# Virginia Election 2020 Analysis

## Abstract

Multiple different aspects of the Virginia 2020 Presidential election results were analyzed to uncover numerous significant irregularities and inconsistencies that warrant additional investigation;   The published official results have clear numerical errors and inconsistencies; The Election Fingerprints at the Per County Congressional District level do not match the expected fingerprint of a free and fair election; There does not appear to be any mathematical way for the official Per-Precinct and Per-County reported data to be reconciled.  All analysis was performed on officially released election results from the Virginia Department of Elections.

# Detailed Analysis Report

There are numerous anomalies and difficulties in trying to analyze and evaluate the VA 2020 Election data, the most significant ones are summarized and presented here. These anomalies are summarized in three categories:

1. The published official results have clear numerical errors and inconsistencies.

2. The Election Fingerprints at the Per County Congressional District level do not match the expected fingerprint of a free and fair election.

3. The use of "virtual" absentee ballot collection and counting precincts at the per county congressional district level makes it impossible to directly account for in-person votes and absentee votes for precincts. Furthermore, there does not appear to be a mathematical way for the Per-Precinct and Per-County data to be reconciled.

## 1. Clear errors and inconsistencies in the published results

Multiple reports on the VA Department of Elections (DoE) give inconsistent results for the 2020 election. Figure 1 shows the total number of votes in the 2020 Presidential Election as 4,486,821 while Figure 2 shows the number of votes cast as 4,460,524. The difference between these two officially reported numbers is **26,297** votes.

Figure 1 also shows the most egregious and obvious example of a data inaccuracy that appears on the VA Dept of Elections official reporting of the election results. They report turnout, or percentage of people voting versus total registered voters, of 81.48%. This percentage is verifiably incorrect by simple
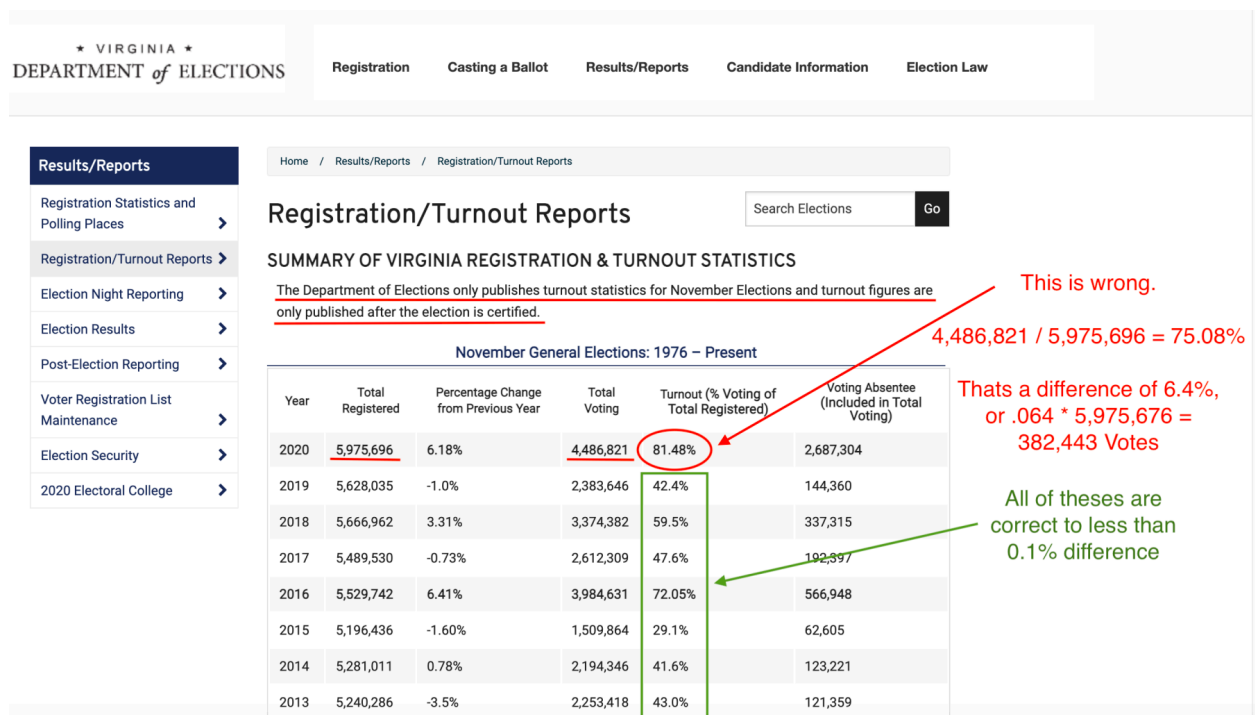


Figure 1: Screen capture (2021-08-10) of https://www.elections.virginia.gov/resultsreports/ registrationturnout-statistics/ showing data discrepancy.
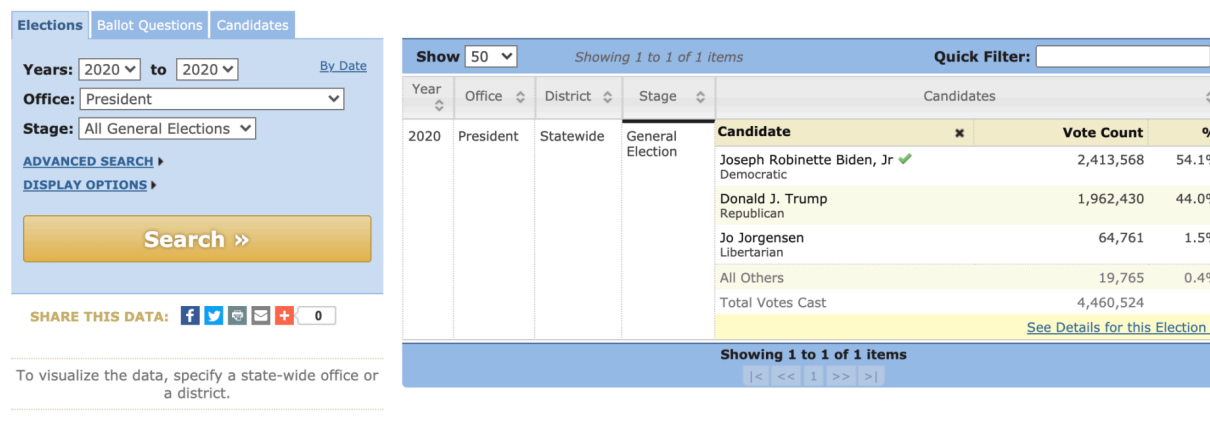
Figure 2: Screen capture (2021-08-10) from https://historical.elections.virginia.gov/elections/search/
year_from:2020/year_to:2020/office_id:1/stage:General

arithmetic.  The actual turnout number using their reported data can be computed as 75.08% as shown in the callouts of figure 1.  The difference in turnout percentage is 6.4%, or the equivalent of approximately **382,443** votes.

The Virginia DoE also publishes the 2020 General Election results (https://apps.elections.virginia.gov/SBE_CSV/ELECTIONS/ELECTIONRESULTS/2020/2020%20November%20General.csv) as a csv file. One would expect the data in the csv and the summary results presented on the VA DoE website to match, or at least be close. However, the data on the CSV file and the Virginia DoE website do not match, and in a peculiar way.  The sum of the values from the 2020 November General csv file, which includes In-Person, Absentee and Provisional votes gives the following results:

```
>> TotalTrumpVotes = 1962430

>> TotalBidenVotes = 2413568

>> TotalJorgensenVotes = 64761

>> TotalWriteInVotes = 21841

>> TotalVotes = 4462600

>> TotalInPersonVotes = 1630833

>> TotalAbsentee = 2831767
```

The official reported data from Figure 1 shows the Absentee vote being 2,687,304. Yet the "2020 November General.csv" file sum shows 2,831,767 absentee votes. Thats a difference of **144,463** votes.

The summed csv file totals for each candidate also exactly matches the reported data in figure 2, save for a **2,076** vote discrepancy in Write-In/Other.

In response to a FOIA request, the VA Dept of Elections supplied a summary file of the counts of all requested and returned absentee ballots per county.  The following file was provided by ashley.coles@elections.virginia.gov.

https://wwrkds.net/wp2/wp-content/uploads/2021/08/OSD-1422-1.xlsx

The noted file contains two data columns: the Total Requested Absentee Ballots and Total Returned Absentee Ballots.

The Total Returned column produces a result of **2,829,037** voted Absentee Ballots, a difference of **2730** votes (< 1%) from the CSV file, but a difference of **141,733** from the certified results in Figure 1.

These four reports are four official sources of Virginia Election data, but the four sources reveal different totals. The discrepancies between the CSV and the data per the FOIA request are a minimal difference (2730, <1%)  but a comparison of the certified election absentee data versus the CSV file and the FOIA information shows discrepancies of **144,463** (5.4%) and **141,733** (5.3%). These latter discrepancies are material differences, and warrant further investigation.

Figure 3 presents the discrepancies noted above in the various Virginia DoE reports, as well as others, into a summary table.

| | VA DoE Certified Results | Biden Vote | Trump Vote | Total Voting | Total Registered | % Change from previous year | (Reported) Turnout (Total Voting / Total Registered) | (Computed) Turnout | Voting Absentee |
|---|---|---|---|---|---|---|---|---|---|
| #1 | VA DoE Report: Registration / Turnout | | | 4,486,821 | 5,975,676 | 6.18% | 81.48% | 75.08% (~382,443 Votes) | 2,687,304 |
| #2 | VA DoE Elections Database | 2,413,568 | 1,962,430 | 4,460,524 | | | | | |
| #3 | CSV: 2020-November-General | 2,413,568 | 1,962,430 | 4,462,600 | | | | | 2,831,767 |
| #4 | CSV: Registrations 11-01-2020 | | | | 5,975,717 | 6.18% | | | |
| #5 | FOIA'd CSV | | | | | | | | 2,829,037 |
| #6 | Daily Absentee List (by Request) | | | | | | | | 2,826,484 |
| | Computed from #3 & #4 | | | | | | 74.68% | 74.68% | |
| | Max Abs Difference | 0 | 0 | 24,221 | 41 | 0 | 6.8% | 0.4% | 144,463 |

Figure 3: Inconsistencies in VA Dept of Elections Provided Data

## 2. Election Fingerprint Analysis

### a.  Background

The US National Academy of Sciences (NAS) published a paper in 2012 titled "Statistical detection of systematic election irregularities." [1]  The paper asked the question, "How can it be distinguished whether an election outcome represents the will of the people or the will of the counters?" The study reviewed the results from elections in Russia and other countries, where widespread fraud was suspected. The study was published in the proceedings of the National Academy of Sciences as well as referenced in multiple election guides by USAID [2][3], among other citations.

The study authors' thesis was that with a large sample of the voting data, they would be able to see whether or not voting patterns deviated from the voting patterns of elections where there was no fraud. The results of their study proved that there were indeed significant deviations from the expected, normal voting patterns in the elections where fraud was suspected.

Statistical results are often graphed, to provide a visual representation of how normal data should look. A particularly useful visual representation of election data is the *election fingerprint*. When used to analyze election data, the election fingerprint typically analyzes the votes for the winner versus voter turnout by voting district. The expected shape of the fingerprint is of that of a 2D Gaussian (a.k.a. "Normal") distribution [4].

See this MIT News article for a great description and primer on the "Gaussian" or "Normal" distribution: https://news.mit.edu/2012/explained-sigma-0209.  From the Article:

> "In many situations, the results of an experiment follow what is called a 'normal distribution'. For example, if you flip a coin 100 times and count how many times it comes up heads, the average result will be 50. But if you do this test 100 times, most of the results will be close to 50, but not exactly. You'll get almost as many cases with 49, or 51. You'll get quite a few 45s or 55s, but almost no 20s or 80s. If you plot your 100 tests on a graph, you'll get a well-known shape called a bell curve that's highest in the middle and tapers off on either side. That is a normal distribution."

In a free and fair election, the plotted graphs of both the Turnout percentage and the Vote Share for Election Winner should both resemble Gaussian "Normal" distributions; their combined distribution should also follow a Gaussian or normal distribution. Computing this 2 Dimensional joint distribution of the Turnout vs. Vote Share is what we refer to as an "Election Fingerprint".

Figure 4 is reprinted examples from the referenced National Academy of Sciences paper:

The actual election results in Russia, Uganda and Switzerland appear in the left column, the right column is the modeled expected appearance in a fair election with little fraud, and the middle column is the researchers' model of the as-collected data, with any possible fraud mechanisms included.

As you can see, the election in Switzerland (assumed fair) shows a range of voter turnout, from approximately 30 - 70% across voting districts, and a similar range of votes for the winner.  The Switzerland data is consistent across models, and does not show any significant irregularities.
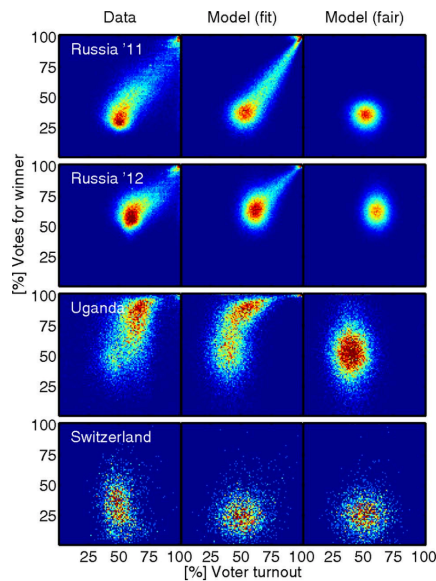
Figure 4: NAS Paper Results (reprinted from [1])

What do the clusters mean in the Russia 2011 and 2012 elections? Of particular concern are the top right corners, showing nearly 100% turnout of voters, and nearly 100% of them voted for the winner.

Both of those events (more than 90% of registered voters turning out to vote and more than 90% of the voters voting for the winner) are statistically improbable, even for very contested elections. Election results that show a strong linear streak away from the main fingerprint lobe indicates 'ballot stuffing,' where ballots are added at a specific rate. Voter turnout over 100% indicates 'extreme fraud'. [1][5]

Election results with 'outliers' - results that fall outside of expected normal voting patterns - while evidentiary indicators, are not in and of themselves definitive proof of outright fraud. For example, in rare but extreme cases, where the electorate is very split and the split closely follows the geographic boundaries between voting precincts, we could see multiple overlapping Gaussian lobes in the 2D image. Even in that rare case, there should not be distinct structures visible in the election fingerprint, linear streaks, overly skewed or smeared distributions, or exceedingly high turnout or vote share percentages.  Additional reviews of voting patterns and election results should be conducted whenever deviations from normal patterns occur in an election.  Additionally it should be noted that "the absence of evidence is not the evidence of absence":  Election Fingerprints that look otherwise normal might still have underlying issues that are not readily apparent with this view of the data.

## b. The US Virginia 2020 Election Fingerprint, Per County Congressional District

The upper right image in Figure 5 is the computed election fingerprint, computed according to the NAS paper and using official state reported voter turnout and votes for the statewide winner and reported per County Congressional District (CCD) with combined InPerson, Absentee and Provisional votes. The color scale moves from precincts with low counts as deep blue, to precincts with high numbers represented as bright yellow. Note that a small blurring filter was applied to the computed image for ease of viewing small isolated CCD results.

The upper right image was computed per the NAS paper; the bottom left image shows what an idealized model of the data could or should look like, based on the reported voter turnout and vote share for the winner.  This ideal model is allowed to have up to 3 Gaussian lobes based on the peak locations and standard deviations in the reported Virginia results.

The top-left and bottom-right plots show the sum of the rows and columns of the fingerprint image. The top-left graph corresponds to the sum of the rows in the upper right image and is the histogram of the vote share for Biden across precincts. The bottom right graph shows the sum of the columns of the upper right image, and is the histogram of the percentage turnout across Virginia voting districts.
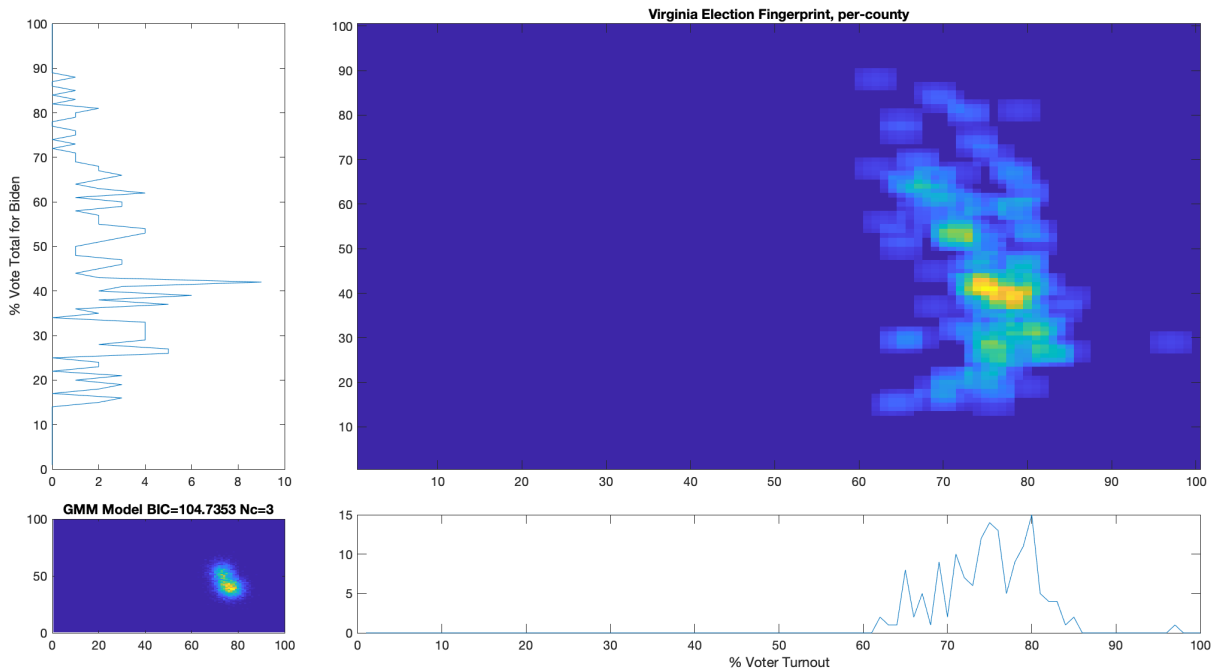
Figure 5 : Per County Congressional District Election Fingerprint
CSV-2020-November-General: https://apps.elections.virginia.gov/SBE_CSV/ELECTIONS/ELECTIONRESULTS/2020/
CSV-Registrations-2020-11-01: https://www.elections.virginia.gov/media/registration-statistics/2020/10/
Daily_Registrant_Count_By_Locality_2020_11_01_054008.csv

**Observations / Conclusions from Fingerprint:**

- There is significant correlated "streaking", distinct structural patterns and deviations from the expected 2D Gaussian distribution.  The fingerprint is, by definition, irregular and does not correspond with the accepted theory of what would be expected in a free and fair election.

- The "Galax City : 09" CCD is a distinct and singular outlier, and has a turnout of 96.67% which is very high (but not impossible) even for contested elections.

- There are a number of CCD's with Vote Share for Biden at or near the 90% mark, which is also very high (but also not impossible) even for contested elections.

- Election results with 'outliers' - results that fall outside of normal voting patterns - are not in and of themselves definitive proof of fraud. But additional reviews of voting patterns and election results should be conducted whenever deviations from normal patterns occur in an election.

## 3. No mathematical way to Reconcile Per-County and Per-Precinct data sets

In theory, each individual vote, whether in person, absentee, or provisional ballot, should be traceable to unique voters in each voting precinct.  Such a mapping would be useful in order to attempt to generate a higher resolution Election Fingerprint image, with each histogram datapoint representing an individual precinct instead of a County Congressional District (CCD).

Virginia uses virtual absentee and provisional ballot counting precincts, and collects all absentee and provisional ballots at the county congressional district level. Some counties or large cities will span multiple congressional districts.  The in-Person ballots on election day are collected and reported at precincts within each CCD, and the absentee and provisional ballots are collected at Central Absentee Precincts (CAPs), one per CCD.  As a result, absentee and provisional ballots cannot be commonly reconciled back to the precincts to which they belong.

If the reported results from the Virginia Department of Elections are accurate, the votes for each candidate attributed to absentee counting district entries in the VA published "2020 November General.csv" file should be able to be "un-aggregated" and attributed back to the real precincts that they originated from without violating any of the known restrictions based on registered voters, total absentee ballots received and approved, etc.  However, important information is missing from the VA reported results, information which is necessary to directly "un-aggregate" the absentee data.

Mathematically, if the Virginia reported datasets are consistent and valid, there _must exist at least one_ such set of possible "un-aggregated" results, if not many. If no such set can be found, then the reported data from the Virginia DoE is by definition inconsistent and mathematically invalid.

Additional calculations were undertaken to discover a solution that satisfies the per precinct per congressional district totals of absentee, provisional and returned ballots allocated to Central Absentee Precincts. In computer science and mathematical theory, this is known as a "multiple constrained subset-sum" problem, and standard techniques can be used to try and discover such solutions.

The assumptions for these calculations, using the certified Virginia Election data sources available, are as follows:

1. The total number of In-Person votes per precinct are taken directly from the published per precinct data and held fixed.

2. The total number of votes (In-Person, Absentee) per precinct cannot exceed the number of registered voters for the precinct.

3. The total number of Absentee votes allocated to a precinct should be equal to the total number of returned and accepted absentee ballots for the precinct, as computed from the number of accepted and counted unique absentee ballots in the Daily Absentee List (DAL) available from the VA DoE.

4. The sum of all vote un-aggregated totals (including attributed absentee ballots) for each candidate for all of the precincts in a given CCD should equal the original published aggregate numbers that used the aggregated CAPs for that CCD.

**Method and Results of Additional Calculations:**

Multiple randomized trials using a non-linear Levenberg-Marquardt optimization algorithm according to the constraints were attempted using all of the publicly available data from Virginia Department of Elections.

No solution was found using the constraints noted above.

Jonathan J. Lareau                                                 version 4

Estimates of the absentee votes per precinct were calculated, but none of the estimates could align with all of the data constraints noted above.  The "best bad" possible mathematical solution that was determined still violated the number of Absentee votes per precinct by over 15,000, The "worst optimal" solution violated by over 550,000.
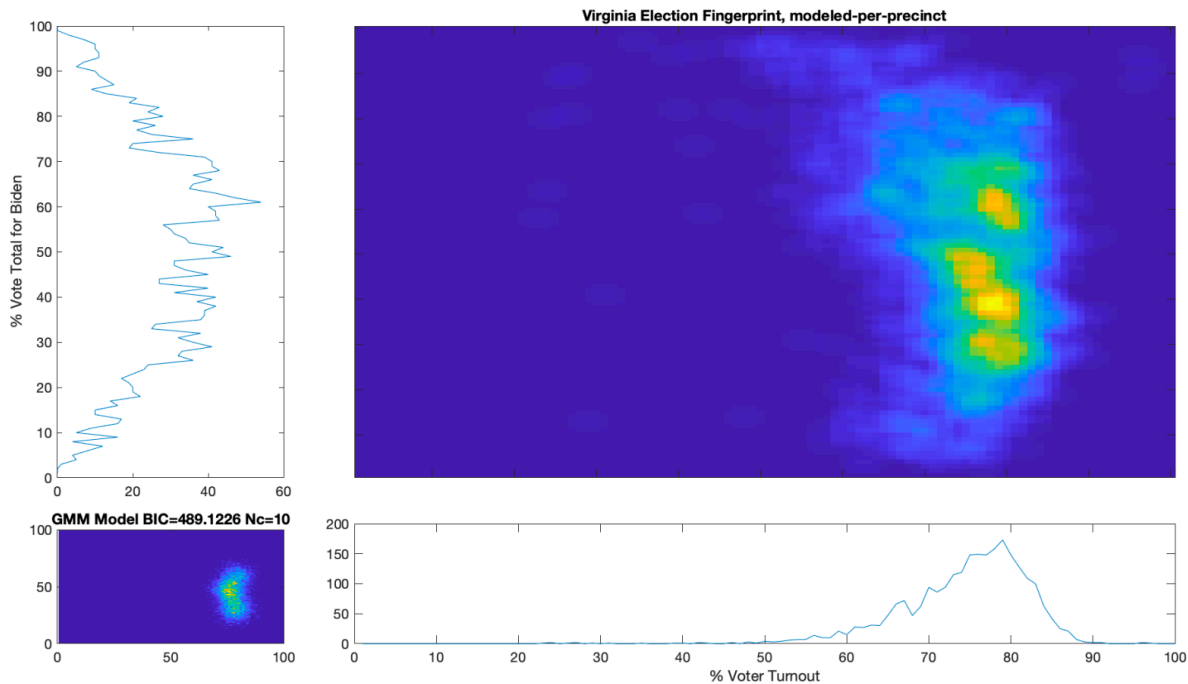


Figure 6: The "best bad" modeled solution to the absentee unwrapping problem.

## Final Conclusion:

◆ The clear discontinuities in reported data, the fingerprint analysis and additional calculations undertaken to attempt to replicate the commingled Central Absentee Precinct data indicate significant irregularities in the Virginia 2020 Election and reported results that warrant additional scrutiny and investigation.

◆ Additional information can be found at https://wwrkds.net/wp2/blog/

**Data sources:**

1. Registration / Turnout Report: https://www.elections.virginia.gov/resultsreports/registrationturnout-statistics/
2. Elections Database: https://historical.elections.virginia.gov/elections/search/year_from:2020/year_to:2020/office_id:1/stage:General
3. CSV-2020-November-General: https://apps.elections.virginia.gov/SBE_CSV/ELECTIONS/ELECTIONRESULTS/2020/2020%20November%20General.csv
4. CSV-Registrations-2020-11-01: https://www.elections.virginia.gov/media/registration-statistics/2020/10/Daily_Registrant_Count_By_Locality_2020_11_01_054008.csv
5. FOIA's CSV: https://wwrkds.net/wp2/wp-content/uploads/2021/08/OSD-1422-1.xlsx
6. DAL (as of 2020-11-30): https://www.elections.virginia.gov/candidatepac-info/client-services/index.html


**References:**

[1] "Statistical detection of election irregularities"
Peter Klimek, Yuri Yegorov, Rudolf Hanel, Stefan Thurner Proceedings of the National Academy of Sciences Oct 2012, 109 (41) 16469-16473; DOI: 10.1073/pnas.1210722109 (https://www.pnas.org/content/109/41/16469)

[2] USAID: Assessing and Verifying Election Results: A Decision Makers Guide to Parallel Vote Tabulation and Other Tools (http://web.archive.org/web/20201118021847/https://pdf.usaid.gov/pdf_docs/PA00KGWR.pdf)

[3] USAID: A guide to Election Forensics (http://web.archive.org/web/20210501091306/https://pdf.usaid.gov/pdf_docs/PA00MXR7.pdf)

[4] Multivariate Normal Distribution - Wikipedia (https://en.wikipedia.org/wiki/Multivariate_normal_distribution)

[5] Mebane, Walter R. and Kalinin, Kirill, Comparative Election Fraud Detection (2009). APSA 2009 Toronto Meeting Paper, Available at SSRN: https://ssrn.com/abstract=1450078