



From SA to the USA: Election forecasting

JP Holloway* HW Ittmann† N Dudeni-Tlhone‡ PMU Schmitz§

Received: 22 November 2017; Revised: 25 April 2018; Accepted: 11 May 2018

Abstract

Elections draw enormous interest worldwide, especially if these involve major countries, and there is much speculation in the media as to possible outcomes from these elections. In many of these recent elections, such as the UK and USA, however, forecasts from market surveys, electoral polls, scientific forecasting models and even exit polls, obtained from voters as they leave the voting stations, failed to predict the correct outcome. Election night forecasts, which endeavour to forecast the ultimate result before the final outcome is known using early results, were also carried out, with some more accurate than others.

After successfully predicting most of the metropolitan region results correctly in the South African local 2016 municipal elections, using an election night forecasting model developed for *South Africa* (SA), the question of adapting the model to work outside of SA on a different electoral system was raised. The focus of this paper is to describe the results obtained for the 2016 USA presidential election, on election night, using an adapted version of the SA model. This paper also addresses the applicability of the model assumptions as well as the data issues involved in forecasting outside of South Africa. It is shown that even with many hurdles experienced in the process the model performed relatively well.

Key words: Elections, forecasting

*Decision Support and Systems Analysis Research Group, CSIR Built Environment, Pretoria, South Africa, email: jhollowa@csir.co.za

†Corresponding author: (**Fellow of the Operations Research Society of South Africa**) *Institute for Logistics and Transport Studies* (ILTS) Africa, University of Johannesburg, Johannesburg, South Africa, email: hittmann01@gmail.com

‡Decision Support and Systems Analysis Research Group, CSIR Built Environment, Pretoria, South Africa, email: ndudenitlhone@csir.co.za

§Sustainable Human Settlements Research Group, CSIR Built Environment, Pretoria, South Africa; Department of Geography, UNISA, Florida, South Africa, email: pschmitz@csir.co.za

1 Introduction

Elections draw enormous interest worldwide especially if these involve major countries such as the United States of America (USA), the United Kingdom (UK), Germany and France. Months before *Election Day* there is much speculation about the outcome of an election. Predictions from market surveys and forecasts are published attempting to indicate what the outcome would be. Election forecasts are based on speculation, non-systematic interviews, *etc.* or scientifically based methods such as simulations, market analysis, sample surveys or statistical models [31]. The 2015 general election in the UK, Brexit, the 2016 USA presidential election, the 2017 presidential election in France as well as the 2017 general election in Germany, are all examples of elections that attracted huge attention. What made these different from previous elections is the fact that the forecasts done before the elections, in almost all of these elections, were wrong. The forecasts from market surveys, electoral polls, scientific forecasting models and even the exit polls, obtained from voters as they left the voting stations, failed to predict the correct outcome.

Over and above these approaches there is a further type of election forecasting model, namely one that is used to forecast the final election results, based on early results as they are released, during election night. These are referred to as election night forecasting models. These models have the benefit of being able to use actual released results and are attractive to the media in terms of being able to offer alternative and updated “live” predictions for discussion purposes.

South Africa (SA) is no exception when it comes to elections. There is huge interest in elections and market surveys are published regularly up to a day before the election. Exit polls, however, are banned under the 1998 Electoral Act [44]. Since 1999 an election forecasting team from the *Council for Scientific and Industrial Research* (CSIR) have consistently, during every election, used an election night forecasting model to forecast the final election outcome using the first batch of voting district results as they are released by the *Independent Electoral Commission* (IEC). This model has performed well except when certain model assumptions were violated, as in 2014 (see [27]). After successfully predicting most of the metro (metropolitan region) results correctly in the 2016 municipal elections, the team felt that the model could possibly be adapted to work outside of SA and decided to test the model on the 2016 USA presidential elections. The focus of this paper is to describe the results obtained for the 2016 USA presidential election, on election night, using an adapted version of a model developed for forecasting the final results for elections in SA.

In SA there is a window of opportunity for forecasting in the time period from when the first voting district results are declared until the final outcome is announced. For example, the first voting district results in the country during the 2016 municipal elections came in shortly after polls closed on the Wednesday night but the Johannesburg results were only finalised late on the Saturday afternoon. Due to this time delay, the election night forecasts become useful to the media as they can be discussed on radio and live TV during the respective special election broadcasts. When considering the USA presidential elections it was felt that there might be a similar window of opportunity (although smaller) for doing election night forecasts in the USA due to time zone differences. The question

was whether the model assumptions would hold outside of SA and whether appropriate data could be obtained.

The paper is structured as follows: first, other election predictions and forecasting models found in the literature are discussed briefly. Forecasts can be wrong and in this regard particular emphasis is placed on the recent USA presidential election where the outcome was totally different from most of the forecasts. It is shown that this was no exception considering the forecasts for two of the recent UK elections. Models used both before elections and during elections are considered. Some issues leading up to these elections are also touched on.

Secondly, a brief overview of the SA model's performance is presented, followed by a description of the differences between the SA and USA electoral systems. Data requirements and the adaption of the SA model are also provided and issues involved in forecasting outside of South Africa are discussed, especially with regard to data. The forecasts obtained for the USA elections during election night are then presented and the model assumptions analysed. Some overall remarks on election forecasting methodologies and suggestions on adapting the model to pre-election forecasts are also given and the paper is then concluded.

2 Review of forecasts for “unexpected outcome” elections

Prior to elections there are polls (surveys) that attempt to measure how the voting population will vote and there are forecasting models, modelling the behaviour of the voting population, all endeavouring to forecast election outcomes. In addition, there are efforts to predict the outcome after voting stations have closed, known as exit polls. Election night forecasts are also carried out to try to give an indication of what the ultimate result could be before the final outcome is known. In the case of exit polls, voters are asked how they voted as they exit the voting stations and this data is used to make projections of what the outcome could be. Election night forecasts on the other hand are endeavours to incorporate early results, as they are made known, to forecast the final election result. Both pre-election forecasts (including exit polls) and election night forecasts are discussed in this section, with particular emphasis given to the UK and USA elections. The 2017 French and German elections are briefly mentioned.

2.1 Pre-election forecasts

A number of recent international elections led to totally unexpected election results where the predictions beforehand turned out to be totally wrong. In the UK it prompted Prime Minister David Cameron to comment, a day after the May 2015 election, that “*pundits got it wrong, the pollsters got it wrong, the commentators got it wrong*” [45]. The outcomes of the 2015 UK election, the Brexit referendum and the 2016 USA presidential election were in every case against almost all expectations. The most recent 2017 presidential election in France was also totally unexpected, but for a different reason, as the French elected a young and relatively unknown president who ran a “*courageous pro-European campaign*” [6]. In the case of the 2017 German election the rise of the *Alternative für Deutschland* (AfD) party was a total surprise as well as the large decrease in support for the two

main parties, the *Christian Democratic Union* (CDU) and the *Sozialdemokratische Partei Deutschlands* (SPD) [5].

2.1.1 The 2015 UK general elections

In 2015 in the UK, the polls and the more sophisticated forecasting models consistently forecasted that no party would get an outright majority (see for example [30] and the forecasts given in Table 1) since an outright majority required 326 seats in 2015.

Party	Forecast		Final Results	
	Seats	%Votes	Seats	%Votes
Conservatives	278	34.4	331	36.9
Labour	267	32.8	232	30.4
SNP	53	4.0	56	4.7
Lib. Democrats	27	11.7	8	7.9
DUP	8		8	0.6
Plaid Cymru	4	0.6	3	0.6
SDLP	3		3	0.3
UKIP	1	10.6	1	12.6
Greens	1	4.1	1	3.8
Other	8	1.7	7	2.2

Table 1: 2015 UK general elections – Forecast [21] and Result [3].

Compared to the final results, in terms of seats, for almost all of the parties these “before election” forecasts (Table 1) were incorrect by a large margin¹. After the election one UK polling company, Survation, claimed that their forecast was almost correct but because it was so different from all the others they decided not to publish it [23]! The Survation forecast was indeed almost spot on.

Cowley & Kavanagh [13] indicated that there were many more polls than during previous elections and since these consistently showed the possibility of a hung parliament it led “to an explosion of coalitionology” in the media. The media debated hung parliaments, possible coalitions, constitutional process and the “legitimacy” of various outcomes, *etc.* All this speculation in the media had a huge influence on the various forecasting models used during that election.

2.1.2 United Kingdom European Union membership referendum, 2016 (Brexit)

The Brexit referendum took place on 23 June 2016 in the UK and Gibraltar (a British Overseas Territory). The aim was to gauge support for the UK remaining a member of the *European Union* (EU) or leaving the EU. The outcome of the referendum was that 51.9% of voters voted in favour of leaving the EU (17 410 742 votes) while 48.1% voted against leaving the EU (16 141 241 votes) [4]. There were 0.08% invalid ballots (25 359 votes).

¹The authors observed a significant difference ($p=0.006$) when using a chi-square test on the data in Table 1.

The 2016 Brexit result was again a surprise and an unexpected outcome, since it was expected that most voters would vote to remain in the EU (see for example [58]). An exit poll on the day of the referendum by YouGov [59] found some correlation between voting preference and age as well as turnout and age (see Figure 1). This could explain the reason for the unexpected outcome. A total of 4 772 voters were interviewed.

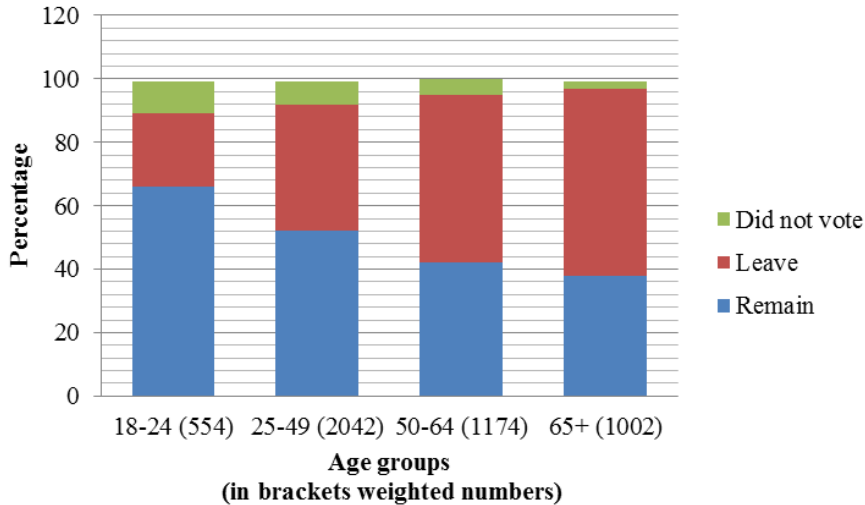


Figure 1: EU referendum exit polls by age group, June 23, 2016 [59].

It is clear that the older generation (50 and older) mostly favoured leaving the EU while the younger generation were happy for the UK to remain. A high proportion of the younger generation did not vote compared to those from the older generation(s). Analysing campaign and survey data, Hobolt [24] shows that globalisation was an issue. In addition those favouring leaving the EU were particularly common among less-educated, poorer and older voters as well as those who expressed concerns about immigration and multiculturalism.

2.1.3 The 2016 United States of America presidential election

The most recent USA presidential election will certainly go down in history as very different to any other election in the past. With two “controversial” candidates, generating a lot of contention even before the election started, the 2016 election was unprecedented in nature, creating even more interest in all the election forecasts. In the end very few forecasters got it right!

Nate Silver, a well-known forecaster, was incredibly successful in the 2012 elections [52]. In that election, the FiveThirtyEight website, where Silver’s forecast was published, never had Obama’s forecasted chances of winning at less than 61.1%. This time around Silver had it all wrong [49]! Silver [50] also indicates that: “to build a model is to measure uncertainty and to account for risk”. The model Silver used is a probability based election forecasting model developed in 2010 [18]. The model makes use of various data sources but relies mainly on the aggregation of pre-election polls using a technique developed by Brown *et al.* [8]. Aggregated poll data are adjusted based on current demographic patterns via

a regression model and simulations are run to quantify both the national and state-level uncertainty around the election outcome. Specifically, a Monte Carlo simulation method is used to draw random samples forming a matrix of, for instance, 51 states and 10 000 observations representing a possible result of the election or uncertainty. Each candidate's share of votes is therefore expressed as a joint distribution with these thousands of simulations representing forecasts of uncertainty about the election outcome. This methodology is applied to estimate both the popular and electoral vote shares for each candidate. The model was, however, not able to predict Trump's unexpected win during the 2016 presidential elections. The bad performance of this model was attributed to imperfect polling data and the non-uniform structure of errors in the poll-based forecast which had not been perceived before. Silver does, however, point out that FiveThirtyEight's forecast gave Trump much better odds, a 29% chance of winning the Electoral College, than any of the other polling-based models. On further reflection Silver [51] addresses the myth that Trump's victory represented some sort of catastrophic failure for the polls. He points out that in the case of the popular vote the polls were out by only 1 to 2 percentage points nationally, predicting correctly that Trump would lose to Clinton, which was slightly more accurate than the predictions in 2012. It is in the swing states and especially in Michigan, Wisconsin and Pennsylvania, where Trump did much better than what the polls forecasted. He then makes the interesting observation that the result was not a massive outlier and points out that the polls were pretty much as accurate as they'd been, on average, since 1968.

Samuelson [45] states that for this USA election only three or four forecasters correctly predicted the outcome. Most of these were not strongly quantitatively based. Lichtman [32], a quantitative historian, has used a "13 Keys" model since 1984 to correctly forecast the USA presidential election. While Lichtman was able to forecast the popular vote correctly in previous elections, he was only able to forecast the electoral vote correctly in 2016.

Zogby [60], a successful pollster for many years in the USA, realised early on the difficulty of making an accurate call on the election. During a presentation he stated: *"I can't tell you who's really going to win. Tell me who will vote, and I'll tell you who will win. If we get around 132 million votes, as we did in 2008 and 2012, Hillary wins. If we get 121 million, as we did in 2004, Trump wins."* In the end the turnout was 126 million, giving Clinton about a 2 million vote lead. Several million non-voters, particularly Democratic-leaning people, in the north central states of the USA helped tip those states to Trump.

A film director, Michael Moore [35], wrote on his website in July 2016 (three months before the elections) that he expected Trump to win the election. He gave five reasons to motivate his "speculation" about this election prediction. The main reason, according to Moore, were the angry and frustrated white male workers, mainly union workers, in the Rust Belt of the Midwest of the US. One or two of the other reasons was "the Hillary problem" and Moore's belief that "people will vote for Trump because they can".

The media, as in the 1960 Kennedy and Nixon television debates, again played a role [34]. Clinton was perceived through media channels by many voters as a "corporate lawyer turned politician", while Trump's peculiar ways during visual engagements intrigued voters, thus making Trump the preferred choice of many voters.

Finally an artificial intelligence system, called MogAI [17], predicted a Trump victory. The system uses 20 million data points from online platforms like Google, YouTube, and Twitter to come up with its predictions. The system had correctly predicted three previous presidential elections. As an AI system, MogAI learns from its environment, develops its own rules at the policy layer and then develops expert systems without discarding any data. In this regard it is generally accepted that President Obama was the first presidential candidate that used social media in a big way and very successfully so [12].

The presidential election shocked many in the USA and worldwide with the most unlikely candidate, according to the majority of people, winning the election. This was a direct result of changes in the winning party in many of the reported swing states, which can be seen when comparing the outcomes in these states between the 2012 and 2016 USA presidential elections, as shown in Table 2 and Figure 2. Note that swing states, also known as “battleground” or “purple states”, are constituent political entities which are closely divided between the Democratic and Republican parties in the US electoral system; such that the candidates contesting an election both have a good chance of winning the presidency [16]. They include several states that have gone both ways in the recent elections and are considered important to the outcome of the election [22].

States	2016		2012		2008		2004	
	Swing state?	Won by	Swing state?	Won by	Swing state?	Won by	Swing state?	Won by
Colorado	✓	Dem	✓	Dem	✓	Dem	✗	Rep
Florida	✓	Rep	✓	Dem	✓	Dem	✓	Rep
Iowa	✓	Rep	✓	Dem	✓	Dem	✓	Rep
Michigan	✓	Rep	✓	Dem	✗	Dem	✓	Dem
Minnesota	✓	Dem	✗	Dem	✗	Dem	✗	Dem
Nevada	✓	Dem	✓	Dem	✓	Dem	✓	Rep
New Hampshire	✓	Dem	✓	Dem	✓	Dem	✓	Dem
North Carolina	✓	Rep	✓	Rep	✓	Dem	✗	Rep
Ohio	✓	Rep	✓	Dem	✓	Dem	✓	Rep
Pennsylvania	✓	Rep	✓	Dem	✗	Dem	✓	Dem
Virginia	✓	Dem	✓	Dem	✓	Dem	✓	Rep
Wisconsin	✓	Rep	✓	Dem	✓	Dem	✓	Dem

Table 2: Reported swing states in the 2016 USA presidential elections and their previous expectations as swing states and their outcomes. Only states appearing in more than one source were included for 2016 [7, 29, 33, 42, 48], only states appearing in more than two sources were included for 2012 [11, 15, 40, 43], only states appearing in more than two sources were included for 2008 [1, 15, 28, 53, 56] and only states appearing in more than two sources were included for 2004 [15, 39, 47].

2.2 Election night forecasts

Election night forecasting models generally become a source of insightful information during election coverage in the media while the election outcome is still unknown. Countries including the UK [8, 9, 37, 38], USA [2, 10], Spain [37], South Africa [19, 20, 27], New Zealand [36, 37], Australia [37], Austria [25], Sweden [37] and Ireland [37] have, across

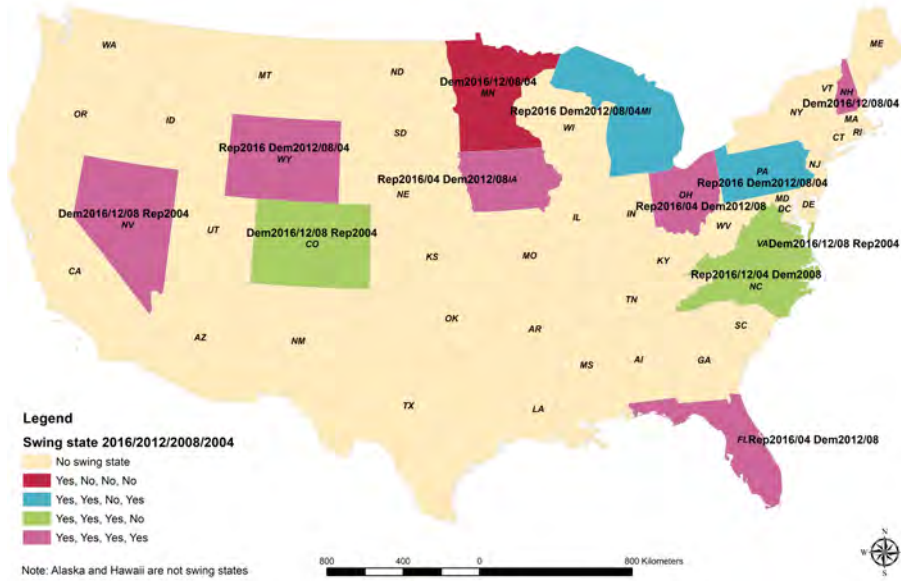


Figure 2: Map of changes in swing states between 2004 and 2016 USA presidential elections.

a number of elections, made use of election night forecasting models to infer an election outcome during hours (and sometimes days) of uncertainty around the final outcome of an election. These models seek to provide an accurate forecast of an election outcome from a small but reasonable sample of the released results. Hence they are considered successful if they can correctly predict the outcome of an election while there is still considerable uncertainty around such an outcome. The public interest is then still high. This subsection particularly focusses on discussing a published UK election night model, as well as some online election night forecasts for the recent USA presidential elections.

2.2.1 The UK general elections

The UK has been using election night forecasting for about 4 decades (since 1975 elections) from which models have been continuously updated to improve the prediction accuracy across relevant election periods. One such model was first developed by Brown and Payne [9] for the BBC during the 1974 British general elections, and has been modified over time to improve forecasting of subsequent elections [8]. This model is based on grouping constituencies into three groups and applying ridge regression models on a set of variables to obtain coefficients from which to forecast the share of votes (seats) for predominant parties as well as the swing in party votes. Priors from ridge regression coefficients, which can be used to represent the mean share of votes in the various seat categories, are created using exit poll data. During election night, these ridge regression coefficients are updated by sequentially adding actual voting results, as they become available, together with other relevant variables, which could include socio-economic variables and the previous election share of votes for each party. This model performed reasonably well for a number of

election years except for the 1987 and 1992 elections. The authors attributed the poor performance to the Gallup judgemental poll-based prior which was used as a new input into the model during these years. Also, instead of the ridge regression that they had previously used, a swingometer-based method was applied. An introduction of a separate grouping of 62 special seats, its estimation from prior fixed probabilities, as well as late declaration of such seats were also among the issues that increased the prediction errors during the 1992 elections. Opinion polls and exit polls both forecasted an overall majority for the Labour party in 1992 when in fact the Conservative party won by a small majority. The election night model was slow to converge to the final outcome, as shown by the absolute values of the errors in Table 3. Note that a total of 651 seats were contested in the 1992 elections and 326 seats were needed for a majority [54].

Election	Number of seats declared						Winner	Actual majority
	0	5	10	50	100	300		
1974 (Feb.)	6	2	16	12	4	2	Lab	4
1974 (Oct.)	28	13	15	7	8	4	Lab	42
1979	4	13	2	0	9	3	Con	70
1983	2	3	7	10	16	6	Con	188
1987	70	39	46	20	20	5	Con	147
1992	62	49	40	30	36	10	Con	65
1997	25	10	15	5	18	3	Lab	254
2001	6	8	5	10	6	4	Lab	253

Table 3: BBC results-based forecasts for general elections 1974–2001: Errors in majority of the winning party over the second party by number of seats declared [38].

Brown *et al.* [8] also mentioned the difficulty in accurately forecasting a closely fought election, such as in 1992. Modifications to this model were made, in addition to reverting back to the pure regression approach, to produce improved forecasts for the 1997 and 2001 elections [8, 38]. According to Payne [38], however, it is unclear whether the latter more accurate forecasts were due to the improved model or due to these elections being landslide victories, and indicates that the real test would be in a future closely fought election. Unfortunately, forecasts from this model, or similar election night models, for the unexpected outcome of the 2015 elections, discussed in Subsection 2.1.1, could not be found.

2.2.2 The 2016 United States of America presidential election

With respect to the USA election night forecasting, the focus is on two models, the one by FiveThirtyEight [10] and the other of Aisch *et al.* [2], applied in the prediction of the presidential candidate for the recent 2016 elections. The first model available on FiveThirtyEight [10] is based on firstly setting up pre-election forecasts using polling data (see Subsection 2.1.3 for details) and then updating the forecasts as the states get called for presidential candidates. The pre-election models, which are developed from simulating election results by aggregating poll data, are updated by applying several regression models to establish a correlation structure between the states in the Electoral College. The correlation patterns between the states play an important role as this enables the model

to make predictions about the outcome of the uncalled states from those whose results have been announced. Such patterns also make it possible to, for instance, forecast an unexpected win of a state by a candidate as soon as that candidate gets an unanticipated win from a similar state. Even though the strength of the relationships between states is essential, the forecast results are derived from simulation runs for different or unexpected scenarios that may arise. For instance, a candidate may win some states (as perhaps expected) but may lose some from the same group of correlated states, which provides some room for the unexpected or changing patterns. The focus of this model is mainly on forecasting the electoral votes for the party candidates and the probabilities associated with either winning or losing an election.

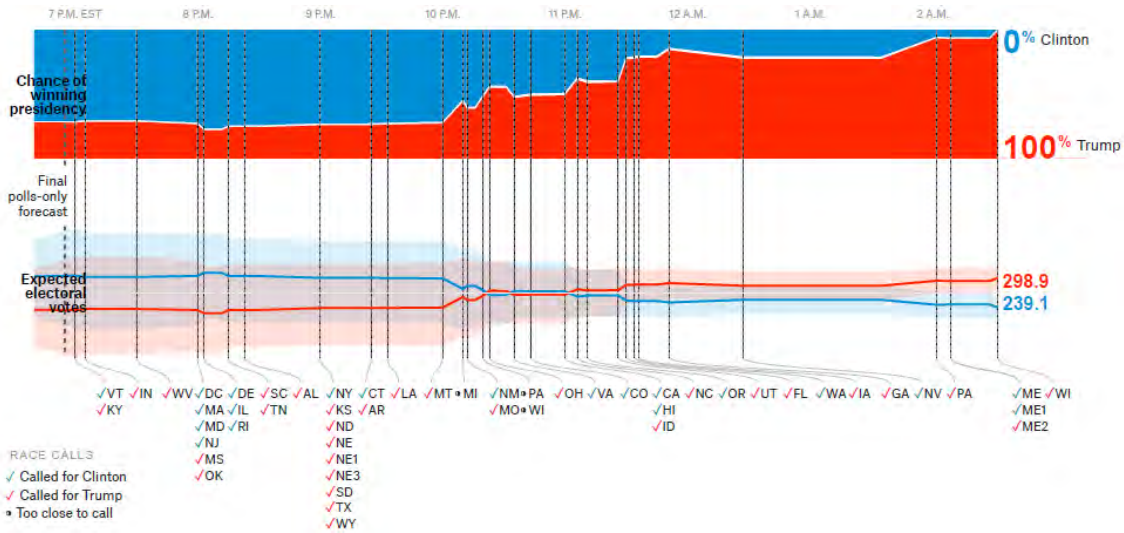


Figure 3: Election night forecasts during the live release of the 2016 USA presidential elections [10].

In Figure 3, the forecasts with respect to the probability of winning the presidency and the share of the electoral votes are shown, including a list of the declared states from which such forecasts were made at the given time intervals. The forecasts were in favour of Clinton from the onset (including the pre-election forecast) until only around 10:23pm Eastern Time.

The election night forecasting model developed by Aisch *et al.* [2] uses historical voting patterns of the states, county-level demographic characteristics and turnout estimates as inputs. As the voting results get reported from respective counties, forecasts about the likely winning party candidate are made. The model forecasts an election outcome using three metrics which include estimating the *chance of winning* the presidency for the main party candidates, the *number of electoral votes*, as well as the *popular vote margin*.

Figure 4 shows the forecast of the total electoral votes and the associated changes over time for the two main party candidates. Uncertainties around the forecasts are also estimated. From the onset, the Democratic Party candidate, Clinton, had a higher chance of winning the presidency, obtaining just over 300 electoral votes while this all changed in favour of the Republican candidate, Trump, from before 6am New York local time. Trump’s predicted

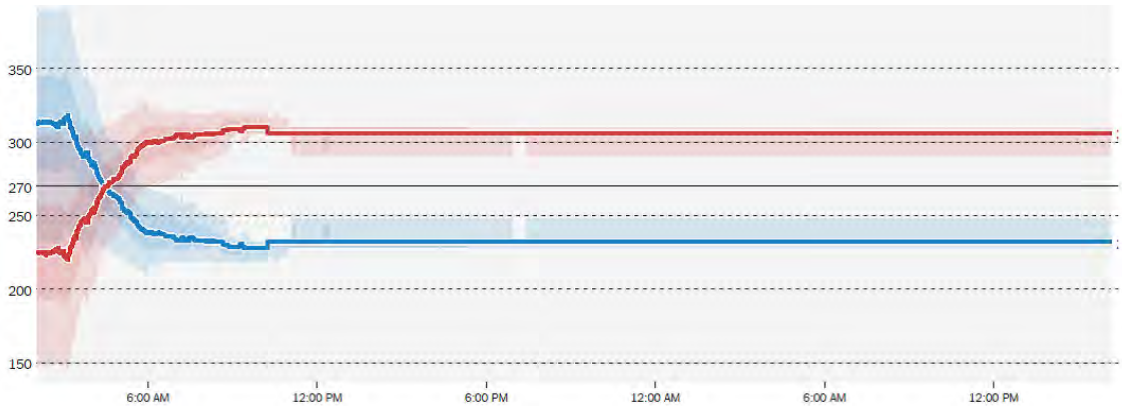


Figure 4: New York Times 2016 election night forecast of the total electoral votes [2].

win was then maintained throughout. Fluctuations or instability in the forecasts can be seen on the graph, particularly in the beginning when fewer counties had reported their results but such instabilities diminished over time.

3 SA model and its adaption to the USA

The focus of the CSIR team, since 1999, has always been on an election night forecasting model. In this section a brief outline is given of differences between the SA and USA electoral systems. An overview of the methodology of the SA model is presented and how it was adapted to fit the USA presidential elections. The model performance in the most recent 2016 South African municipal elections is discussed together with the data needs and requirements. Challenges faced are also discussed in this section.

3.1 Differences between SA and USA electoral systems

In South Africa the vote in the national and provincial elections is for a political party. The party will receive a proportion of the seats available in parliament based on the number of votes garnered during an election. This is commonly known as the *proportional representation* (PR) voting system. However, this system changes during local government elections, namely that voters cast two votes, one vote for the party and one vote for a ward councillor. The first vote is a PR vote, where the party will receive a certain amount of seats, and the rest are allocated to ward councillors [26]. The election model only forecasts the PR vote during a local government election.

In an outline of the USA presidential election process [57] an infographic is given on the process followed by a presidential hopeful in becoming the president of the United States. The process before the actual election of the president is outside the scope of this paper as this paper focuses on the election itself. The president and the vice-president get elected via the Electoral College and not by the popular vote. The number of Electoral College votes is based on the number of members of Congress that has been allocated to each state. Forty-eight states, as well as the District of Columbia (which is assigned its own Electoral

College votes), have a winner-takes-all policy and therefore even if the vote is 50.1% to 49.9%, the party candidate with 50.1% takes all the Electoral College votes for that state. Nebraska and Maine work more on a PR system, with one Electoral College vote for each congressional district and two for the overall state result. In special circumstances, as in the 2016 US presidential election, a candidate may win the popular vote but not have enough Electoral College votes to become the new president [57].

3.2 Overview of methodology

The mathematical details of the model developed for SA can be found in Greben *et al.* [19] and Greben *et al.* [20] but a very simplified high level overview is given here. The model works primarily on creating sets of fuzzy clusters at a selected spatial aggregation level and in SA this is usually 20 clusters per province and/or 20 clusters for the entire nation. The clusters are created using the voting behaviour (election results) of all *voting districts* (VDs) from a previous election, after these results have been adjusted for some VD demarcation changes. The use of fuzzy clustering ensures that each VD has a membership in each of the clusters of its spatially associated set of clusters, for example, the set of clusters associated with the VD's province. After clustering has taken place, the previous election results are discarded and only the cluster memberships and number of registered voters per VD are retained. On election night, once the results from a sample of results from counted VDs becomes available in the database, the model then uses these results (valid vote count and percentage per party), together with the VDs cluster memberships and registered voter count, to predict the party percentages and turnouts per cluster. These cluster predictions are in turn used to predict the uncounted VDs. Finally the known counted VD results and the predicted uncounted VD results are aggregated together to provide an overall prediction at the required spatial aggregation unit (municipality/province/nation).

Since the South African model merely works on the PR system using total vote counts, analogous to what is referred to as the popular vote in the USA, the model had to be adapted to allow for the Electoral College vote count per state (see Subsection 3.1). In the model, the predicted party percentages per state had to be calculated and the associated number of electoral votes per state then had to be assigned to the party with the highest predicted percentage in that state. These electoral votes for the predicted winner in each state then had to be combined to calculate the predicted national electoral vote count per party (presidential candidate) while still maintaining an overall national percentage to allow for the prediction of the popular vote.

3.3 Background on model performance in SA context

The original model was developed, by the CSIR, for the South African elections and has been applied during the 1999, 2004, 2009, 2014 general elections and the 2000, 2006, 2011, 2016 municipal elections. Up to this point it has only been used as an "election night" model which therefore only provides predictions of the final election results as counted and validated results from individual voting districts become available. Although initially contracted by the IEC to predict voting district results in order to help the IEC detect

any potential election fraud, it soon became clear that the model was also able to predict the aggregated results of the elections with a great deal of accuracy at a very early stage during the vote count. These predictions were particularly useful to the South African media due to the long time delays between the time that the first results get reported until the final counts were in, often spanning a few days. Consequently, from 2004 onwards, the *South African Broadcasting Corporation* (SABC) contracted the CSIR team to provide prediction results that could be discussed on live TV and on radio during their special election broadcasts.

This election night model worked well in the South African context, except when certain model assumptions were violated, as in 2014 (see [27]). A summary of most of the results up to the 2014 elections, as well as issues regarding the model assumptions, can be found in Ittmann *et al.* [27]. Typically the prediction results at a municipal level are less accurate than those at a national or provincial level since one is working with a smaller sample of voting districts, but the findings presented in Ittmann *et al.* [27] were used to make improvements to the 2016 municipal election clusters with the result that better prediction accuracy was achieved. Overall, the 2016 municipal predictions for the metros' PR votes proved to be a great success with the model accurately predicting that coalition governments would take power in the Johannesburg, Tshwane (Pretoria), Ekurhuleni and Nelson Mandela Bay (Port Elizabeth) metros; that the DA would be the largest party in the latter metro; that the *Democratic Alliance* (DA) would consolidate their control over Cape Town; and that the *African National Congress* (ANC) would retain control over eThekweni (Durban), Buffalo City (East London) and Mangaung (Bloemfontein). While the Nelson Mandela Bay prediction was not unexpected, even though the DA thought they might get an outright majority, the prediction of the ANC not getting an outright majority in Ekurhuleni was a complete surprise.

With voting booths having closed late (7pm), on Wednesday 3rd of August 2016, the model was able to predict, at a very early stage on Thursday morning 4th of August (5am), when there was still a great deal of uncertainty about where things would end, the final vote share for all the major parties in the metros of interest to within 1.7%. In the case of Tshwane, however, the model did not pick up until very late in the counting process that the DA would end up winning the largest share of the metro, although this winning margin was extremely small. The most impressive early prediction was for eThekweni – even though no voting district results had been reported from this municipality (due to technical issues) the model, using results from the rest of the province, determined not only that the ANC would retain its control of the metro but got within 0.3% of the final result for the top 4 parties (see Table 4). The prediction for Johannesburg was also very accurate, particularly since the first prediction released to the media at 5am on Thursday morning, after 13% of Johannesburg VDs had been declared, ended up being within 0.4% of the final result for both the ANC and DA, while the actual counts kept showing a very different result until late on Saturday afternoon when the final count was declared (see Table 4).

Metro (Metropolitan region)	Party	Predicted PR % (at 5am on Thursday)	Actual PR % (at 5am on Thursday)	Final PR %
eThekweni (Durban)	ANC	58.8	0	59.1
	DA	27.7	0	27.5
	EFF	3.5	0	3.6
	IFP	4.0	0	4.3
Johannesburg	ANC	44.5	39.5	44.9
	DA	38.9	45.3	38.5
	EFF	10.7	9.8	10.9
Tshwane (Pretoria)	ANC	42.8	44.2	41.5
	DA	41.5	43.0	43.1
	EFF	10.7	8.6	11.6
Ekurhuleni	ANC	47.8	37.2	48.8
	DA	36.0	50.8	34.1
	EFF	10.7	7.1	11.1
Nelson Mandela Bay (Port Elizabeth)	ANC	42.2	31.4	41.5
	DA	48.4	59.8	46.7
	EFF	3.9	3.8	5.0
Cape Town	ANC	25.1	18.8	24.5
	DA	65.7	72.4	66.8
	EFF	2.9	3.1	3.1

Table 4: Comparison of actual results and forecasts for some metros in the 2016 South African local elections – using PR vote only.

3.4 Data requirements of the model

For the South Africa elections the model uses the ID (Identification code assigned by the IEC) of the lowest spatial unit at which results are obtained as its primary key to link all data files and the data required prior to the elections are the following:

- Past results containing percentage per party per spatial unit (used for clustering);
- Registered voter population for new elections per spatial unit;
- Spatial aggregation data – linking lowest spatial unit to its higher spatial units, *e.g.* province; and
- List of parties contesting new elections and associated party IDs or codes that will be used when new data is released.

For the USA presidential elections the above data were also required for the selected spatial unit (see Subsection 3.5 on the selection of the spatial unit), with the province being replaced by state and the parties being reduced to only 3, namely; Democratic, Republican and “Other”. The “Other” was an aggregation of all smaller parties since the decision was made to only focus on the predictions for the two major parties. There was, however, the additional requirement for data on the number of electoral votes per state (and per congressional district for Maine and Nebraska). Note that the District of Columbia had to be treated as a county and a state, and therefore only a single data point (spatial unit) was available for this proxy state.

On election night, in addition to the prepared data files discussed above, the model required

regularly updated data for every spatial unit that had been counted and released. These data per spatial unit had to contain, in a text format, the total number of valid votes counted and the percentage per party ID.

3.5 Data challenges and testing

In South Africa, the lowest spatial unit used in the model is the *voting district* (VD). Every registered voter belongs to a VD and results are released at a VD level via the IEC database. Consequently, the previous results of VDs are clustered together to represent groups of similar voting behaviour, as mentioned in Subsection 3.2. In adapting the model to the USA elections, it was important to select the correct spatial unit to use in the model. Data for all US voting precincts was not available, both historically and on election night, and it was also not feasible to use data at this level due to spatial changes between election periods. Consequently, the county was chosen as the lowest spatial unit [46]. Historical data for the 2008 and 2012 presidential election results were obtained from the Data.Gov website [14]. The only problem with using counties was that it reduced the sample size of available data for prediction purposes, especially since some states have very few counties. For example, the District of Columbia had to be regarded as a state and county, Delaware has only 3 counties and for Alaska results are only released at a state level. In SA the data for the 2016 elections contained a total of 22 612 VDs, while in the USA, where the population is considerably larger, there were only 3 116 counties/spatial units, when considering Alaska as a single entity.

Accessing data on registered voters per county was not straightforward since not all states kept up-to-date information on registered voter counts prior to the elections. Data had to be sourced on a state by state basis from the relevant state websites and for those states where no registered voter records were available, the number of registered voters was estimated using the turnout and turnout percentages from the 2012 presidential elections.

Before moving straight into forecasting the 2016 USA presidential elections, it was imperative to test whether the adapted model would work in the USA context, using county data. Consequently, a test simulation of the 2012 USA presidential election was run using county data clustered from the 2008 presidential election results. Although no actual “time-stamp” could be sourced regarding the order in which counties were declared in 2012, an estimated “time-stamp” was created by using poll closure times, taking into account the time zones of the various states and then adding some randomised order of all counties within each of those closure times. Results from the test run proved to be very accurate (see Figure 5) in predicting the state winners and the overall electoral vote count for the Democrats (332 for Obama) and the Republicans (206 for Romney) after about 15% of the county results had been counted and released.

However, referring back to Table 2, one can see that apart from North Carolina, there were no changes in the winning party in the reported swing states between the 2008 and 2012 elections. Potentially, this indicates that, with president Obama standing for election a second time, the 2012 elections were relatively easy to predict correctly.

One of the major data challenges was to obtain “live” actual votes (total valid votes and party percentages per county) on election night for the USA 2016 presidential elections.

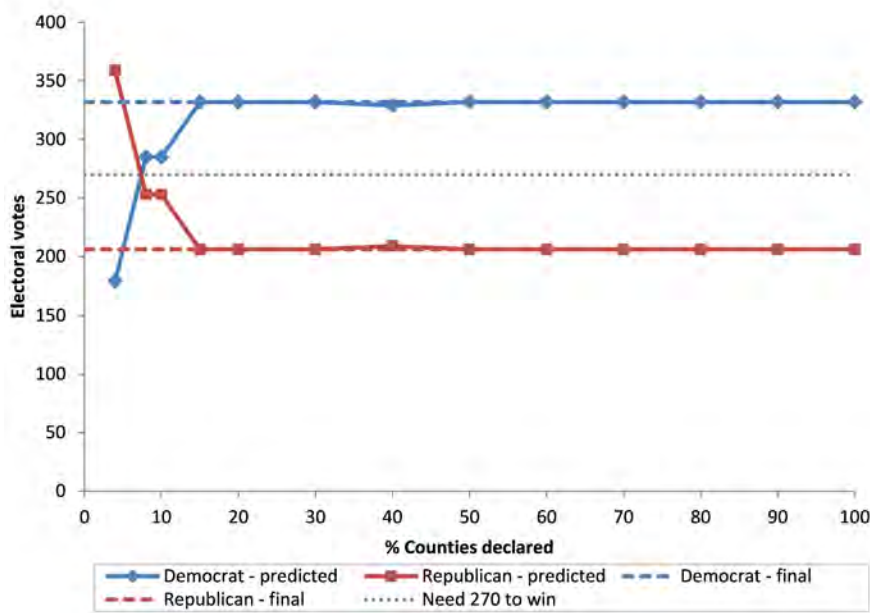


Figure 5: Simulated predicted electoral vote counts for the 2012 USA presidential elections at various percentages of counties declared.

In the SA context, the IEC has a national database containing all validated and released VD results, to which the CSIR team is given access, but to obtain results per county for the USA, the team had to rely on the Politico website [41]. Apart from the delay between when the county’s results were declared and when the results appeared on the website, the format in which the data would appear was unknown and therefore a software tool (“data scraper”), to automatically read the data off the website and output it to a formatted text file, could only be written once results started appearing on the website. This led to early data having to be captured manually, with some associated data errors, while various software bugs and data alignment issues had to be checked and sorted out once the “data scraper” was implemented.

3.6 Results of SA model applied to USA 2016 presidential elections

It is important to note that the model retains no prior knowledge of how people have previously voted in each state for either the swing states shown in Table 2 or the non-swing states and consequently predictions rely purely on new results per county, cluster memberships per county and registered voters per county. The following graph shows how the overall count of electoral votes, for both actual and predicted, changed as more county data became available. Due to data issues, the first corrected forecast could only be captured for the 8.7% sample of counties (see graph in Figure 6 and map in Figure 7) and, in addition, forecasts were stopped once it became obvious that Trump would undeniably be the final winner. Unfortunately, data was captured in terms of % counties declared (Figure 6) rather than time and is thus not directly comparable to Figure 3 and 4 given in Subsection 2.2.2.

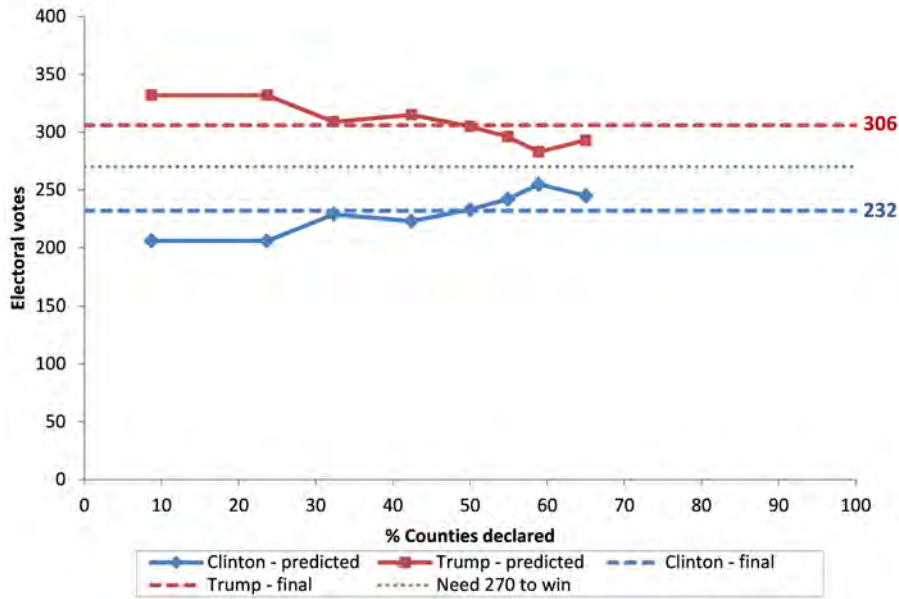


Figure 6: Final and predicted electoral vote counts for the 2016 US presidential elections at various percentages of counties declared.

Overall, the team was satisfied that the model had worked reasonably well once the correct data was used, especially in light of the many data issues that were encountered, and considering that it continuously predicted Trump as the winner. However, unlike the SA elections where the total percentages are reported to decide the winning party, maintaining the stability in the predictions for the USA elections was more difficult because of the electoral votes per state system where winner-takes-all and because of several very closely contested states. For example, Colorado, a swing state with 9 electoral votes, was initially predicted by the model to go to Trump since the initial predicted percentages were 47.8% for Clinton and 48.2% for Trump. It was only after 50% of all national counties and after about 15% of Colorado counties were declared that it swung the other way to predict Clinton to win the state, thus creating an overall substantial change in the total electoral vote counts per presidential candidate. Other swing states where the predictions changed over time were Iowa, Michigan, Minnesota, Nevada, New Hampshire, Virginia and Wisconsin. However, overall the predictions for 44 states (including the District of Columbia) remained stable throughout, despite no prior voting preferences having been included in the predictions, and these 44 stable state predictions comprised of 190 electoral votes for Clinton and 279 electoral votes for Trump – a stable overall predicted win for Trump.

4 Impact of model assumptions in USA predictions

As discussed in Ittmann *et al.* [27], the election night model relies on two main assumptions. First, it assumes that voting behaviour is not random but has a statistical pattern that can be modelled. As a result, the spatial units are clustered together into groups based

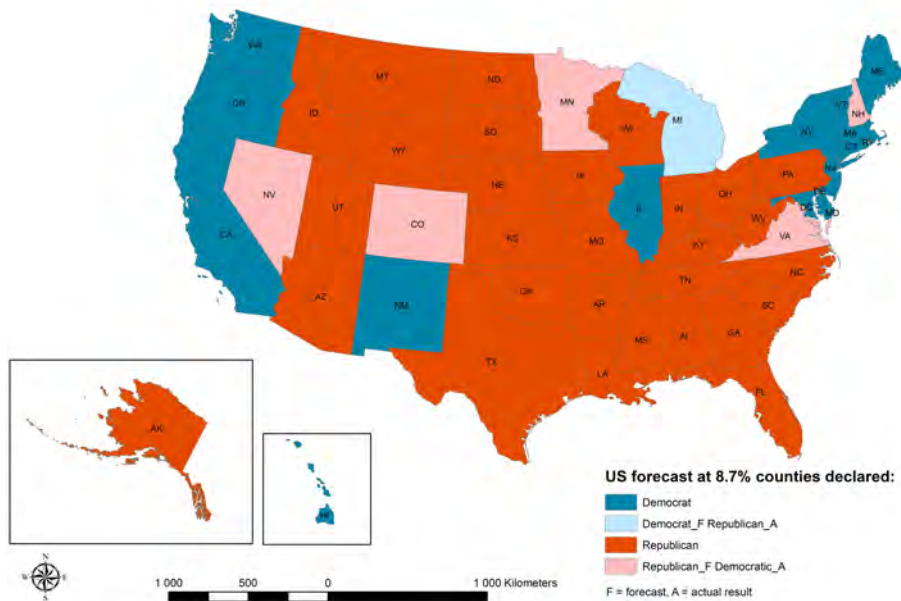


Figure 7: Model forecast of the 2016 USA presidential elections: Winning party per state in first prediction of electoral votes, with 8.7% counties declared nationally – Trump predicted to get 332 electoral votes and Clinton 206 electoral votes.

on similar past voting behaviour. Linked to this assumption is the belief that changes in voting behaviour are also non-random and therefore groups that voted in a similar way before may experience similar changes in their voting preferences due to various political and socio-economic influences. In testing the 2012 USA presidential elections using past voting behaviour from the 2008 elections, this assumption proved to hold true, due to reasons mentioned in subsection 3.5, thus resulting in very stable predictions. However, in 2016, the political influences within states varied substantially due to the choice of presidential candidates as well as their campaigns and since clusters were created using counties across the whole nation based on the 2012 election results, it was to some degree more difficult for this assumption to hold true. However, despite these influences and the fact that the predictions were less stable, the clusters in the model still managed to predict the overall winner correctly.

The second assumption of the model involves the order in which the voting results are released, wherein the model is designed to adjust for moderate non-random patterns in the order. The initial sample of county results that were released and used in the model were heavily biased in favour of Trump (245 of the 275 counties declared in the initial sample were in favour of Trump – see map in Figure 8) and although this resulted in an over-estimation of electoral votes for Trump, the model was still able to adequately adjust for some of this order bias (see graph and map showing first prediction in Figure 6 and 7 respectively). Although it was expected, due to time zones, that the results from the far eastern half of the US would come in first, thus creating an extreme spatial bias, in reality this did not prove to be quite as extreme as expected and the model was able to adjust for some of the actual spatial bias in the order of counties released.

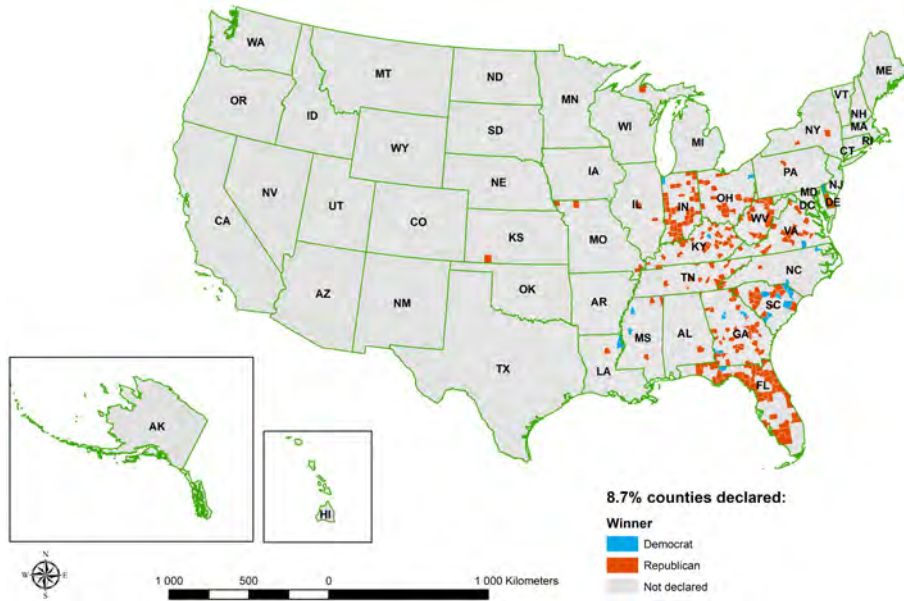


Figure 8: Spatial distribution and political party preference of counties declared in the first prediction (8.7% sample of counties).

A third model assumption, which has never been an issue in the South African elections, is the availability of correct numbers of registered voters per spatial unit, since in SA there is stringent control of the voter registration roll across the country. In the case of the USA, however, where registered voters had to be estimated for counties in certain states, there was a concern that this would play a significant role on either under or over estimating the vote count and voter turnout for some areas in the model, thus affecting the predictions. In order to do a post-election analysis to test this assumption, test simulations using the 2010–2014 estimated *Census Voting Age Population* (CVAP) for all counties across all states obtained from the United States Census Bureau website [55], in place of the registered voter population figures used on election night, was carried out. The differences between these two sets of voter count values was quite substantial since the percentage difference between the two counts per county revealed an overall average difference of 11.3% and an overall standard deviation of 26.8%. For certain states the values were fairly similar but for some states these differences were quite large. The test simulation showed that the initial prediction, at 8.7% of counties declared, using the CVAP in place of registered voters was slightly worse (by another 16 electoral votes) than the one obtained on election night but after this time point the predictions were almost exactly the same. The implications of this test are that the correctness of the registered voter population can affect the early predictions from the model when the sample size of counted spatial units is small but it becomes less of an issue as more election result data becomes available.

5 Remarks on election forecasting and potential adaptations to the SA model

In the “unexpected outcome” elections, discussed in Section 2, forecasts from polling data were predominantly found to be inaccurate, even though the methodologies sometimes differed substantially, and there has been an abundance of literature that have delved into the issues surrounding these inaccuracies. However, pre-election forecasts do play a vital role in terms of stimulating public interest and driving party campaigns in the lead up to the elections. In some cases it could even be said that forecasts can affect the outcome by encouraging the turnout or non-turnout of certain groups of people or people in certain contentious areas. The South African model discussed in this paper has so far only been applied to election night forecasting but going forward there is indeed scope to apply this same model to a sample of spatial units taken from polling data to obtain pre-election forecasts. This would involve using the same set of clusters that would be setup for election night and extrapolating cluster predictions, from the polling sample, to non-sampled spatial units. It would, however, rely strongly on the sampling methodology of the polling within each spatial unit which would need to be sufficiently representative of that entire voting area.

Election night forecasting models also have their value in capturing public interest but in a limited time frame. The common issues arising from the election night forecasting models found in the literature in Subsection 2.2, as well as the South African model discussed in Section 3, include the problem of accommodating changes to the demarcation of voting booths or voting districts, as well as the non-random manner in which the elections results are released during the election night. In terms of methodology, these models all have a component of grouping voters even though the manner in which the groups are constructed varies. Unlike the South African model, the other election night models predominantly incorporate polling data to initialise the model predictions on election night. In many cases, such as the USA examples in subsection 2.2.2, these priors actually negatively affected the starting election night forecasts causing a complete swop in the forecasted winning party at some point during the election night. The South African model could potentially be adapted to incorporate pre-election forecasts from polling data as initial priors in the election night model but it carries the risk of initialising the starting points of the forecasts to unreliable values, as seen in the recent UK and USA elections, that could then cause the model to converge more slowly to the correct outcome. On the other hand, if pre-election forecasts are reasonably correct, a much faster convergence to the final outcome on election night could result. One of the strong points of the SA model has always been its independence of prior data and its reliance on only initialising the forecasts from actual released results. Therefore, such adaptations need to be considered with caution and tested for robustness before being implemented.

6 Conclusion

In summary, it was found that the SA election night forecasting model could be successfully adapted to different electoral systems such as the one used in the USA. The model

assumptions did have some influence on the stability of the forecasts but this impact was limited, allowing the model to still produce reasonable forecasts for the 2016 USA presidential elections. Forecasting outside of South Africa, however, does come with a lot of data problems and is therefore difficult unless access to appropriate data at the desired spatial level can be obtained both historically and in “real time” during the election night.

With growing technology, which has sped up the collection of voting results, the “window of opportunity” for such election night forecasting has, in some countries, become considerably smaller and therefore going forward, serious consideration needs to be given to testing the applicability of this SA model on polling data collected at relevant spatial units. Should such an exercise be successful, it would certainly increase the value of the model by increasing the time window of its forecasts, allowing the model to be used for both pre-election and election night forecasting.

Acknowledgments

The authors would like to acknowledge Dr Jan Greben who developed the initial methodology as well as Brenwen Ntlangu, our indispensable software developer, who was able to keep his cool no matter what data challenges were thrown his way. The authors would also like to thank the other members of the project team who worked in the lead up to, as well as during, the local elections and those who worked through the night (South African time) on the USA Election Day in order to help test out the methodology on the 2016 USA presidential elections. Special thanks go to Dr Zaid Kimmie who led the authors through both elections and who inspired the authors to take on this new challenge. The authors would also like to thank the SABC for giving the CSIR the opportunity to be involved in the South African local elections.

References

- [1] ABRAMOWITZ A, 2017, *The Color Purple: An Examination of the 2008 Presidential Election in the Swing States*, [Online], [Cited July 8th, 2017], Available from https://www.huffingtonpost.com/alan-abramowitz/the-color-purple-an-exami_b_115033.html.
- [2] AISCH G, COHN N, COX A, KATZ J, PEARCE A, & QUEALY K, 2016, *Live Presidential Forecast*, [Online], [Cited September 28th, 2017], Available from <https://www.nytimes.com/elections/forecast/president>.
- [3] BBC, 2015, *Conservatives win 12-seat majority*, [Online], [Cited May 18th, 2015], Available from <http://www.bbc.com/news/election/2015/results>.
- [4] BBC, 2016, *EU Referendum Results*, [Online], [Cited May 18th, 2015], Available from http://www.bbc.com/news/politics/eu_referendum/results.
- [5] BERLIN J, 2017, *Pessimism and optimism on Germany after its election*, [Online], [Cited September 27th, 2017], Available from <https://www.economist.com/blogs/kaffeeklatsch/2017/09/germany-different-angles>.
- [6] BOFFEY D, 2017, *World leaders congratulate Macron on victory as EU breathes sigh of relief*, [Online], [Cited September 3rd, 2017], Available from <https://www.theguardian.com/world/2017/may/07/theresa-may-congratulates-macron-on-victory-as-eu-breathes-sigh-of-relief>.

- [7] BOMBOY S, 2016, *What are the real Swing States in the 2016 election?*, [Online], [Cited October 5th, 2017], Available from <https://constitutioncenter.org/blog/what-are-the-really-swing-states-in-the-2016-election>.
- [8] BROWN P, FIRTH D & PAYNE CD, 1999, *Forecasting on British election night 1997*, Journal of the Royal Statistical Society: Series A (Statistics in Society), **162(2)**, pp. 211–226.
- [9] BROWN P & PAYNE CD, 1975, *Election night forecasting*, Journal of the Royal Statistical Society: Series A (General), pp. 463–498.
- [10] BYCOFFE A, SCHEINKMAN A, & SILVER N, 2016, *Live Election Night Forecast*, [Online], [Cited September 1st, 2017], Available from <https://projects.fivethirtyeight.com/election-night-forecast-2016/>.
- [11] CILLIZZA C, 2012, *The 9 swing states of 2012*, [Online], [Cited 11 June 2017], Available from https://www.washingtonpost.com/blogs/the-fix/post/the-9-swing-states-of-2012/2012/04/16/gIQABuXaLT_blog.html?utm_term=.28bfe4c29449.
- [12] CLINTON HR, 2017, *What happened*, Simon & Schuster, New York.
- [13] COWLEY P & KAVANAGH D, 2016, *The British general election of 2015*, Springer.
- [14] DATA.GOV, 2016, *Datasets*, [Online], [Cited September 28th, 2017], Available from <https://catalog.data.gov/dataset>.
- [15] DAVISS C, 2015, *Does the Candidate Determine the Battleground States in Presidential Elections?*, [Online], [Cited July 10th, 2017], Available from <http://www.fairvote.org/does-the-candidate-determine-the-battleground-states-in-presidential-elections>.
- [16] DICTIONARY.COM, 2016, *Swing State*, [Online], [Cited October 20th, 2017], Available from <http://www.dictionary.com/browse/swing-state>.
- [17] ENGEL P, 2016, *An artificial intelligence system that correctly predicted the last 3 elections says Trump will win*, [Online], [Cited September 6th, 2017], Available from <http://www.businessinsider.com/artificial-intelligence-trump-win-2016-10?IR=T>.
- [18] GELMAN A, & SILVER N, 2008, *What will we know on Tuesday at 7pm?*, Mathematics Magazine, **83(4)**, pp. 258–266. Available from <https://academiccommons.columbia.edu/catalog/ac:125231>.
- [19] GREBEN JM, ELPHINSTONE C, HOLLOWAY J, DE VILLIERS R, ITTMANN H & SCHMITZ P, 2005, *Prediction of the 2004 national elections in South Africa*, South African Journal of Science, **101(3-4)**, pp. 157–161.
- [20] GREBEN JM, ELPHINSTONE C & HOLLOWAY J, 2006, *A model for election night forecasting applied to the 2004 South African elections*, ORION, **22(1)**, pp. 89–103.
- [21] HANRETTY C, LAUDERDALE B & VIVYAN N, 2015, *Final 2015 UK Parliamentary Election Forecast*, [Online], [Cited September 27th, 2017], Available from <http://www.electionforecast.co.uk/2015/index.html>.
- [22] HENDERSON D Barney, Lawler, 2017, *How does the US election work and what is a swing state?*, [Online], Available from <https://www.telegraph.co.uk/news/0/how-does-the-us-election-work-and-which-swing-states-will-determ/>.
- [23] HIRST T, 2015, *This polling company claims it knew the General Election result ahead of time but ‘chickened out’ of publishing it*, [Online], [Cited 27 September 2017], Available from <http://uk.businessinsider.com/survation-unpublished-poll-general-election-result-2015-5>.
- [24] HOBOLT SB, 2016, *The Brexit vote: a divided nation, a divided continent*, Journal of European Public Policy, **23(9)**, pp. 1259–1277.

- [25] HOCHREITER R & WALDHAUSER C, 2015, *Evolving accuracy: A genetic algorithm to improve election night forecasts*, Applied Soft Computing, **34**, pp. 606–612.
- [26] INDEPENDENT ELECTORAL COMMISSION OF SOUTH AFRICA (IEC), 2013, *Election types*, [Online], [Cited October 10th, 2017], Available from <http://www.elections.org.za/content/Elections/Election-types/>.
- [27] ITTMANN HW, HOLLOWAY JP & DUDENI-TLHONE N, 2017, *2014 Election forecast- A post-election analysis*, ORIION, **33(1)**, pp. 35–55.
- [28] KEETER S, HOROWITZ J & TYSON A, 2008, *Young Voters in the 2008 Election*, [Online], [Cited July 11th, 2017], Available from <http://www.pewresearch.org/2008/11/13/young-voters-in-the-2008-election/>.
- [29] KENTISH B, 2016, *US election 2016: 15 states that will decide everything*, [Online], [Cited July 7th, 2017], Available from <https://www.independent.co.uk/news/world/americas/us-elections/us-election-2016-states-to-win-florida-ohio-colorado-iowa-a7402871.html>.
- [30] LEBO MJ & NORPOTH H, 2016, *Victory without power: The PM-Pendulum forecast*, Electoral Studies, **41**, pp. 255–259.
- [31] LEWIS-BECK MS, 2005, *Election forecasting: Principles and Practice*, The British Journal of Politics & International Relations, **7(2)**, pp. 145–164.
- [32] LICHTMAN AJ, 2016, *Predicting the next president: The keys to the White House*, Rowman & Littlefield, Lanham, Md., USA.
- [33] MAHTESIAN C, 2016, *What are the swing states in 2016?*, [Online], [Cited July 7th, 2017], Available from <https://www.politico.com/blogs/swing-states-2016-election/2016/06/what-are-the-swing-states-in-2016-list-224327>.
- [34] MACLUHAN M, 1964, *Understanding Media: The Extensions of Man*, revised, critical addition, 2003, Gingko Press, Corte Madera, Calif., USA.
- [35] MOORE M, 2017, *5 Reasons Why Trump Will Win*, POSITIF, (**674**), pp. 60–64.
- [36] MORTON RH, 1988, *Election night forecasting in New Zealand*, Electoral Studies, **7(3)**, pp. 269–277.
- [37] MORTON RH, 1990, *Election night forecasting*, The New Zealand Statistician, **25**, pp. 36–84.
- [38] PAYNE CD, 2003, *Election forecasting in the UK: The BBC's experience*, Euramerica, **33**, pp. 193–234.
- [39] PFEIFFER A, 2004, *The Election of 2004 – Battleground and Swing States*, [Online], [Cited July 10th, 2017], Available from <http://cphcmp.smu.edu/2004election/battleground-swing-states/>.
- [40] POLITICO, 2012, *Swing-State Map, List & Polls*, [Online], [Cited July 10th, 2017], Available from <https://www.politico.com/2012-election/swing-state/>.
- [41] POLITICO, 2016, *2016 Presidential Election Results*, [Online], [Cited November 8th, 2016], Available from <https://www.politico.com/mapdata-2016/2016-election/results/map/president/>.
- [42] REALCLEAR POLITICS, 2016, *Battle for White House*, [Online], [Cited July 8th, 2017], Available from https://realclearpolitics.com/epolls/2016/president/2016_elections_electoral_college_map.html.
- [43] SAAD L, 2012, *Obama-Romney Race Competitive in 2012 Swing States*, [Online], [Cited July 7th, 2017], Available from <http://news.gallup.com/poll/154502/obama-romney-race-competitive-2012-swing-states.aspx>.
- [44] SOUTH AFRICAN GOVERNMENT GAZETTE, 1998, *Act 73 of 1998: Electoral Act*, [Online], [Cited October 12th, 2017], Available from <https://www.gov.za/sites/default/files/Act73of1998.pdf>.

- [45] SAMUELSON DA, 2016, *The election that confounded everybody*, ORMS-Today, **43(6)**, pp. 28–31.
- [46] SCHMITZ PM, HOLLOWAY JP, DUDENI-TLHONE N, NTLANGU MB & KOEN R, 2017, *Could the outcome of the 2016 US elections have been predicted from past voting patterns?*, in Paper presented at the 28th International Cartographic Conference 2017, Washington DC, USA.
- [47] SEELYE K, 2004, *The 2004 Campaign: The Strategy, Some Swing States Appear to Be Swinging to President*, [Online], [Cited July 11th, 2017], Available from <https://www.nytimes.com/2004/09/29/us/the-2004-campaign-the-strategy-some-swing-states-appear-to-be-swinging.html>.
- [48] SILVER N, 2016a, *Election Update: The Swing States Are Tightening, Too*, [Online], [Cited July 7th, 2017], Available from <http://fivethirtyeight.com/features/election-update-the-swing-states-are-tightening-too/>.
- [49] SILVER N, 2016b, *Who will win the presidency?*, [Online], [Cited September 6th, 2017], Available from <https://projects.fivethirtyeight.com/2016-election-forecast/#plus>.
- [50] SILVER N, 2016c, *Why FiveThirtyEight gave Trump a better chance than almost anyone else*, [Online], [Cited September 6th, 2017], Available from <http://fivethirtyeight.com/features/why-fivethirtyeight-gave-trump-a-better-chance-than-almost-anyone-else/>.
- [51] SILVER N, 2017, *The Real Story of 2016*, [Online], [Cited October 12th, 2017], Available from <http://fivethirtyeight.com/features/the-real-story-of-2016/>.
- [52] TERDIMAN D, 2012, *Obama's win a big vindication for Nate Silver, king of the quants*, [Online], [Cited September 6th, 2017], Available from <https://www.cnet.com/news/obamas-win-a-big-vindication-for-nate-silver-king-of-the-quants/>.
- [53] THE WASHINGTON POST, 2008, *Key Swing States in the 2008 elections*, [Online], [Cited July 11th, 2017], Available from http://www.washingtonpost.com/wp-srv/politics/interactives/campaign08/battleground_cheat_sheet.html.
- [54] UK POLITICAL INFO., 1992, *1992 General election results summary*, [Online], [Cited October 2nd, 2017], Available from <http://www.ukpolitical.info/1992.htm>.
- [55] UNITED STATES CENSUS BUREAU, 2016, *Voting Age Population by Citizenship and Race (CVAP)*, [Online], [Cited September 14th, 2017], Available from <https://www.census.gov/programs-surveys/decennial-census/about/voting-rights/cvap.html>.
- [56] URBINA I, 2008, *States' Actions to Block Voters Appear Illegal*, [Online], [Cited July 7th, 2017], Available from <https://www.census.gov/programs-surveys/decennial-census/about/voting-rights/cvap.html>.
- [57] UNITED STATES GOVERNMENT, 2008, *Presidential Election Process*, [Online], [Cited October 10th, 2017], Available from <https://www.usa.gov/election>.
- [58] YOUGov UK a, 2016, *YouGov / Times Survey Results, Eve of Poll Voting Intention*, [Online], [Cited September 6th, 2017], Available from <https://yougov.co.uk/publicopinion/archive/>.
- [59] YOUGov UK b, 2016, *YouGov / Times Survey Results, On the Day Vote*, [Online], [Cited September 6th, 2017], Available from <https://yougov.co.uk/publicopinion/archive/>.
- [60] ZOGBY J, 2016, *Election 2016: How Did We Get To This?*, Podcast, Center for Strategic and International Studies, [Online], Available from <https://www.csis.org/events/election-2016-how-did-we-get>.