioning System (GPS) coordinates. Ideally, the GPS coordinate and manual coordinate (calculated from the manual map spot assignment) for a housing unit would be on top of each other, because the lister was standing on the doorstep per procedures for spotting the unit in the correct location on the map, assuming the map data they were using was accurate and the GPS location was accurately measured by the technology.

The quality control program in place for the Address Canvassing operation was a batch sampling system called Dependent Quality Control (DQC). After the lister was finished, the assignment area was sent to a quality control (QC) lister. The QC lister’s handheld computer would select a random starting housing unit in each assignment area, and instruct the QC lister to verify the next several units on the ground (the sample size varied by the number of housing units in the assignment area). After the QC lister finished reviewing those units, the handheld computer would indicate whether the assignment had passed or failed DQC, and the QC lister would either check the rest of the assignment area (in the case of a failure), or he/she was finished with the DQC (pass). The overall DQC failure rate for the 2010 Address Canvassing operation was 8.43%. While this program is statistically reliable, it does not use GPS technology. So, it is easy to envision a neighborhood with few changes that would likely pass the DQC even if the lister never visited it – opening up the possibility of missing new units, or failing to remove units that had been demolished or no longer met the description of a housing unit (e.g., now a business address).

Indeed, we found some evidence that the listers did not always follow procedures. Prior to the 2010 Census, we ran a small test in 2007 that used the same handheld computers. After that test, members of the Census Bureau’s Geography Division discovered clusters of GPS map spots centered on coffee shops, restaurants, and individual houses. We dubbed these “curbstoning[[1]](#footnote-1) clusters,” meaning that the enumerators were falsifying data or at least not following procedures for the manual map spot collection. The test was too small to draw any conclusions, so we resolved to research this issue after the 2010 Address Canvassing operation was completed.

We also realized there might be benefit in simply calculating the distance between the manual map spot and the GPS map spot for the same unit, which was called the “strand length.” We wondered if “long strands” would provide similar information as the curbstoning clusters, while being easier to detect in the data.

These two avenues of research presented an entirely new way of checking the listers’ work, using technology already at our disposal. In addition, if the research was successful, it might be possible to incorporate appropriate checks into the software ahead of time. For example, if the lister was about to spot a unit far from where they are located, the software could give them a chance to confirm they are in fact at the correct location. This check could provide a drastic improvement in quality for little cost.

# Methodology

## Defining Curbstoning Cluster

Our first task was to define a curbstoning cluster. Ideally, the manual map spots recorded by listers should directly, or very closely, coincide with the GPS map spots reported by the handheld computer. Significant discrepancies between the two types of map spots could indicate violations of canvassing procedures or even data falsification. However, exact definitions must be established between divergences that occur as a result of GPS resolution issues[[2]](#footnote-2) and those due to poor lister performance.

Our initial definition of curbstoning clusters relies on the fact that townhouses, row houses, etc. would be the densest structures listers would typically encounter without being multi-unit structures. We decided having six or more housing units (excluding multi-units) within a 7.6 meter (25 feet) diameter circle would be extremely unlikely. This rule is somewhat arbitrary.

That rule is also difficult to implement. It is relatively easy to detect that a given housing unit has one or more neighbors within the specified distance, but calculating that information for every housing unit means that duplication and overlaps in the curbstoning clusters are detected, which then have to be resolved. So, for purposes of our analysis, we instead:

1. Divided the country into squares of length 0.0001 degrees on each side (the exact distance varied, but each side was approximately 40 feet or 12.2 meters).
2. Counted the number of GPS map spots within each square.
3. We considered a square that had six or more GPS map spots a curbstoning cluster.
4. In addition, we considered any adjacent square with six or more GPS map spots part of the same curbstoning cluster.

## Curbstoning Cluster Analysis

With a working definition of a curbstoning cluster in place, we were able to generate a list of all of the curbstoning clusters in the United States and begin our analysis. To evaluate the usefulness of the curbstoning cluster data, we compared AAs containing curbstoning clusters to the pass/fail decision of the DQC.

This analysis compared the DQC outcome for AAs containing curbstoning clusters with those that did not contain curbstoning clusters. If the AAs containing curbstoning clusters failed more often, then this would imply the presence of curbstoning clusters within a block is an indicator of poor lister performance.

For example, consider a situation where an AA has not had many new housing structures built or demolished in recent years. A lister could have sat at a local coffee shop and marked each housing unit as “Verified” and the AA would have probably passed the current DQC because, although the lister may not have done his/her work as expected, all units in the DQC sample were in the initial list provided to the lister. However, the curbstoning cluster analysis could detect that the lister was not following procedures and may not have checked all the structures thoroughly.

We selected a sample of curbstoning clusters and plotted them using Google Earth, which provides overhead imagery (usually taken from aircraft) of the entire world. We assigned each curbstoning cluster a code based on what we were able to determine from the overhead imagery. Each code had an associated decision to indicate whether the curbstoning cluster was “Suspicious” (lister appeared to not follow procedures), “Not suspicious” (lister seemed to follow proper procedures), or “Unable to Determine” (used when we were unable to make a judgment about the situation, due to factors such as outdated imagery).

## Long Strand Analysis

We also considered the distance between the manual map spot and the GPS map spot coordinates (or “strand length”) as an alternative indicator of poor lister performance.

Use of strand lengths represented a more general approach for detecting poor lister performance than curbstoning clusters, since we could consider discrepancies between manual and GPS map spots in isolation, rather than as a clustered group. For each map spot, we calculated the distance between the GPS and manual coordinates. For each AA, we calculated the average strand length, then tested that using statistical formulas to indicate AAs that had an unusually large average strand length. We compared the AA average strand length to:

* The overall average strand length in the USA (35.68 meters): We flagged the AA if its average strand length was greater than 35.68 meters + 3δ, where δ is the standard deviation based on the number of strands in the AA.
* The 90th percentile overall strand length of 13.68 meters: We flagged the AA if its average strand length was greater than 13.68 meters + 3δ.

We then compared the DQC fail rate for AAs flagged by one of these tests to AAs that were not flagged.

# Assumptions

## GPS quality

We assumed the GPS receivers in the field achieved a high level of accuracy relative to the true physical location. Without a proper level of fidelity with the true physical coordinates, GPS technology would be unsuitable for this application. We plan to test any developed algorithms in the field to minimize the possibility that any coordinates with poor GPS accuracy incorrectly trigger closer review.

## Multi-Unit Structures

Due to limitations in the data from 2010, we removed any units flagged as being part of a multi-unit structure from the analysis. We have little reason to believe multi-unit structures would be any different from single-family housing units with respect to the map spot accuracy, but we must test this in the future.

## Definition of Curbstoning Cluster

We assumed there are few instances where a square with 0.0001 degrees on each side contains six or more (non-multi-unit) housing units. We did not spot any situations that would violate this assumption in our review of the curbstoning clusters in the overhead imagery.

## AAs could be split

The Local Census Offices had the option of splitting AAs that were too large, but the curbstoning cluster analysis used un-split AAs. However, the split AAs were flagged individually as AAs with curbstoning clusters or not. For example, if AA 101 was split into 101A and 101B, and a curbstoning cluster was discovered in AA 101, we would have counted AA 101A as having a curbstoning cluster, and 101B as an AA without a curbstoning cluster.

# Results

## Using Curbstoning Clusters to Predict Falsification

Overall, we flagged 2.7% (19,556) of the 733,636 AAs as containing one or more curbstoning clusters. The flagged AAs had a DQC failure rate of 13.7%, compared to a DQC failure rate of 8.4% overall. Clearly, we can use the curbstoning cluster analysis to predict AAs with falsification and/or procedural violations. The AAs with no curbstoning clusters had a fail rate of 8.3%, which is very similar the overall failure rate.

Of the 19,556 AAs with curbstoning clusters, we selected a sample of 621 (3.2%) to take a closer look at the situations surrounding some of these AAs:

* 33.2% (206) were situations where the GPS coordinates were centered on one house (or other structure) far from the manual coordinates of the units;
* 32.7% (203) were situations where the GPS coordinates were clustered for houses on a given street (or streets) – for example, the lister was parked near the end of a street, doing the work there;
* 10.6% (66) had GPS coordinates that were clustered on a road and manual coordinates were not at housing units (i.e., in the woods or forest); and
* The remaining 23.5% (146) of the clusters were distributed among nine other categories.

The 621 sample curbstoning clusters were distributed among 988 AAs. The AAs assigned a “Suspicious” classification had a weighted DQC fail rate of 13.8%, and the AAs that were assigned a “Not Suspicious” classification had a weighted DQC fail rate of 15.1%. Both of these rates are higher than the overall DQC fail rate for Address Canvassing (8.4%), so the presence of a curbstoning cluster gives an indication that the listing was of lesser quality than the norm, and it does not matter whether the lister had a legitimate reason for not being able to complete the assignment as prescribed. This result is logical: Regardless of the reason, listers are more likely to make mistakes when they are not at the unit they are supposed to work while updating the information.

## Using Strand Length to Predict Falsification

Our first analysis flagged AAs with an average strand length that was significantly greater than the overall average strand length of 35.68 meters. This first analysis flagged 2.2% (16,113) of the total AAs. The flagged AAs had a DQC fail rate of 14.6%, compared to 8.3% for AAs that were not flagged by this analysis.

Our second analysis flagged AAs with an average strand length that was significantly greater than the 90th percentile of overall strand length (13.68 meters). This second analysis flagged 5.6% (40,991) of the total AAs. The flagged AAs had a DQC fail rate of 12.6%, compared to 8.2% for AAs that were not flagged by this test.

Both of these tests indicate that we can use average strand length as a predictor of falsification and/or procedural violations.

## Comparing Curbstoning Clusters and Strand Length Approaches

Only 0.8% (5,925) of the AAs had one or more curbstoning clusters and were also flagged for having an average strand length significantly greater than the overall average strand length. These AAs had a DQC fail rate of 15.7%.

Similarly, only 1.2% (8,797) of the total AAs had one or more curbstoning clusters and were also flagged for having an average strand length significantly greater than the 90th percentile of overall strand lengths. These AAs had a DQC fail rate of 15.2%.

We concluded that both curbstoning clusters and the long strand approaches have merit and usually flag different AAs, so using both methods may yield better results. Therefore, our plans for future operations will consider both methods.

# Next Steps

We are taking two approaches to implementing what we have learned: detection and prevention.

## Detection

Our sampling system for listing QC operations will incorporate both the curbstoning clusters and the strand length detection algorithms, which will flag AAs for further review. As part of our testing, we will need to compare the flagged AAs to the results during the DQC to determine whether we are correctly flagging AAs that appear falsified. Since the instruments warns the user about potential long strands and perhaps curbstoning clusters (see “Prevention”, below), the data we analyzed above may not be applicable – i.e., the warnings may change the behaviors of the enumerators enough to invalidate the detection algorithms we used in this analysis. Therefore, we will analyze and continue to refine our detection algorithms in future applications.

## Prevention

As we redesign our data collection instruments in preparation for the 2020 Census, we have been incorporating edit checks in the data-collection instruments (both listing and enumeration) to warn the user if the long strand issue arises. For example, during the Nonresponse Follow-Up operation, if the user is too far from the unit when they are about to start the interview, the instrument asks, “You may be too far from <address>. Continue interview?” Our listing instrument will ask a similar question when the user places a manual map spot far from their current location. Even if the warning is triggered, neither instrument prevents a user from continuing the interview/placing the map spot in the current location.

We have several issues to research regarding the strand length:

* A poor GPS fix could cause the warning to trigger incorrectly, which could result in the enumerator moving from a correct location to satisfy the warning.
* The optimal strand length is an open issue. Too many warnings will cause enumerators to ignore them, while too few warnings may affect the quality of the data. We need to strike a balance.
* Related to the prior bullet, the optimum strand length for urban areas versus rural areas is likely to be different.
* For enumeration, there is an assumption that the map spot we have in our records is correct. If the map spot is wrong, the strand length warning could be triggered incorrectly.

Our next step is to research and test all of these issues out in the field. To mitigate the difficulty of implementing different strand lengths for testing purposes, both instruments use parameters to specify the threshold strand length on an assignment-by-assignment basis – the value is not hard coded in the software.

We have not incorporated the curbstoning cluster checks into our plans yet. The biggest issue is that we have not had enough programming resources and time to implement it. Additionally, we need to devise an efficient way to detect the situation, and we may need to refine the definition of a curbstoning cluster.

References:

Cecchi, R., and Marquette, R. (2012), 2010 Census: Global Positioning System Evaluation, 2010 Census Program for Evaluations and Experiments.

1. “Curbstoning” is a term meaning, “falsifying data.” However, these curbstoning clusters may not, in fact, be falsification – for example, a lister may have encountered a locked gate that prevented them from getting to the front door of one or more housing units. [↑](#footnote-ref-1)
2. Due to atmospheric interference and other factors, the GPS coordinates collected by the handheld computer can be several meters or more from the correct location on the ground. In addition, confusing situations can arise for the listers; for example, if a road is placed incorrectly on the map, the GPS may appear to be giving an incorrect location when in fact it is the map that is wrong. [↑](#footnote-ref-2)