

Urban Growth Modeling in the Denver Area

Johnathan Clementi and Henry Feinstein

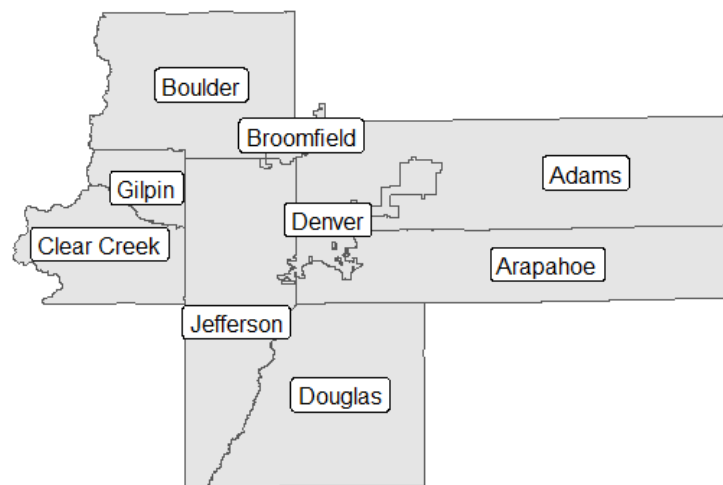
CPLN 675 Homework 5

May 10, 2022

Introduction

This report presents the findings of an urban growth model for the Denver Regional Council of Governments (DRCOG), a group of nine counties in the Denver area. This area is experiencing consistent population growth which has contributed to sprawl and the destruction of vulnerable landscapes such as forests and wetlands. In order to better understand how these issues might be avoided in the future, the model presented here predicts the quantity and location of development over the next ten years. These findings can be leveraged in DRCOG's upcoming comprehensive planning process to help mitigate sprawl.

Study Area Counties



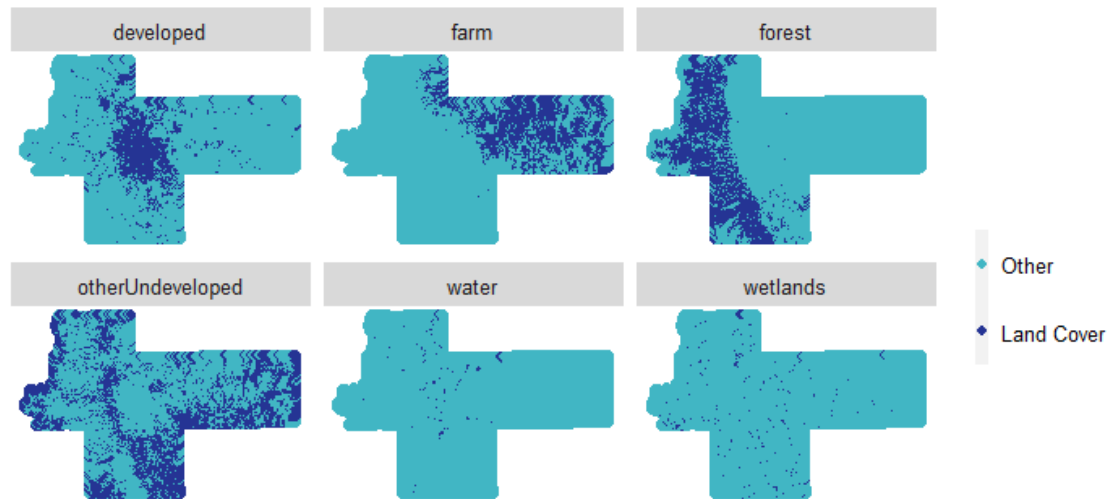
Model Structure and Key Variables

In this report, new development is modeled by dividing the DRCOG region into evenly-sized cells in a grid or “fishnet” pattern. Each cell is given a value for each of the variables included in the model. The model’s dependent variable is whether a given cell changed land cover type (which, for our purposes, almost always means the cell changed from “undeveloped” to “developed”) between two time periods—2009 and 2019. The model returns a probability that a given cell changed land cover and uses the following independent variables to make that prediction:

Land cover in 2009: The existing land cover of a cell is a strong indicator of whether it is likely to change status. The model includes variables indicating whether a cell was wetlands, farm, forest, or other undeveloped land in 2009.

Land Cover Types, 2008

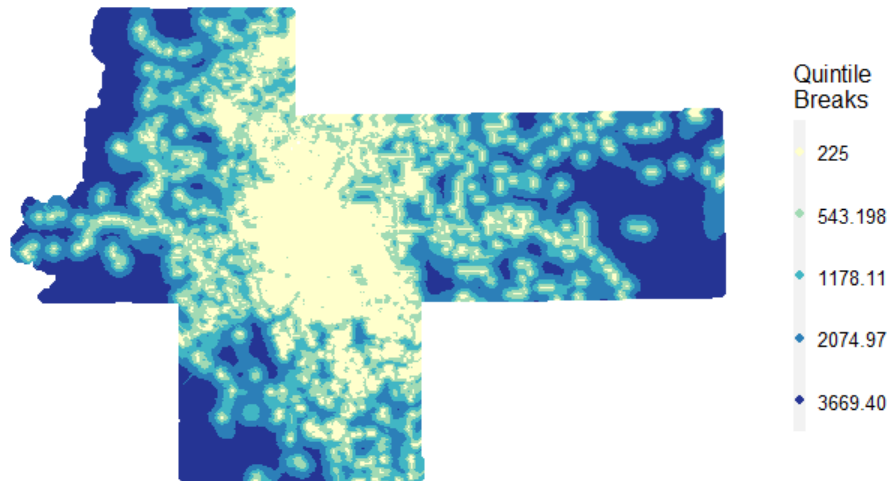
As fishnet centroids



Spatial lag of development: “Spatial lag” is a term that refers to the relationship between a cell and its neighboring cells. The spatial lag of development demand for one cell captures the development demand of the cells nearby.

Spatial Lag to 2008 Development

As fishnet centroids



Population change: We pulled population at the Census tract level for both 2009 and 2019 and distributed it proportionally among the fishnet cells. The model includes population change per cell between the two years, anticipating that population increases lead to more development.

Population Change, DRCOG: 2009-2019

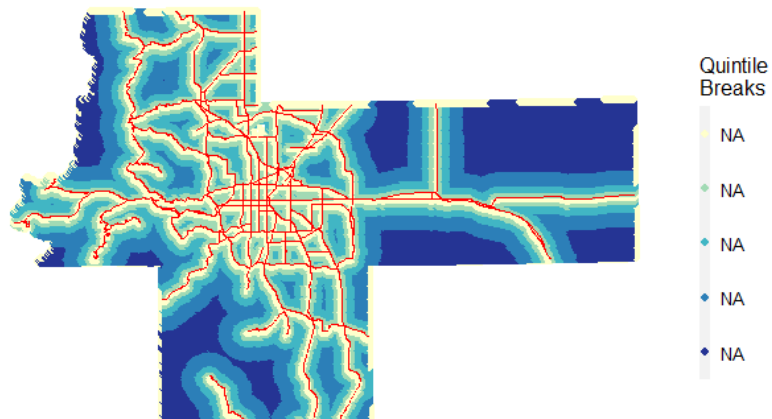
Change per fishnet cell



Distance to highways: The model includes the distance to the closest highway for each cell, given that development often appears around existing infrastructure.

Distance to Highways

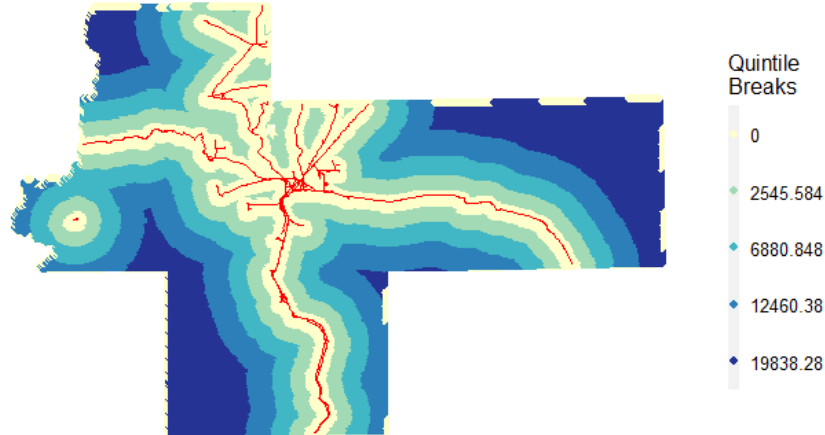
As fishnet centroids; Highways visualized in red



Distance to rail: Similar to highways, the model includes the distance of each cell to the closest rail line.

Distance to Rail

As fishnet centroids; Rail visualized in red



Model Development and Validation

Using the variables presented above, our model predicts the likelihood of a given cell changing land cover category between 2009 and 2019. The next step is to validate the model by comparing its predictions to the actual land cover changes over that time period. This process gives a sense of the model's overall accuracy and generalizability to different scenarios.

```
Call:
glm(formula = lc_change ~ wetlands + forest + farm + otherUndeveloped +
    lagDevelopment + pop_Change + dist_to_highway + dist_to_rail,
    family = binomial(link = "logit"), data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0136  -0.6530  -0.2824  -0.0046   4.9983

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.939738809  0.033797658  27.805 < 0.0000000000000002 ***
wetlands1    -0.424841509  0.102755699  -4.134  0.00003557573 ***
forest1      -0.817466145  0.055222820 -14.803 < 0.0000000000000002 ***
farm1        -0.677756734  0.054638042 -12.404 < 0.0000000000000002 ***
otherUndeveloped1 -0.698647906  0.045946502 -15.206 < 0.0000000000000002 ***
lagDevelopment -0.000960015  0.000024631 -38.977 < 0.0000000000000002 ***
pop_Change    0.002256203  0.000377415   5.978  0.00000000226 ***
dist_to_highway -0.000108450  0.000005521 -19.643 < 0.0000000000000002 ***
dist_to_rail   0.000013074  0.000002480   5.271  0.00000013551 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

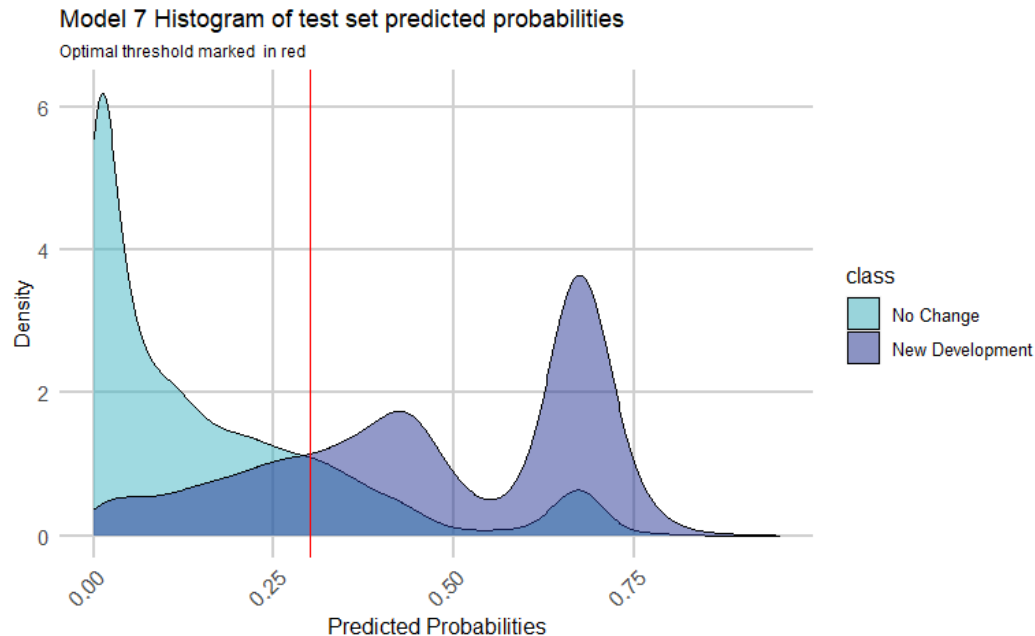
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 36419  on 33336  degrees of freedom
Residual deviance: 25724  on 33328  degrees of freedom
AIC: 25742

Number of Fisher Scoring iterations: 7
```

The table above presents the regression results of the model, where every independent variable is a statistically significant predictor. One key takeaway is that cells with wetlands, forest, and

farm in 2009 were less likely to become developed in 2019; this makes sense given that many of these “natural” cells are in parks or otherwise difficult to develop. Population increases and proximity to existing development and highways, on the other hand, are associated with a greater likelihood of development.



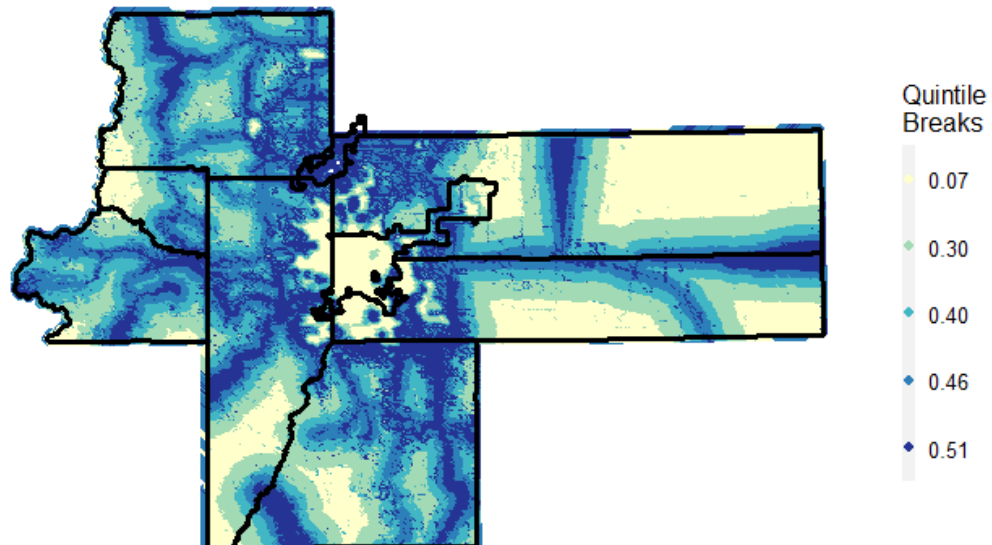
To evaluate the model’s accuracy, a threshold must be decided at which the probability of a cell being developed is high enough to be considered an official “developed” prediction. The chart above plots the probabilities of development for cells that *actually* developed between 2009 and 2019 (in dark blue) and cells that didn’t (in light blue). We can see that the model predicts the two categories differently, giving cells that actually developed generally higher probabilities. We set the prediction threshold at 0.3 given that this is where the two categories intersect on this plot. Using this threshold, we can calculate that the model performs well, achieving 80% accuracy overall. The model, however, is better at predicting when cells will not change in land use (Specificity: 81%) versus predicting when cells will change in land use (Sensitivity: 76%).

To get a sense of how the model performs across different areas, an approach called “spatial cross-validation” is used. This method cycles through different areas—in our case, the nine DRCOG counties—holding one area out and training the model’s predictions on the remaining data. The model then predicts for the holdout region; if it predicts well across all areas, the model is considered generalizable.

county	Observed_Change	Sensitivity	Specificity	Accuracy
Adams	3118	0.74	0.87	0.84
Arapahoe	2303	0.81	0.83	0.83
Boulder	1964	0.71	0.78	0.77
Broomfield	292	1.00	0.12	0.77
Clear Creek	573	0.58	0.92	0.88
Denver	1416	0.91	0.19	0.65
Douglas	2781	0.75	0.78	0.77
Gilpin	336	0.56	0.88	0.83
Jefferson	3105	0.78	0.75	0.76

The table above shows that the model is consistent across all counties except Denver. This makes sense given that Denver is uniquely dense and urban while the other counties are suburban or rural and more similar to each other than to the large city.

Development Demand in 2030: Predicted Probabilities



The model's predictions are unsurprising, following the general patterns of development seen in the training period: highways, rail, and existing development strongly inform where new development is likely to happen. In the second scenario, the model predicts slightly more development south of Denver where the new highway was inserted.

Predicting for 2029: Two Scenarios

The trained model can now be used to predict which cells will become developed in 2029. Here, two scenarios are presented: one in which population increases based on projections published by DRCOG; and a second with the same population projections and an added infrastructural change, which is a new beltway connecting towns to the south of Denver including Roxborough Park, Castle Pines, and Parker. This infrastructural project was chosen given the lack of existing connections between these towns and the likelihood of development in this area in the next ten years. The next section details the methodology for deciding on the proposed locations of development under these two scenarios.

Allocation

Merely predicting which cells will develop has limited use, given that some cells cannot or should not be developed on. This final section presents an allocation schema which takes into account the model's results along with other factors to suggest which fishnet cells should be developed to accommodate the increasing population in the coming decade. The number of cells to allocate was determined by dividing the 2019 population by the number of developed cells in 2019, then dividing the projected change in population in 2029 by that "per-cell" rate. This calculation resulted in needing to allocate 1,253 cells (5,824 acres) to new development. To determine which cells should be allocated, we normalize, weight, and add the following values to create an allocation score:

Development demand: The model's probability that a cell will be developed.

Distance to sensitive regions: The distance of each cell to forest or wetland. The score seeks to maximize distance from these vulnerable natural areas.

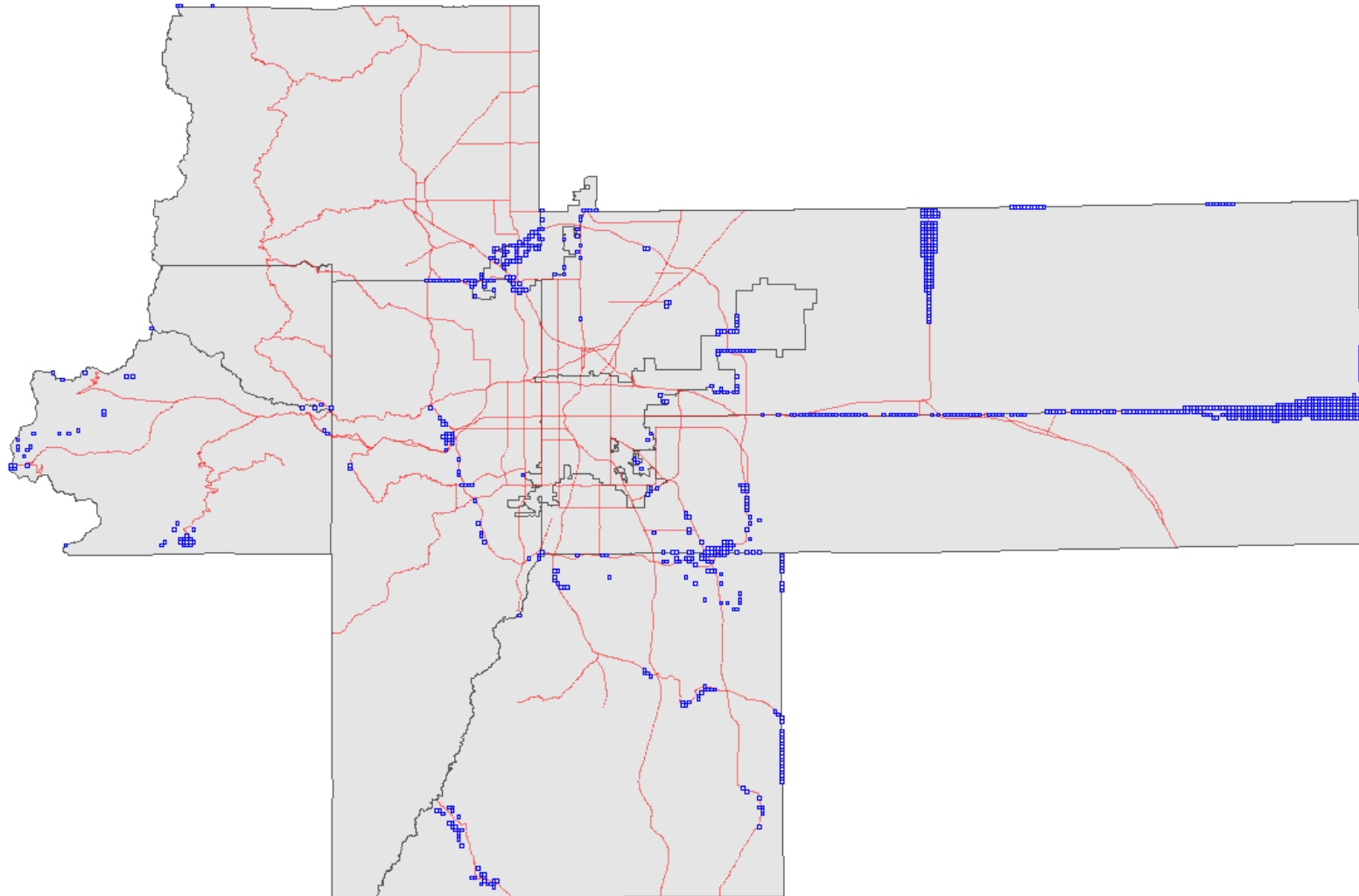
Distance to highways: The distance of each cell to existing highways. The score seeks to minimize distance to highways, since it is efficient to build new development near existing infrastructure.

Spatial lag of development: The distance of each cell to the closest already-developed cell. Similar to highways, the score seeks to minimize the distance between new development and existing development.

After the cells are scored, cells that are "impossible" to develop because they are sensitive land or already developed are removed. The following maps present the remaining top 1,253 cells by allocation score for the two planning scenarios:

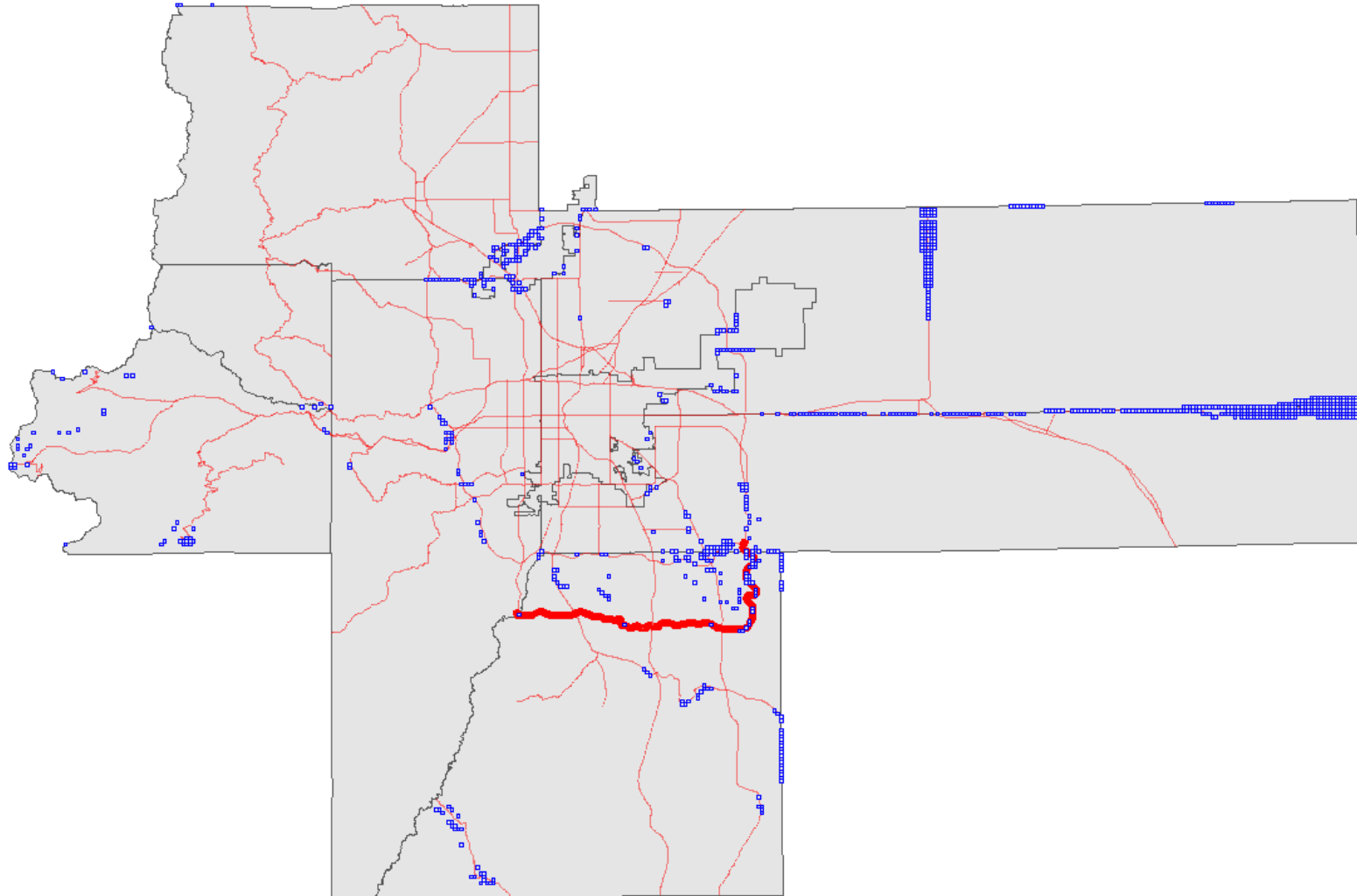
Demand-side Allocation

Land development cells are allocated based on development trends alone.
Roads in red



Demand and Supply-side Allocation

Land development cells are allocated based on development trends and proposed infrastructure
Roads in red, Proposed road highlighted



The model clearly puts a high value on developing near highways, as the strip of new development to the east shows. The model also suggests a lot of development should happen in and around Denver on cells surrounded by existing development. In the second scenario, the model suggests allocating slightly more cells near the new highway compared to the first scenario. In general, the model's results resemble the patterns of development seen in the Denver area today.

Conclusion

Unsurprisingly, the model presented here reflects the history of development in the Denver metro area. If current practices continue, Colorado is likely to see a worsening problem of sprawl around its capital, threatening to disrupt ecological systems and encourage unsustainable, car-dependent living. This report can be used to inform how local governments might leverage zoning and policy strategies to reduce sprawl while accommodating future population growth.