The Predictors of the Dutch Municipality Election Turnout

Universiteit Utrecht

ADS - Spatial Statistics and Machine Learning

Max van den Elsen, Sander Engelberts, Thom Venema

1. The Case

Dutch municipal council election turnout has been decreasing since the eighties, with a record low of 50.3% earlier this year [1]. In order to explain this trend, we investigated socio-economic and socio-demographic determinants of turnout prediction. Turnout research on these determinants has been done before [2, 3, 4, 5], but not on neighbourhood level for the full extent of The Netherlands.

2. The Data

Three main datasources were used to allow for this: a dataset containing the number of people that voted at each polling station [6], a dataset containing the locations of these stations [7], and lastly and most importantly a dataset by CBS (Dutch Statistics Bureau) [8] containing over 30 different socio- and economic variables for each neighbourhood. On this poster you will find an explanation of our methods, and our results displayed in various ways.

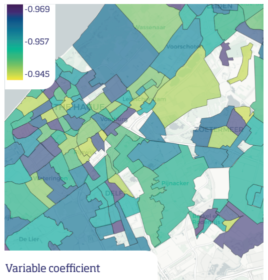
3. Methods

1. Confirm spatial correlation between the independent variables and the election turnout by estimating a simple linear regression model using spatial cross validation, and calculating its Moran's I metric. This we also did for spatial lag and error models.
2. Select independent variables due to the presence of multicollinearity. This was done by repeatedly calculating the VIF value of all variables, dropping the one with the highest VIF.
3. Fit a geographically weighted regression (GWR) model to explain how the relations are locally varying. For this, an adaptive kernel was utilized with a bisquare shape to weight the N closest polling stations based on their distance from a regression point. Further, its optimal bandwidth was determined using cross validation. The significance of the local variations of relationships were checked with a t-test, and the local explained variance R^2 by the model was mapped.
4. Fit a random forest regression on the data using spatial cross validation. This allowed for the inspection of feature importance based on permutation.

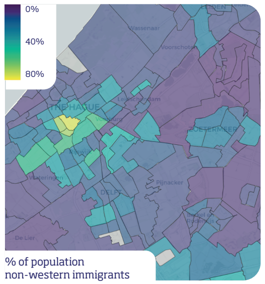
All these models were trained for different combinations of predictor variables. In specific these were grouped into socio-economic, socio-demographic, and spatial related predictors, and a composite of the most relevant

GWR - Coefficients

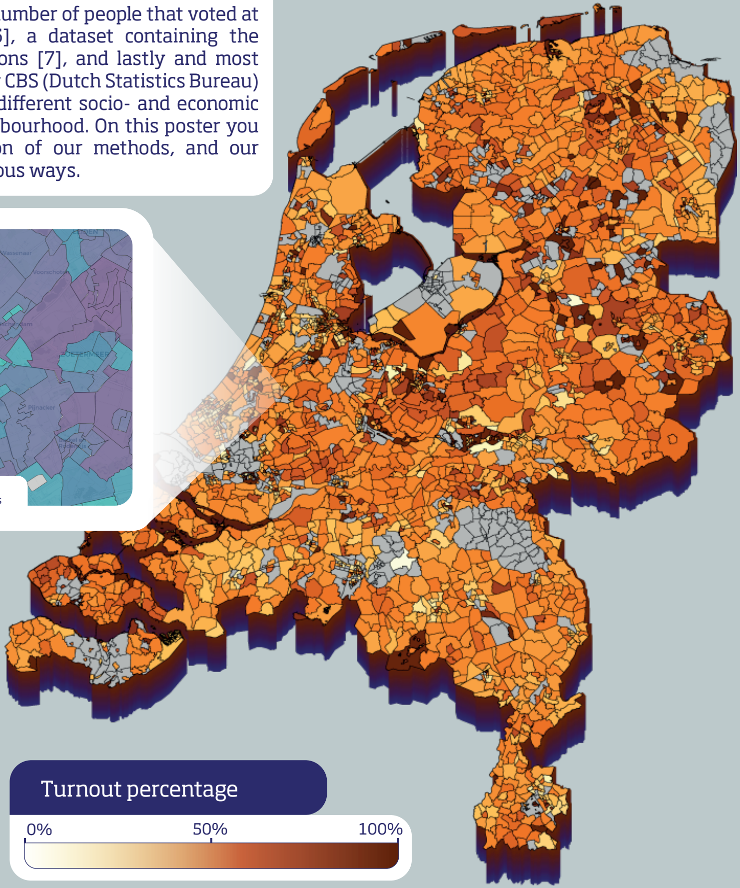
The maps represent Spatial distribution and coefficients of one of our variables, "%non-western immigrants", and its coefficient values, the latter being significant for (almost) all neighbourhoods in The Netherlands. While the **variable** is more clustered, the coefficients of this and other predictors vary over space, with nearby neighbourhoods often containing relatively similar values.



Variable coefficient



% of population non-western immigrants

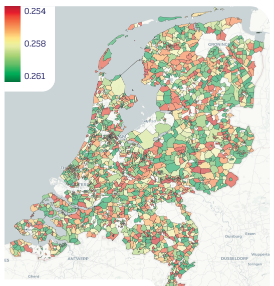


Turnout percentage

0% 50% 100%

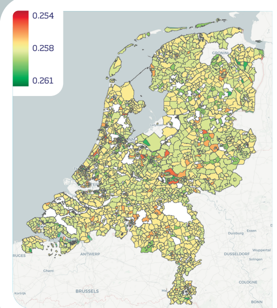
GWR - Local R^2

Local explained variance R^2 of GWR model based on all predictors. These values differ over space but are all very low (around 30%), also for the random forest models. This indicates a complex system of voter turnout on neighbourhood level, and the for further investigation at different levels and with different predictor variables.



Local explained variance (R^2) of GWR

Linear model - Residuals



Logistic regression residuals using an inverse distance weight matrix. These values vary over space, with some neighbourhoods having large negative or positive values. Also, a significant Monte-Carlo simulation of Moran's I p-value indicates the existence of spatial autocorrelation.

GWR - Coefficients

Median GWR coefficient values of predictors that are significant in (almost) all neighbourhoods, with negative values reducing the predicted turnout.

Variable	Median GWR coef.
% of population 0-15 y/o	-5.24 E-1
% of population 45-65 y/o	-1.09
% Non-western immigrant	-4.83 E-1
% Western immigrant	-9.54 E-1
Number of companies	-3.85 E-5
Average house value	5.30 E-4
% of population Low income	-5.47 E-1
Distance to voting station	-3.95 E-5

Discussion and Conclusion

Not all neighbourhoods contain turnout values due to not having had municipal elections or there being no polling stations in the respective area. Also, some neighbourhoods had more votes than eligible voter residents, which may be explained by the significant effects of distance to polling stations and number of companies, because people do not have to vote in their own neighbourhood. Further, multicollinearity and spatial autocorrelation was accounted for, improving the model performances with GWR, spatial error, and random forest models performing the best.

To conclude, municipal council 2022 election turnout behaviour is a complex system of which this study could not well describe the variance on neighbourhood level in the Netherlands. However, this can best be done using a combination of socio-economic, socio-demographic and location predictors. As most important factors to increase the voting turnout, the models showed that decreasing the distance to the nearest voting station, and focusing on low income households may have a positive effect.

Variable importance

Ranking of the variables most important for prediction, which our the result of Random Forest models.

1. Voting station Distance
Importance: 0.0079

2. High School within 3km
Importance: 0.0065

3. % of pop. low income
Importance: 0.0058

4. No. income earners
Importance: 0.0031

References

[1] Politieke redactie AD, "Historisch lage opkomst verkiezingen, minister onderzoekt oorzaak," 03 2022. https://www.ad.nl/politiek/historisch-lage-opkomst-verkiezingen-minister-onderzoekt-oorzaak a491fac6/?referrer=https%3A%2F%2Fwww.google.com%2F.
[2] N. Fiorino, N. Pontarollo, and R. Ricciuti, "Spatial links in the analysis of voter turnout in european parliamentary elections," Letters in Spatial and Resource Sciences, vol. 14, 04 2021.
[3] E. Mansley and U. Demšar, "Space matters: Geographic variability of electoral turnout determinants in the 2012 london mayoral election," Electoral Studies, vol. 40, pp. 322-334, 2015.
[4] L. Manoel, A. C. Costa, and P. Cabral, "Voter turnout in portugal: A geographical perspective," Papers in Applied Geography, vol. 8, no. 1, pp. 88–111, 2022.
[5] D. Koerntjes, "Het verschil in opkomst tussen tweede kamerverkiezingen en gemeenteraadsverkiezingen," 03 2022. https://www.cbs.nl/nl-nl/longread/statistische-trends/2022/het-verschil-in-opkomst-tussen-tweede-kamerverkiezingen-en-gemeenteraadsverkiezingen.
[6] K. (Rijk), "Verkiezingsuitslagen gemeenteraad 2022," 2022. https://data.overheid.nl/dataset/verkiezingsuitslagen-gemeenteraad-2022#panel-resources.
[7] ckan.dataplatform, "Stembureaus tweede kamerverkiezingen 2021," 2021. https://ckan.dataplatform.nl/dataset/stembureaus-tweede-kamerverkiezingen-2021/resource/eb2c1546-7f8d-41d4-9719-61b53b6d2111.
[8] PDok, "Cbs wijken en buurten 2019 (wfs)," 2019. https://www.pdok.nl/geo-services/-/article/cbs-wijken-en-buurten#68b8f7e2848a93f685cdf399e3fb2ef2.