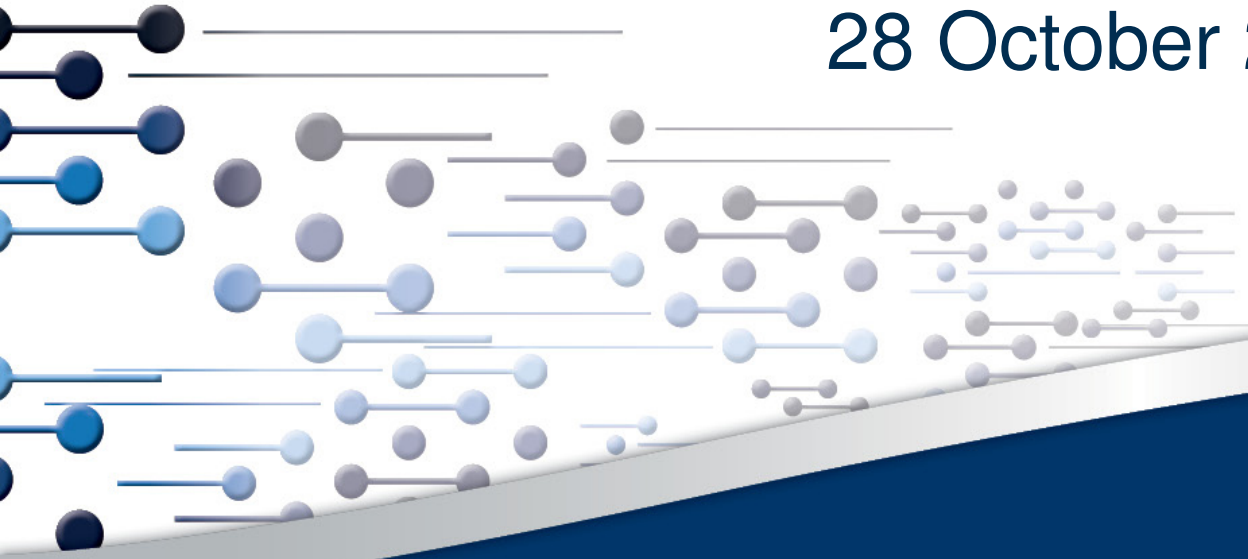
CSIR Elections Forecasting

2016 Local Government Elections

Zaid Kimmie

28 October 2016



Overview

1. Team Members
2. Some History
3. Why Forecast?
4. Methods: Clustering
5. Methods: Predictions
6. Model Performance
7. What Next?

The CSIR Team

Statisticians, computer scientists, and programmers . . .

- Peter Schmitz, Jenny Holloway, Nontembeko Dudeni-Thlone
- Brenwen Ntlangu, Tyrone Naidoo
- Zaid Kimmie, Ndumiso Cingo and Luyanda Vappie
- Paul Mokilane, Quintin van Heerden, Sumarie Meintjes
- Hans Ittmann, Jan Greben, Renee Koen

Some History

- Worked with IEC for 1999 national 2000 municipal elections
  - Checking inconsistencies in voting patterns
  - Forecasts a “by-product” of this methodology
- Worked with SABC for all elections since 2004
  - Produce a forecast of the final results

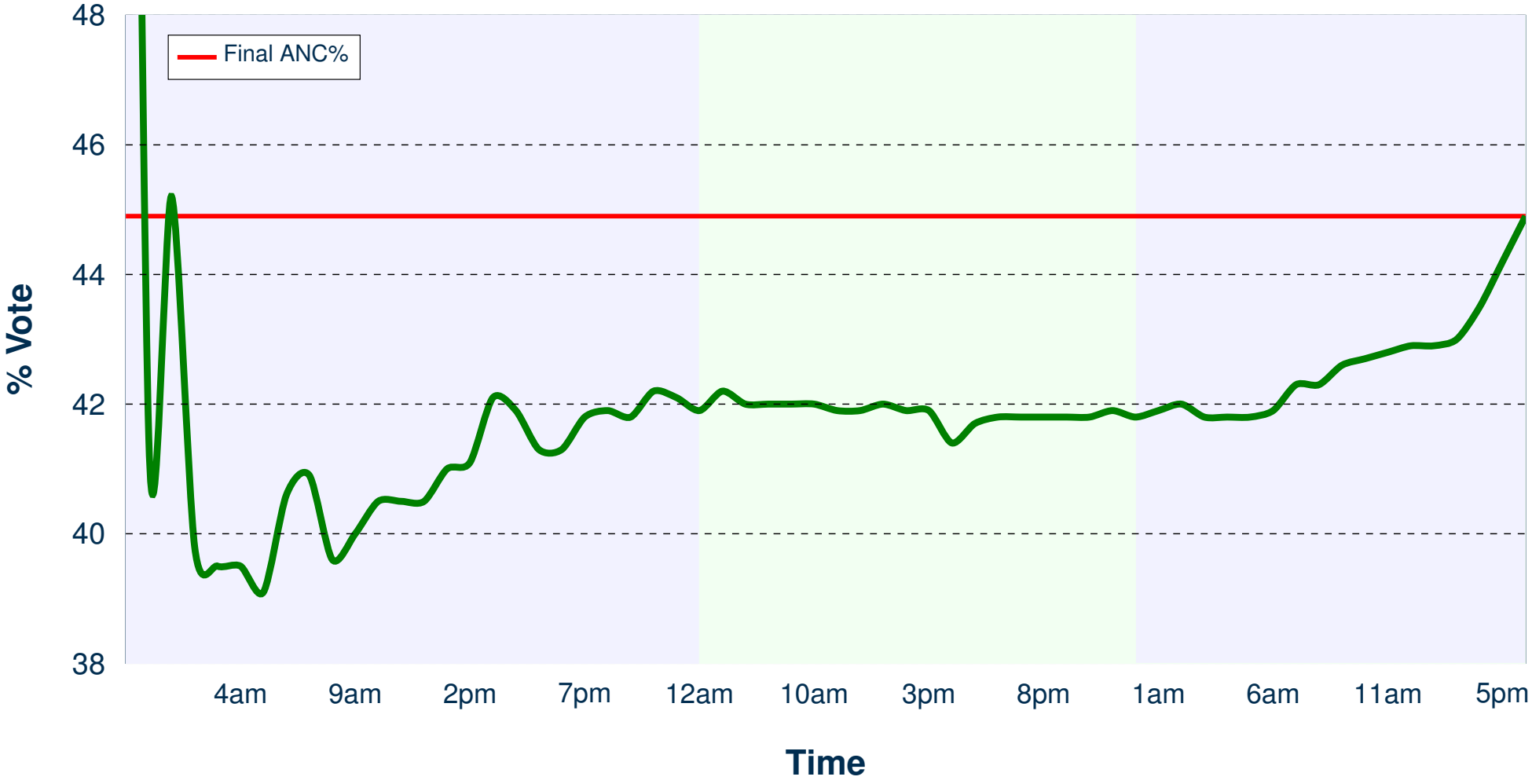
Why Forecast?

- Election results are released by Voting District (VD)
  - Some 22,500 VDs in total
- If the VDs reported in a random order there would be nothing much to do
  - The final result would become clear relatively quickly
  - E.g. 5% of VDs have reported the tabulated results are within a couple of percentage points of their final values

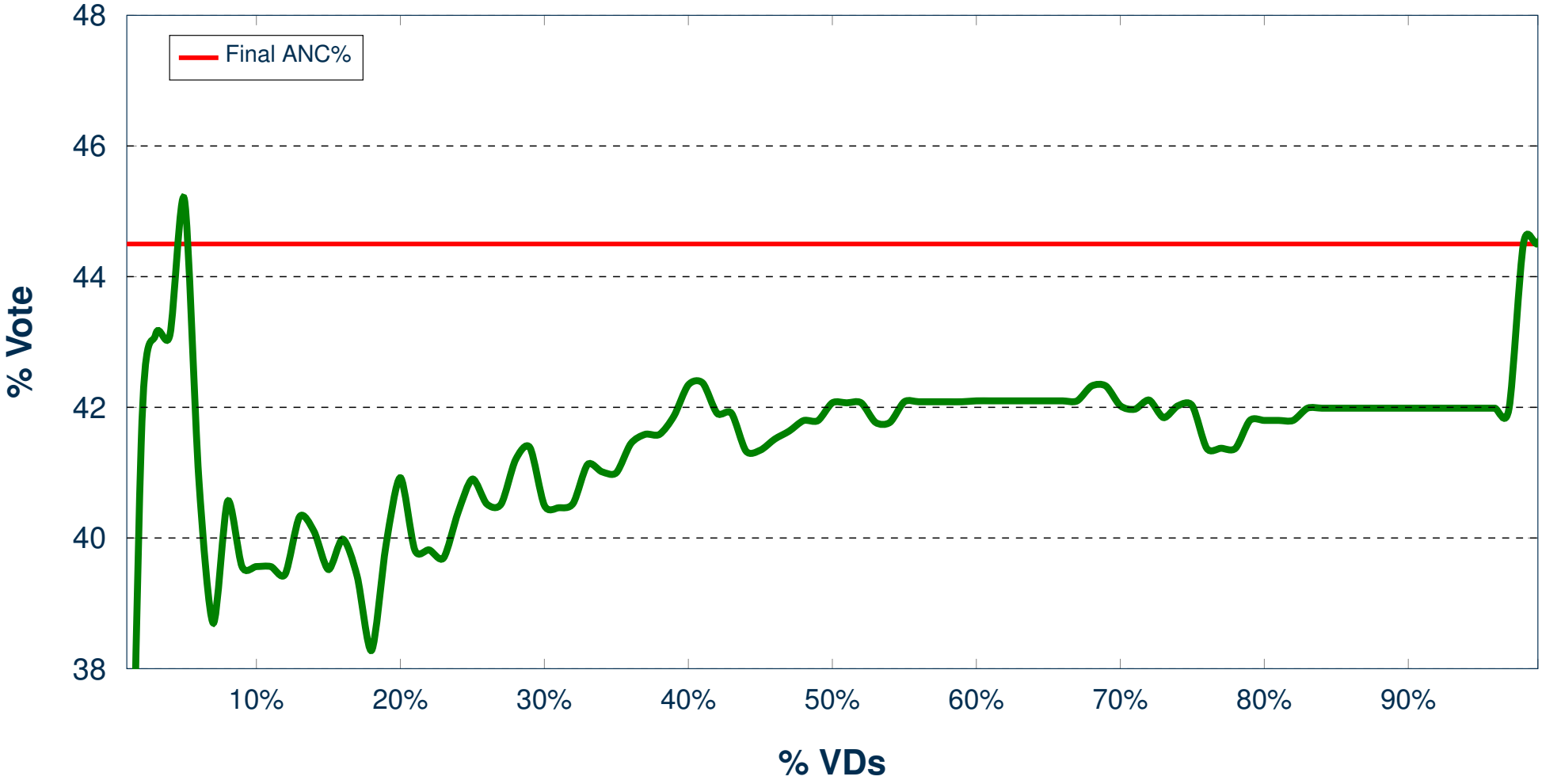
Why Forecast?

- Fortunately (for us) VDs do not report randomly
  - There is in fact a systematic bias in the reporting order
  - The difference, in the early stages, between the “live” and final results are often substantial
  - These differences may persist . . .
- People (including political analysts and the curious member of the public) looking at the “scoreboard” will find that it gives them very little useful information

ANC – Johannesburg Metro



ANC – Johannesburg Metro



Why Forecast?

- In this “window of opportunity” election forecasts can provide useful insights
  - What do the the initial results really mean?
  - Identify interesting patterns that are emerging
- The combination of pre-election polling data and exit polls can get it wrong …
  - Brexit, UK 2015 general election

Why Forecast?

- In this “window of opportunity” election forecasts can provide useful insights
  - What do the the initial results really mean?
  - Identify interesting patterns that are emerging
- The combination of pre-election polling data and exit polls can get it wrong …
  - Brexit, UK 2015 general election
- It can make you look smarter than you actually are …

Why Forecast?

Forecasting Model: Basics

- Method published by Greben, Elphinstone & Holloway, 2006 in ORiON: The Journal of ORSSA

Forecasting Model: Basics

- Method published by Greben, Elphinstone & Holloway, 2006 in ORiON: The Journal of ORSSA

There are a couple of basic principles:

1. Voters do not randomly allocate their electoral preferences – they are influenced by political, socio-economic and demographic factors, as well as past voting history;
2. Changes in voting behaviour between one election and the next are also not random, but are correlated with past voting behaviour, demographic and socio-economic factors.

Example

Suppose our area of interest consists of 200 VDs, and that in the previous election party A has obtained 70% of the vote in the area, with relatively small variation between VDs

- When the first VD reports . . .
- When 10 VDs have reported . . .
- When 30 VDs have reported . . .

Methods: Clustering

The first step is to create clusters of VDs based on previous voting results.

- Fuzzy-c-means
- Fixed number of clusters, c
- Fuzzy clustering performs better than other methods (k-means, k-means with discriminant analysis) – smallest prediction error
- How many clusters?

Methods: Predictions

Two-step process:

- Estimate turnout for outstanding VDs
- Assign fuzzy-cluster estimates to VDs

Methods: 2016 Predictions

- Metro predictions based on **provincial** clusters
- This method allowed us to (accurately) predict eThekwini when **no** results had been released
- But this setup can let us down when inter-provincial variations do occur, as was the case with Tshwane

Model Performance

There are two aspects of model performance – the technical performance of the model and our ability to communicate the model output to the general public

Model Performance

- Assuming that IT snafus have not rendered us mute …
- Early on the Thursday morning after election day – some-where between 5am and 9am, when only about 10% of all VDs have reported – we forecast the final results
- We continually update our forecasts, but the numbers do not change all that much, and the level of interest in the forecast declines as the “scoreboard” starts to match the final score

How did we do?

- Pretty well!
- By 5am on Thursday we identified the major trends well before they could be inferred just by looking at the data
  - That the DA would be the largest party in NMB, but not achieve a majority
  - That the ANC would lose its majorities in all the Gauteng metros
  - That the DA would increase its majority in Cape Town
  - That the ANC would continue to hold a majority in Buffalo City, Mangaung and eThekwini

How did we do?

- We were able to predict that the ANC's share of the national vote would fall to 54%
- We did not get the final result in Tshwane right – our model predicted (and continued to predict until quite late into the reporting) that the ANC would be the largest party
- In general we were able to get within 1.5 percentage points of the final result for the larger parties, and in most cases within 0.5 percentage points

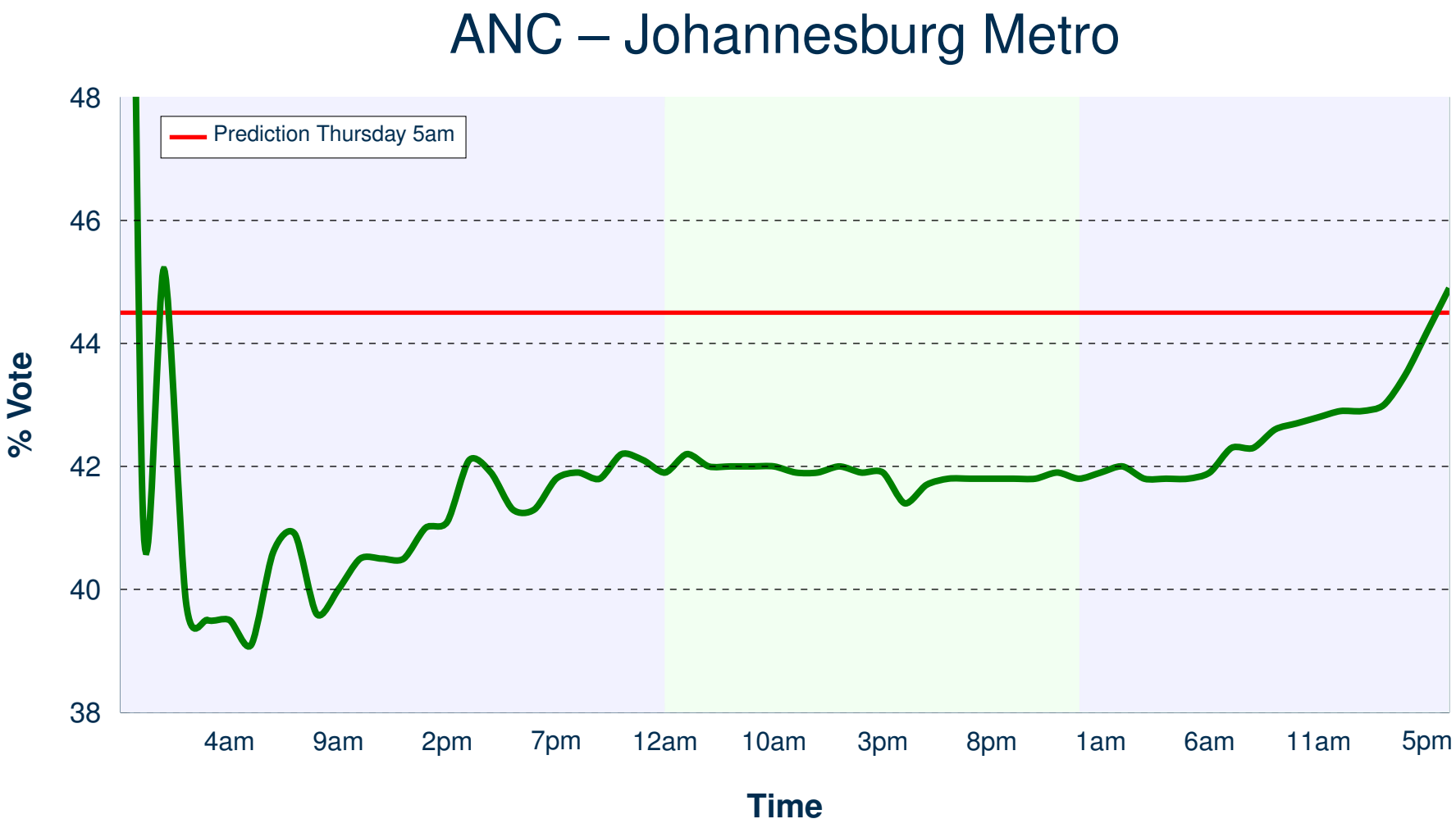
Forecasts

Metro	Party	Predicted 5am	Final	Actual 5am
Johannesburg	ANC	44.5	44.9	39.5
	DA	38.9	38.4	45.3
	EFF	10.7	10.9	9.8
Tshwane	DA	41.5	43.1	47.0
	ANC	42.8	41.5	41.0
	EFF	10.7	11.6	7.8
Ekurhuleni	ANC	47.8	48.9	38.8
	DA	35.8	34.2	50.0
	EFF	10.7	11.1	7.8

Forecasts

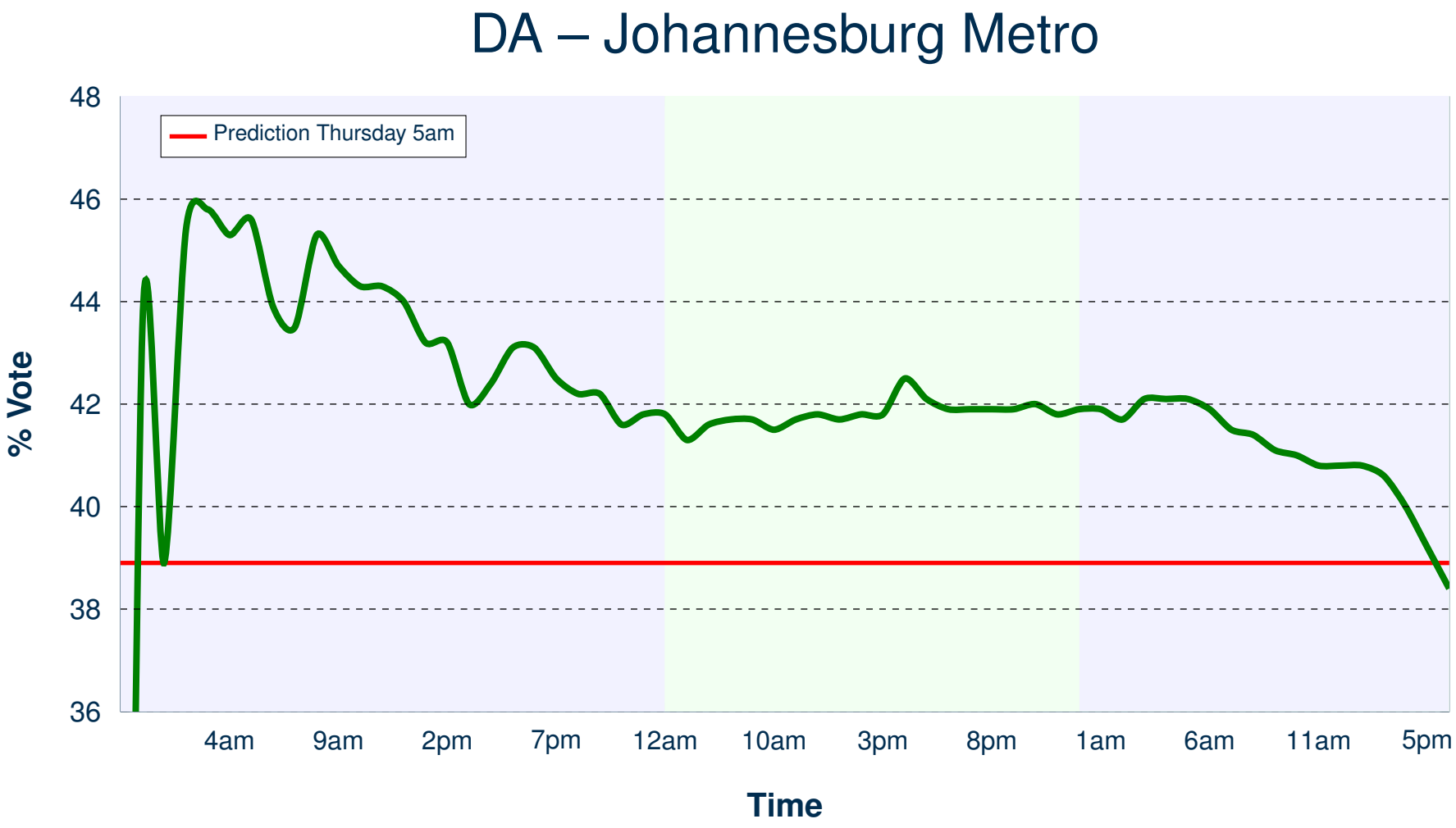
Metro	Party	Predicted 5am	Final	Actual 5am
Cape Town	DA	65.7	66.8	72.5
	ANC	25.1	24.5	18.8
	EFF	2.9	3.1	3.1
Nelson Mandela	DA	48.3	46.6	58.6
	ANC	42.2	41.5	32.9
	EFF	3.9	5.0	3.8
eThekwini	ANC	58.8	60	0
	DA	27.7	27.5	0
	IFP	4.0	4.3	0

Johannesburg



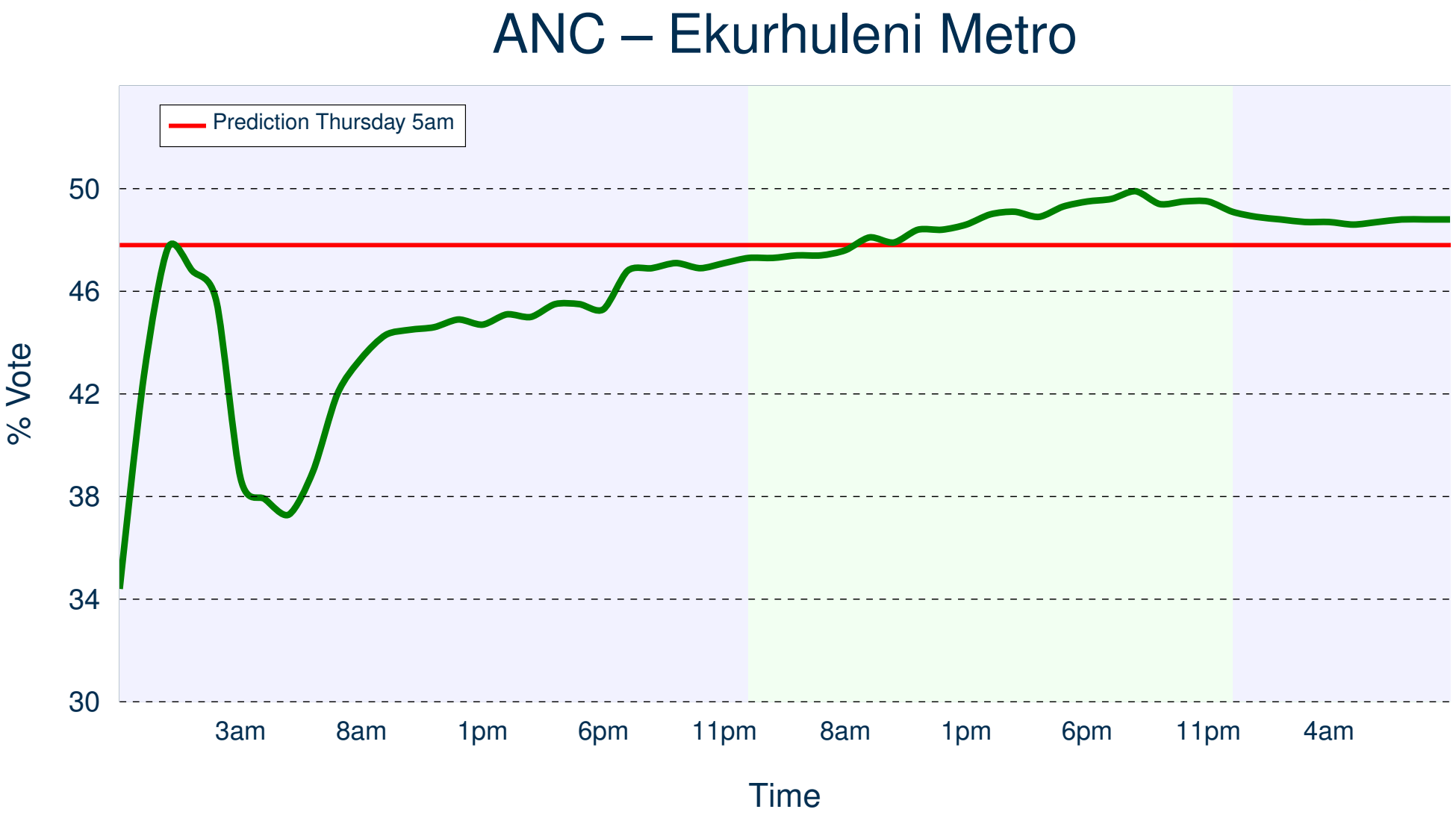
CSIR
our future through science

Johannesburg



CSIR
our future through science

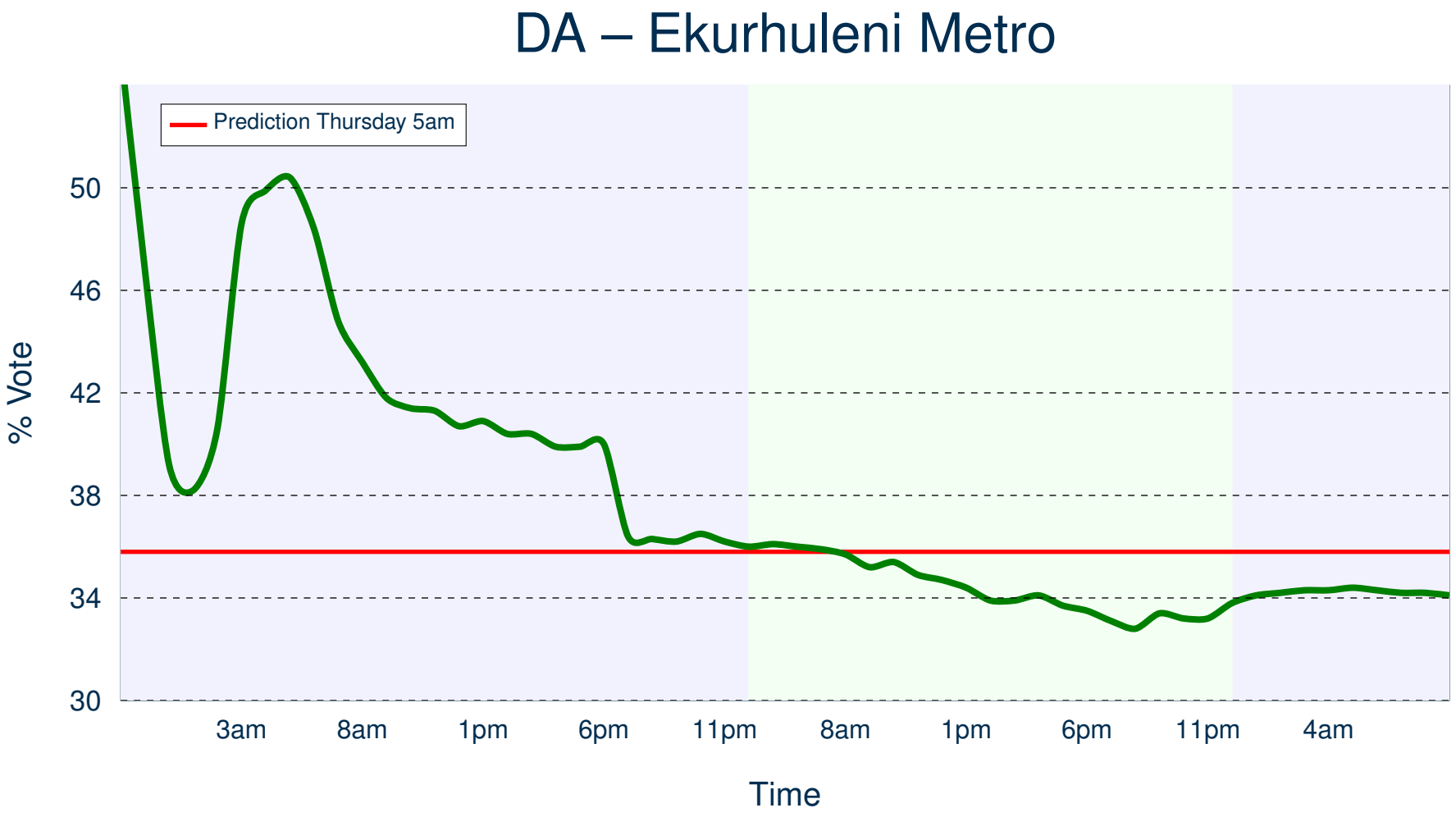
Ekurhuleni


ANC – Ekurhuleni Metro
Prediction Thursday 5am
50
46
42
38
34
30
% Vote
3am
8am
1pm
6pm
11pm
8am
1pm
6pm
11pm
4am
Time

Ekurhuleni

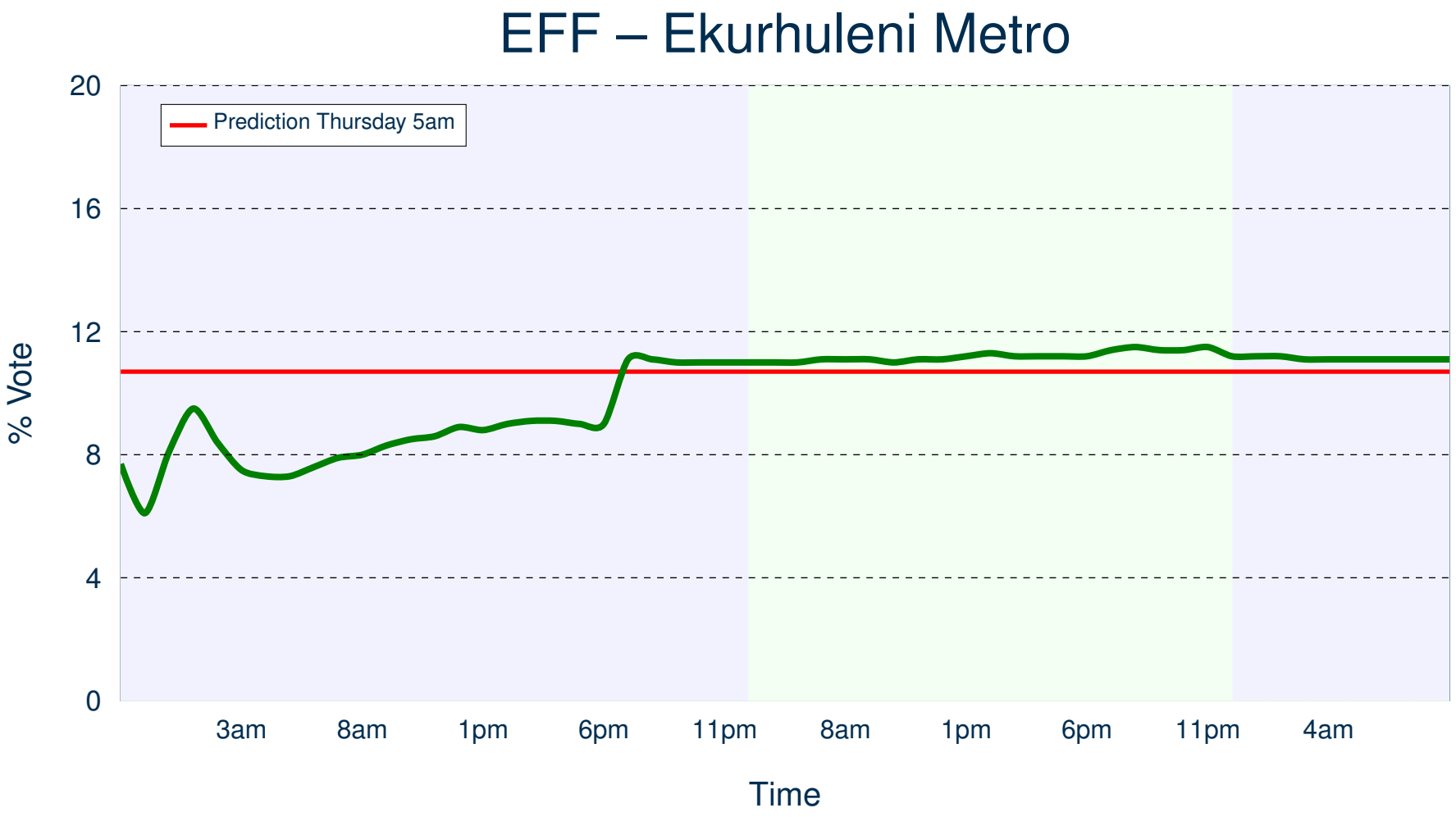
- The difference between the predicted and actual ANC vote count in Ekurhuleni was less than 9,000 votes – just under 900,000 people voted in this metro.

Ekurhuleni



CSIR
our future through science

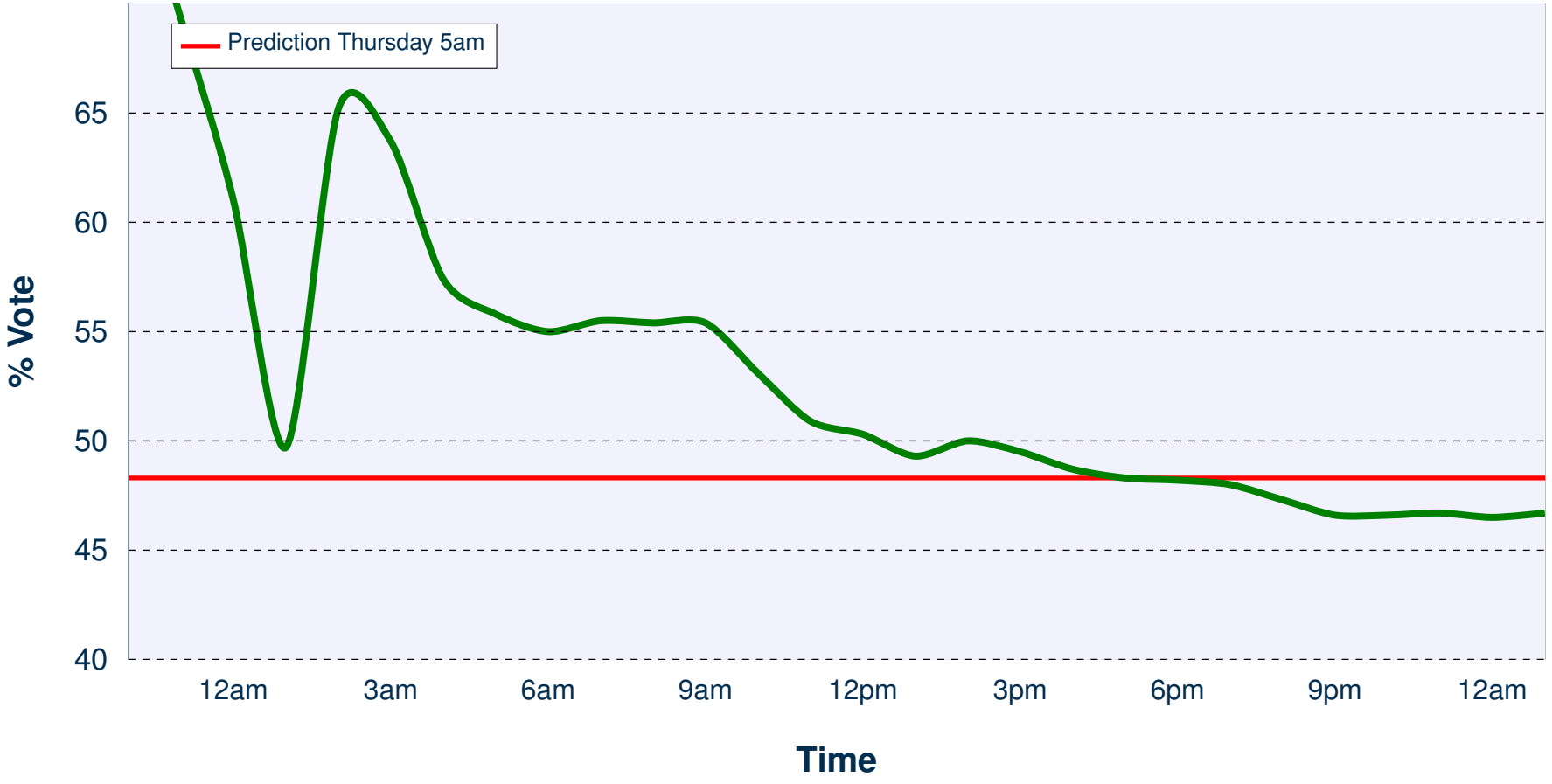
Ekurhuleni



CSIR
our future through science

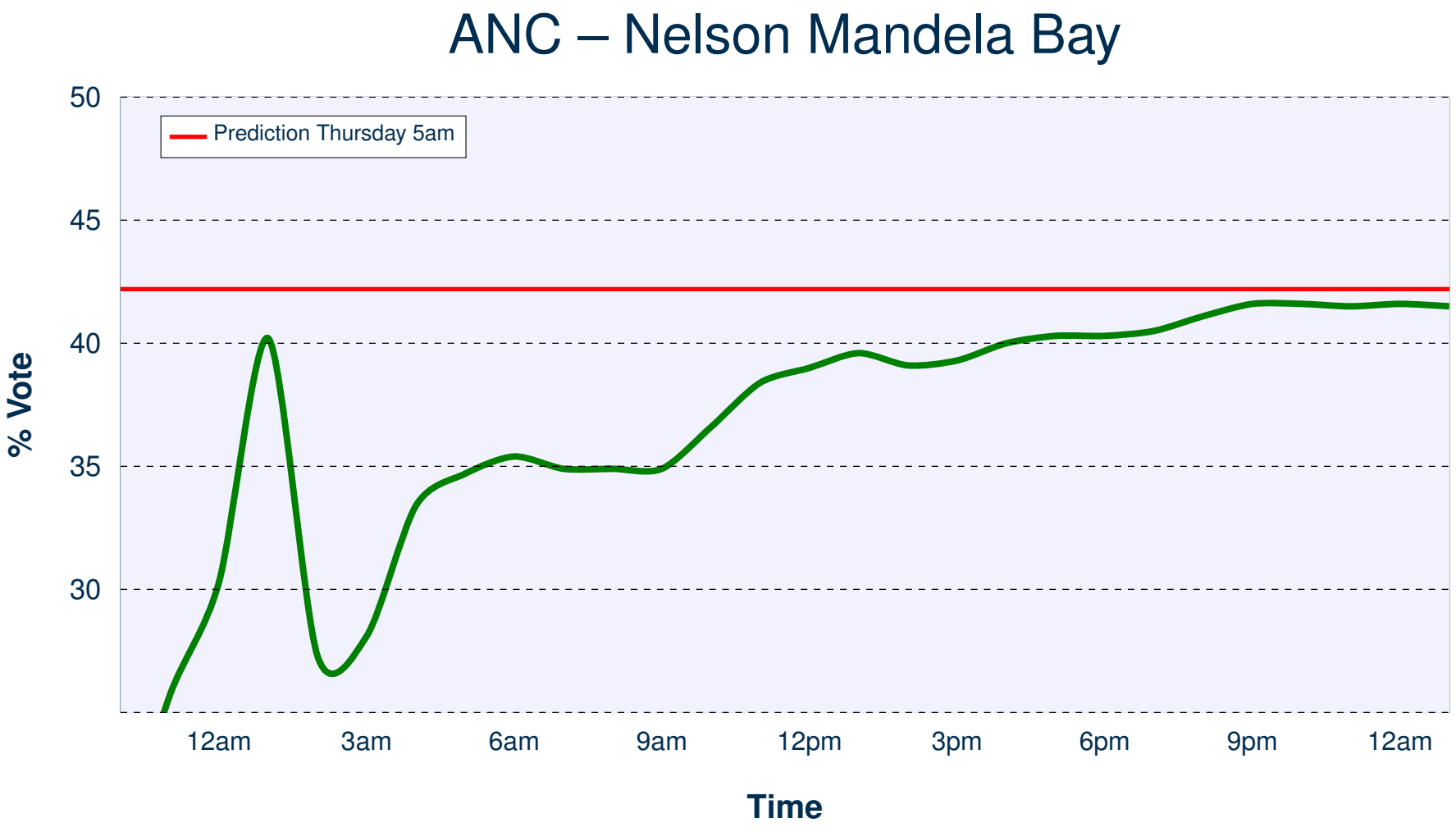
Nelson Mandela Bay

DA – Nelson Mandela Bay

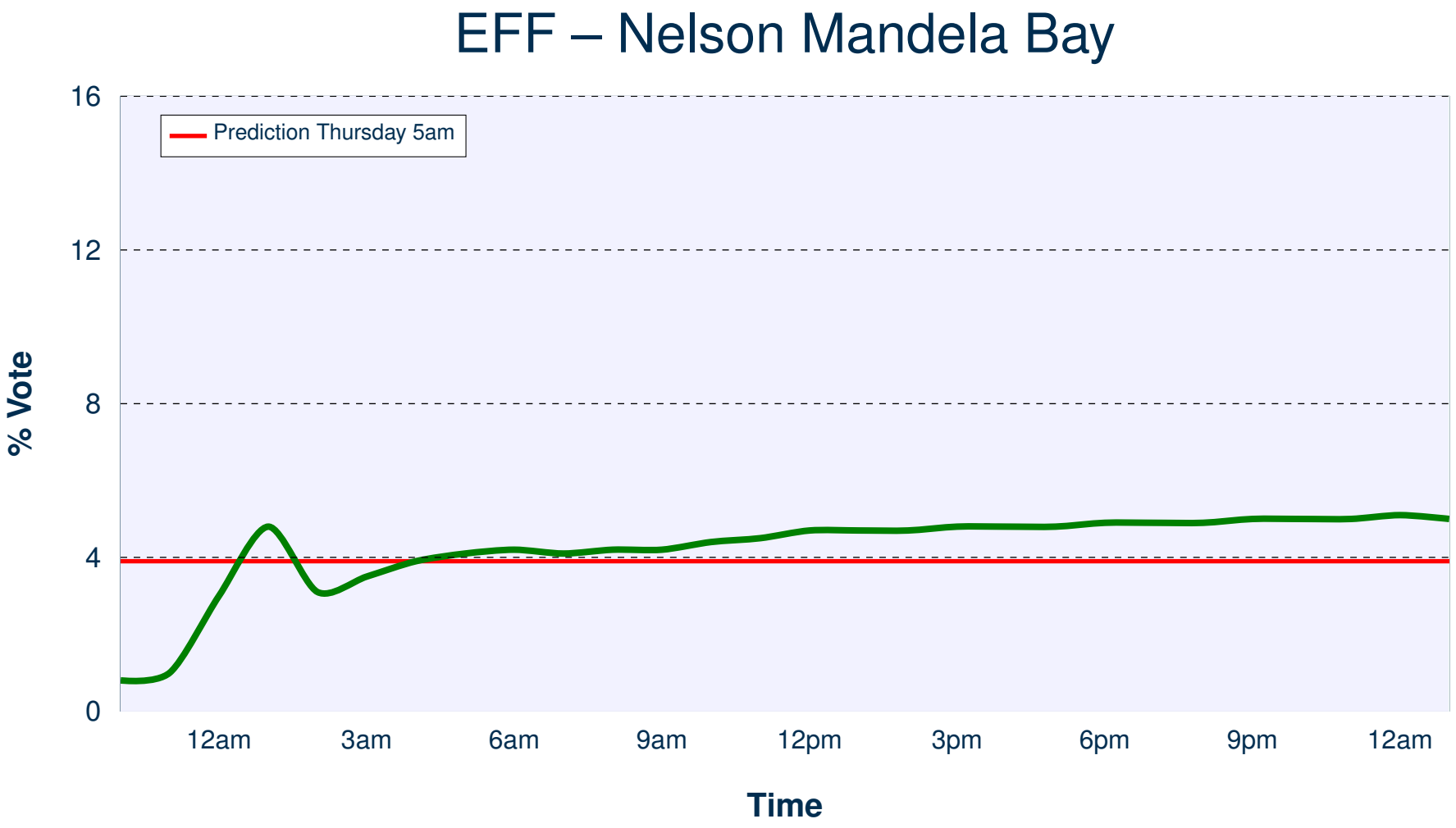


CSIR
our future through science

Nelson Mandela Bay

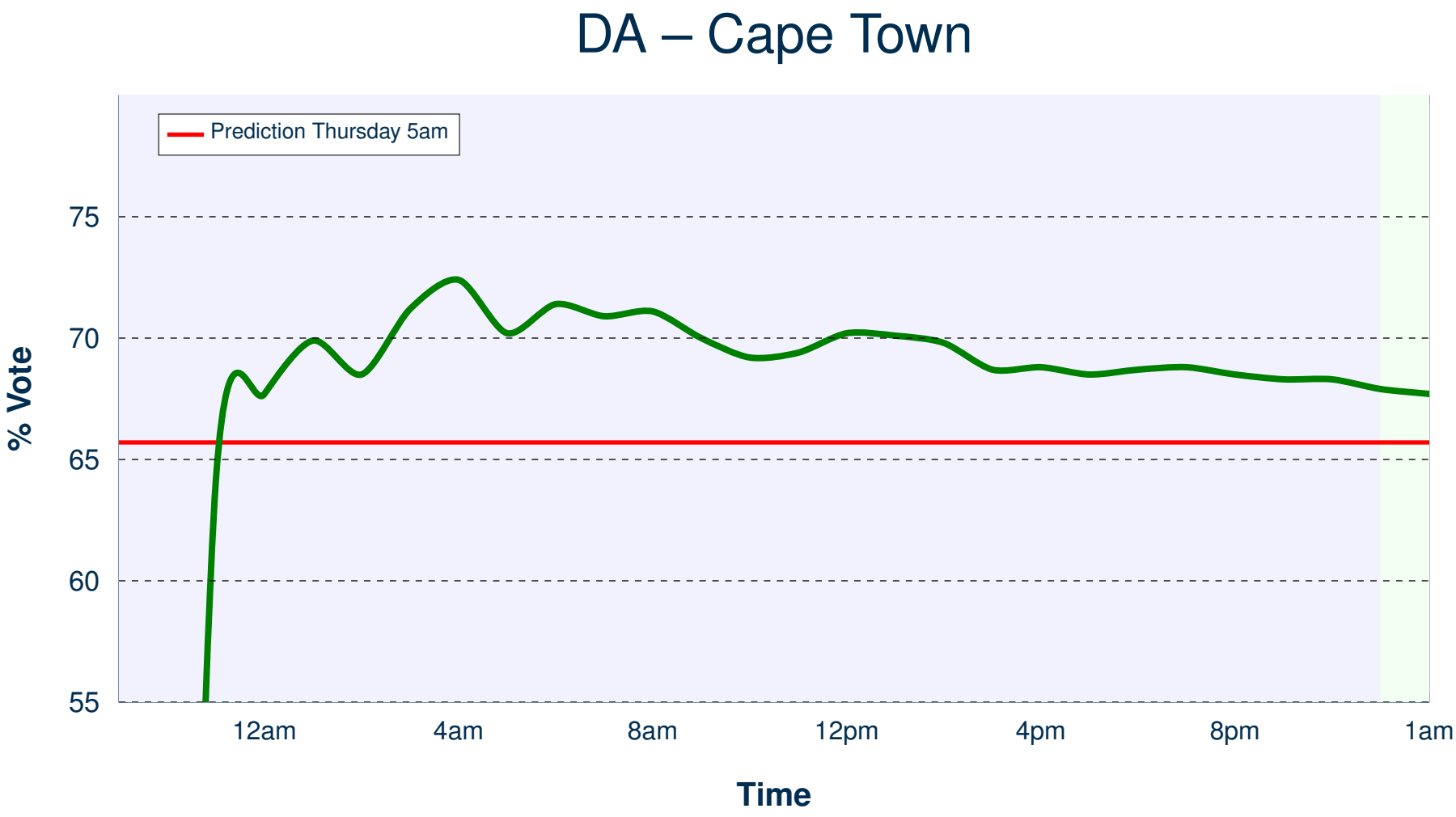


Nelson Mandela Bay



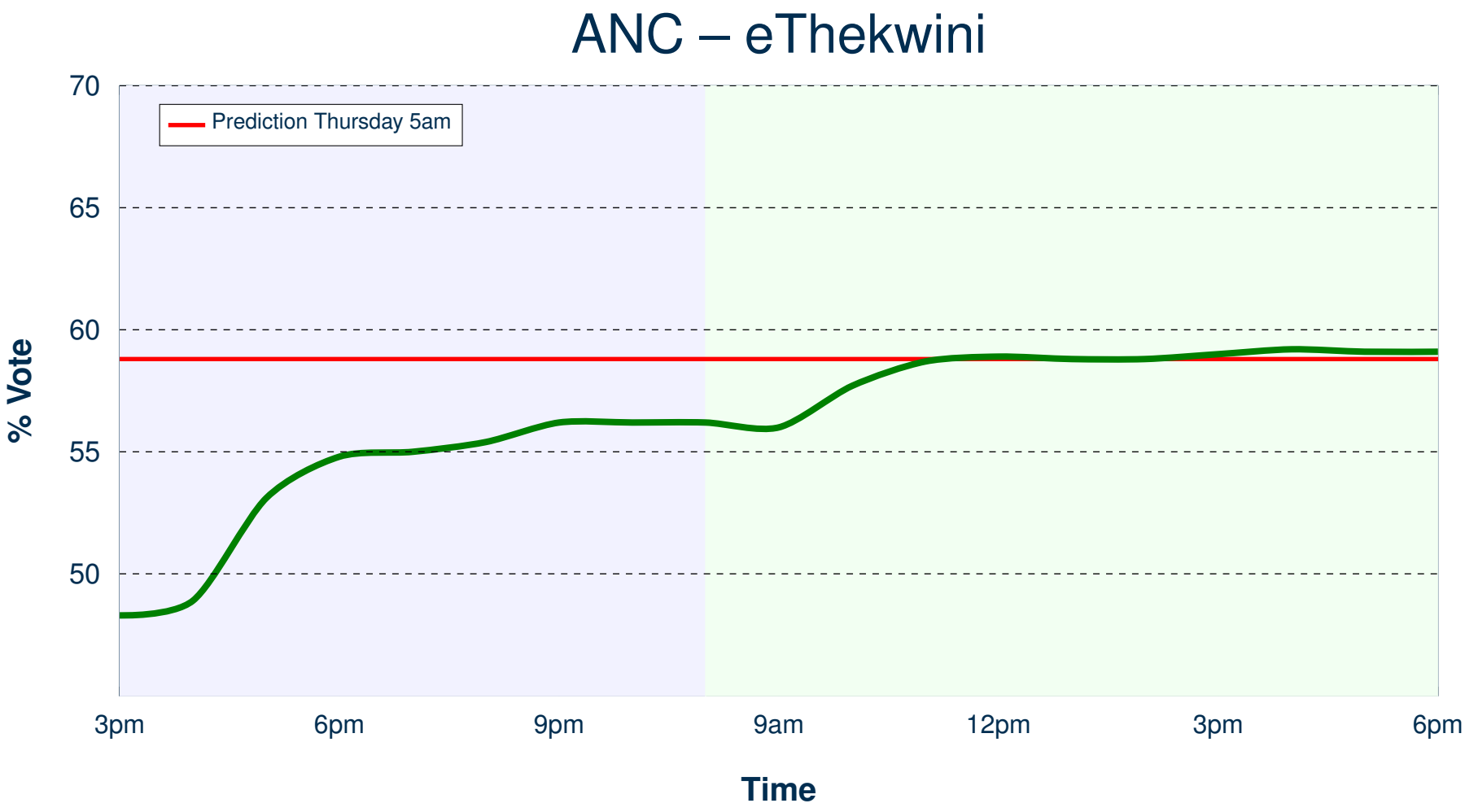
CSIR
our future through science

Cape Town



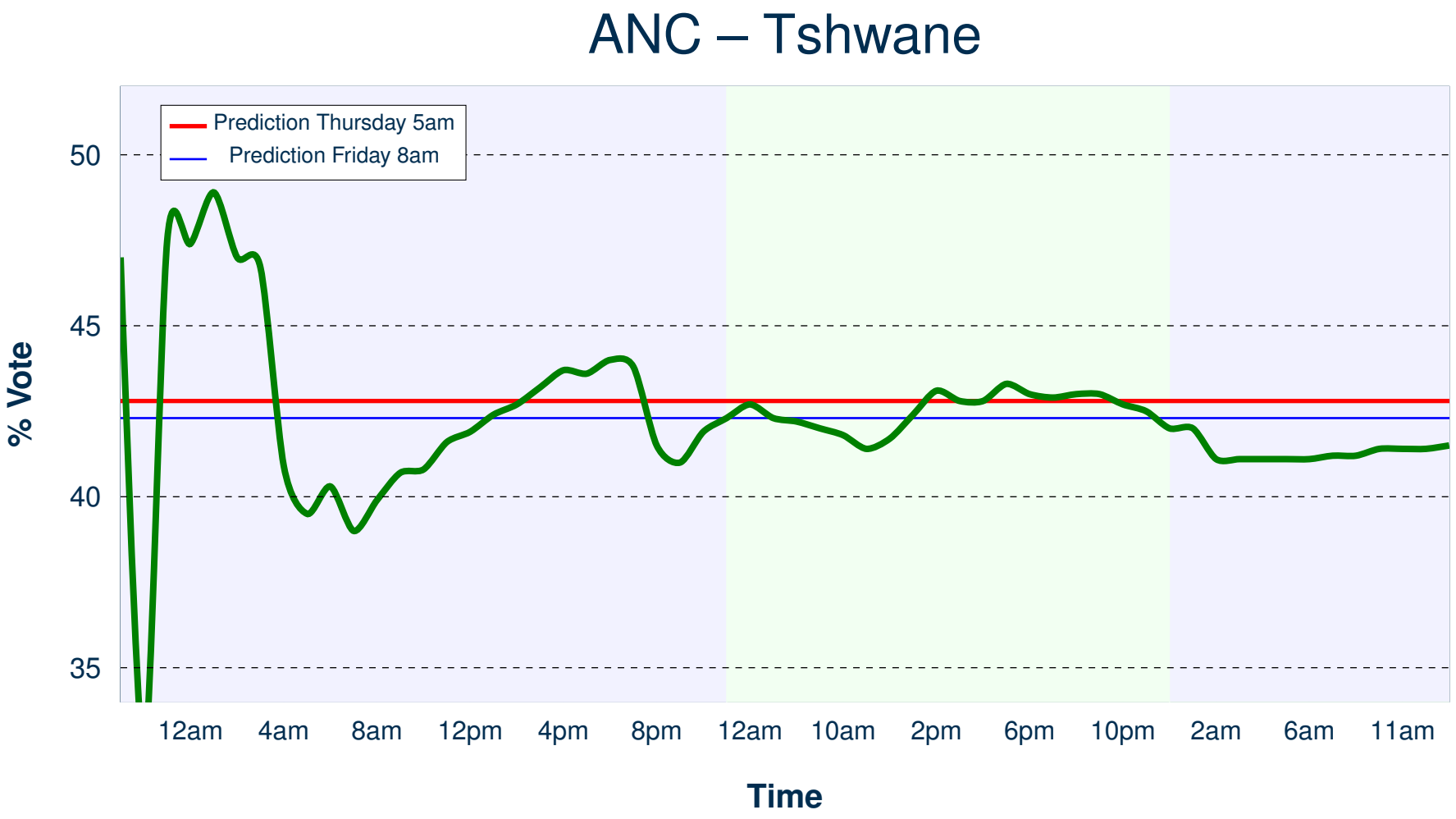
CSIR
our future through science

Ethekwini

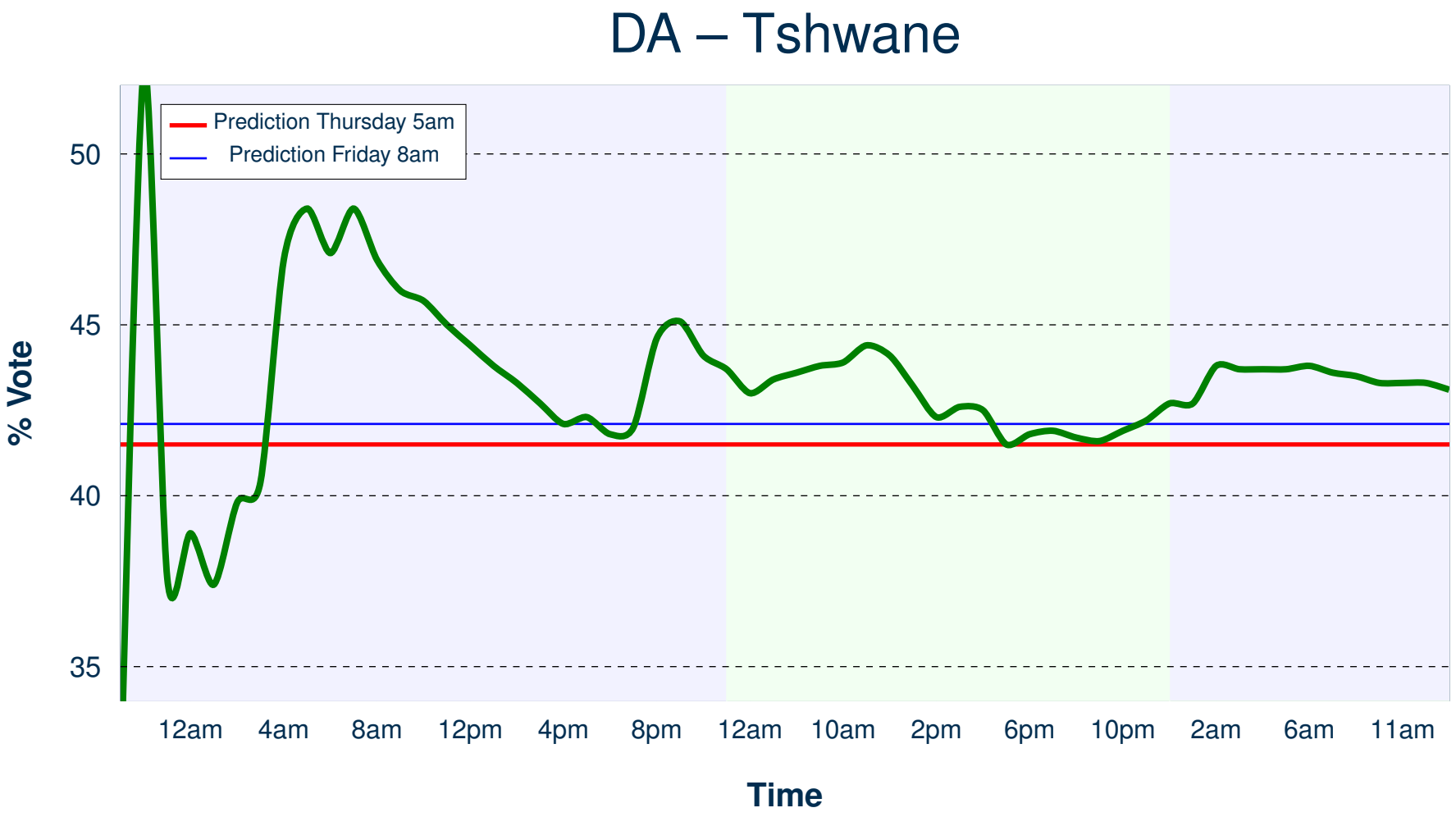


CSIR
our future through science

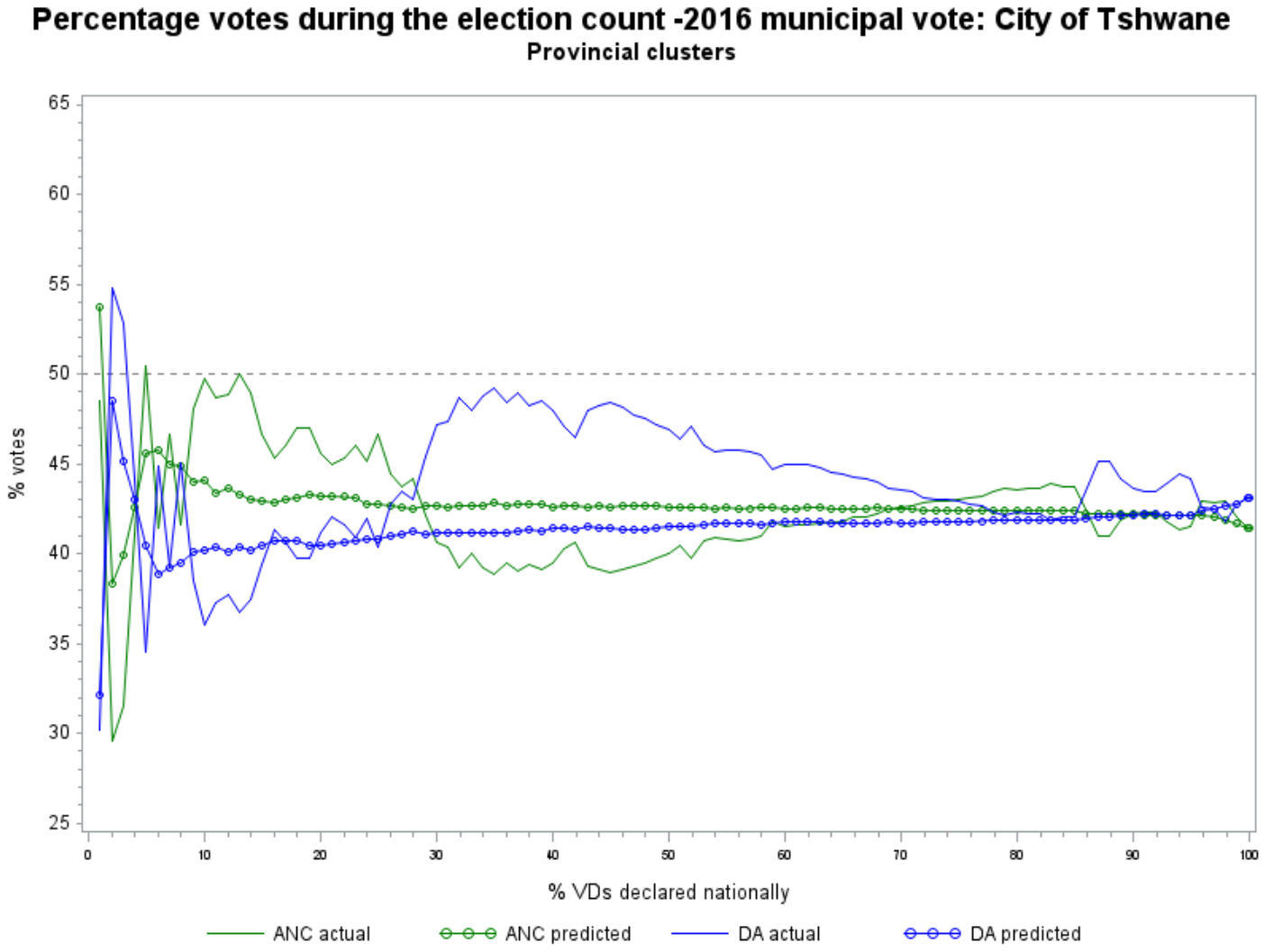
Tshwane



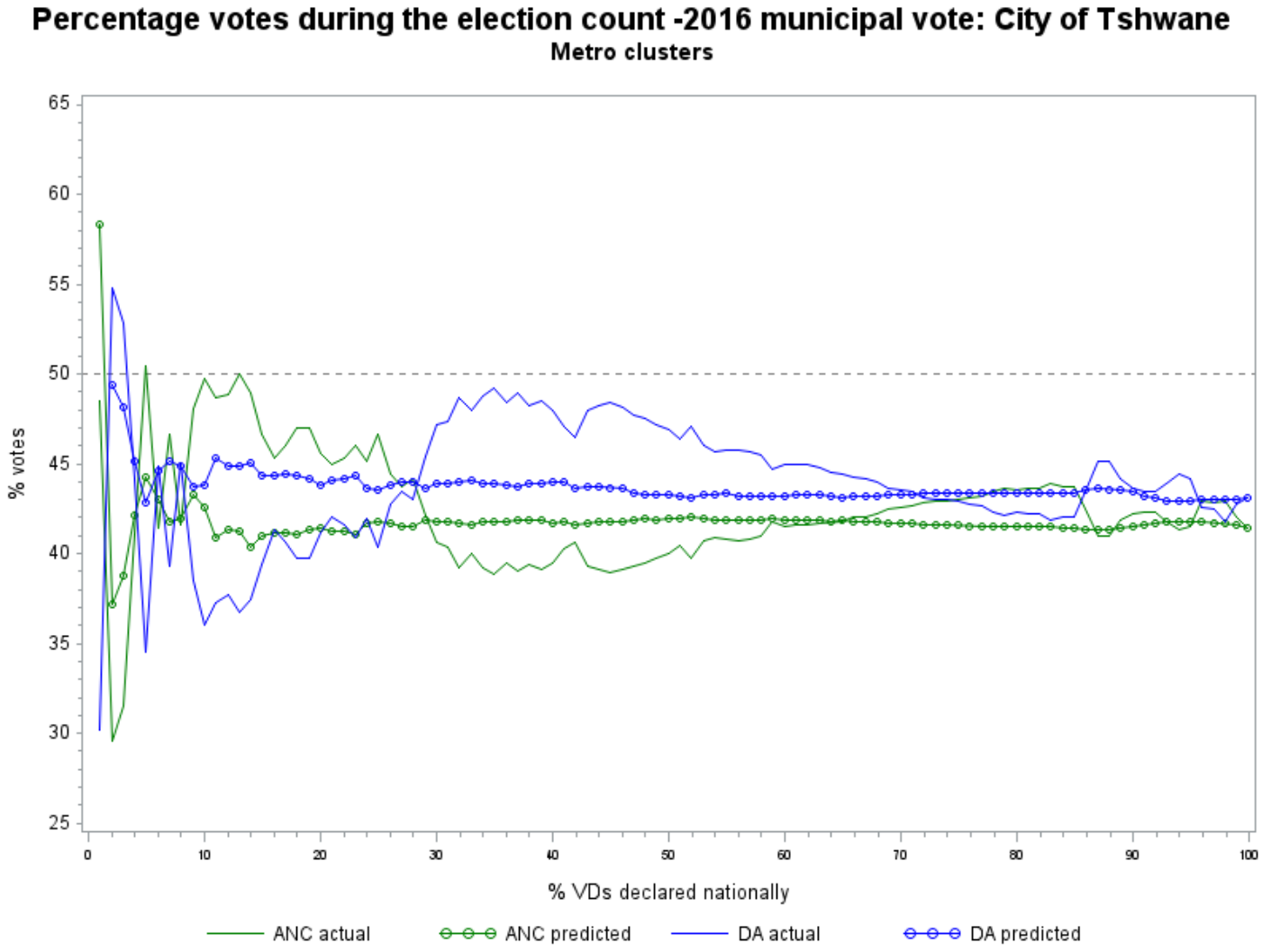
Tshwane



Cluster Comparisons



Cluster Comparisons



CSIR
our future through science

What Next?

- Running multiple models with different clustering options
- Improved diagnostics
- 2019 SA National Election
- 2016 US Election …
- 2020 UK Election