
COMMONWEALTH of VIRGINIA

Virginia Conservation Vision Development Vulnerability Model 2022 Edition

Virginia Department of Conservation and Recreation
Division of Natural Heritage
Natural Heritage Technical Report 22-13
June 2022

**Virginia Conservation Vision
Development Vulnerability Model
2022 Edition**

Prepared by:

Kirsten R. Hazler and David N. Bucklin
Virginia Department of Conservation and Recreation
Division of Natural Heritage
600 East Main Street
Richmond, Virginia 23219

June 2022

Suggested citation:

Hazler, K.R. and D. N. Bucklin. 2022. Virginia Conservation Vision Development Vulnerability Model, 2022 Edition. Natural Heritage Technical Report 22-13. Virginia Department of Conservation and Recreation, Division of Natural Heritage, Richmond, Virginia.

Abstract

The purpose of the Virginia ConservationVision Development Vulnerability Model is to quantify the risk of conversion from greenspace (natural, rural, or other open space lands) to urbanized or other built-up land uses. The model output is a raster dataset in which the relative vulnerability of lands ranges from 0 (least vulnerable) to 100 (most vulnerable). Vulnerability values are not probabilities, but should be interpreted as a relative measure of development potential. Conservation lands on which biodiversity preservation is believed to be the primary goal are considered undevelopable and are coded with the value -1, while areas in which development has already occurred are coded 101.

The basis of the model is a Random Forests machine-learning model, used to relate a suite of predictor variables representing conditions at an initial time to outcomes (developed or not) a decade later. The predictor variables can be grouped into three categories based on their spatial focus: (1) local site characteristics, (2) neighborhood characteristics, and (3) travel time or distance to development “attractors”. The sampling frame was limited to areas that were undeveloped at the initial time but that could potentially be developed in the future.

To train and test the model, predictor variables representing conditions in the year 2006 were connected to outcomes in 2016, with the data split into independent subsets used for training and testing. Once the prediction model was finalized, variables representing conditions in 2019 were used to predict the relative potential for development by the year 2029. To produce the final vulnerability map, raw prediction values were adjusted to reflect current conservation and development status.

This model is one of several in a suite of conservation planning and prioritization models developed by the Virginia Natural Heritage Program and partners, known collectively as Virginia ConservationVision. It is intended for use in conjunction with other data to help target lands for protection. The model can also serve as an input for simulating future land cover change and its consequences under different planning scenarios.

Table of Contents

Introduction.....	1
Methods.....	2
Spatial data processing.....	2
Input data	3
Predictor variables	3
Sample data.....	7
Modeling approach	8
Random Forests	8
Exploratory phase	9
Model testing and prediction	10
Results.....	11
Discussion.....	12
Model interpretation.....	12
Limitations and opportunities	13
Model applications.....	14
Tables	16
Figures.....	23
Maps.....	31
Acknowledgements.....	54
References.....	54

Introduction

As human populations and demand for resources expand, natural areas and rural lands are increasingly threatened by encroaching development. The mission of the Virginia Department of Conservation and Recreation (DCR), as the state's lead natural resource conservation agency, is to provide opportunities that encourage and enable people to enjoy, protect, and restore Virginia's natural and cultural resources (Virginia DCR Staff 2016). The Virginia Land Conservation Foundation (VLCF) provides state funding to purchase or establish conservation easements on various lands of conservation concern (Virginia DCR Staff n.d.). Given limited funds, it is essential to have a means of prioritizing lands worthy of preservation. As part of its work, DCR's Division of Natural Heritage (DNH) and its partners develop and maintain a suite of geospatial models intended to guide strategic land conservation and management decisions. This suite of models is known as Virginia ConservationVision. The models under the ConservationVision umbrella address a variety of conservation issues and priorities, including natural landscapes, watersheds, agriculture, forestry, cultural resources, rare species richness, outdoor recreation, and development vulnerability.

The purpose of the Virginia Development Vulnerability Model is to quantify, statewide, the relative risk of conversion from greenspace (natural, rural, or other open space lands) to urbanized or other built-up uses. It is intended for use in conjunction with other datasets to help strategically target unprotected lands for conservation action. Development vulnerability specifically addresses the time-sensitive nature of conservation decisions. If the relative potential for development is not taken into account, conservation actions may be initiated too late to preserve ecologically valuable areas before damage is done.

As part of their 2008 Resource Lands Assessment, the Chesapeake Bay Program defined vulnerability as a function of suitability for development and proximity to growth “hot spots” (CBP, 2008). Their subsequent work has resulted in a dataset representing predicted development pressure through the year 2055 (CBP, 2020). Although their focus is on the Chesapeake Bay watershed, the later dataset fully covers Virginia. Similar recent modeling efforts covering Virginia include urbanization predictions for the Northeast (McGarigal et al. 2018) and Southeast (Terando et al. 2014) regions of the U.S.

The model described in this report replaces a previous version released in 2015 (Hazler et al. 2016), which in turn replaces an earlier edition (Virginia DCR Staff 2008). Significant differences in model inputs and methodology render direct comparisons between outputs from these model editions inadvisable. With the release of the current edition, the previous edition should be considered obsolete. Improvements since the previous edition include updated input datasets; identification and development of a broad suite of predictor variables; and a formal inductive, machine-learning approach to estimating vulnerability.

Methods

The basis of the Development Vulnerability Model is a Random Forests machine-learning model (Breiman 2001), used to relate a suite of predictor variables representing conditions at an initial time to outcomes (developed or not) a decade later. The sampling frame was limited to areas that were undeveloped at the initial time but that could potentially be developed in the future. To train and test the model, predictor variables representing conditions in the year 2006 were connected to outcomes in 2016, with the data split into independent subsets used for training and testing. Once the prediction model was finalized, variables representing conditions in 2019 were used to predict the relative likelihood of development by the year 2029. To produce the final map of vulnerability, raw prediction values were adjusted to reflect current conservation and development status.

Spatial data processing

ArcGIS Pro software (ESRI 2021) was used for most spatial data processing and development of model inputs. In addition to using standard ArcGIS tools, we developed a set of custom Python scripts¹ to carry out necessary procedures. Since calculations for several predictor variables include distance or neighborhood analyses which are sensitive to boundary effects, we used a processing extent equal to a 50-mile buffer around the state border of Virginia when developing those datasets. For all other datasets, we used a processing extent equal to a 1-mile

¹ https://github.com/VANatHeritage/ConsVision_DevVulnModel

buffer around the border of Virginia.

Many of the model's inputs were developed using datasets from the National Land Cover Database (NLCD), 2019 edition (Dewitz & USGS, 2021; Yang et al., 2018), which include standard 30-m resolution Land Cover, Percent Imperviousness, and Impervious Descriptor raster datasets at 2- to 3-year intervals from 2001 to 2019. Because of this, we used the NLCD raster's coordinate system for all products, and the raster cell size (30-meter) and alignment for all raster datasets, including the final model.

Input data

Predictor variables

Predictor variables were developed as raster datasets with 30-m resolution. We compiled a list of 39 variables to potentially include as predictors, but many were eliminated from the final model (Table 1). In compiling the list, we drew heavily on previous models of urban growth (Irwin et al. 2003, Sohl and Sayler 2008, Jantz et al. 2010, Westervelt et al. 2011, Meentemeyer et al. 2013, Chaudhuri and Clarke 2013, Terando et al. 2014), while taking into account the practicality of deriving the variables from existing, freely available, statewide datasets. The predictor variables can be grouped broadly into three categories based on their spatial focus: (1) local site characteristics, (2) neighborhood characteristics, and (3) travel time or distance to development “attractors”.

Some predictor variables were assumed to be static, in that they were not expected to change significantly over the time frame of interest. Other variables, depending on data from different time periods, were dynamic. For each dynamic variable, two rasters were developed to reflect conditions in both 2006 (to train and test the model) and 2019 (to predict forward). A subset of dynamic variables were also multi-temporal, requiring data from two time periods to reflect changes over a 5- or 6-year time span ending in the nominal date. Multi-temporal variables include distance to urban-growth hotspots and distance to new roads.

Local Site Characteristics

Local site characteristics reflect conditions at a specific location represented by a single raster cell. Elevation and slope were derived from a digital elevation model (USGS, 2017). Land cover development cost was a unitless value between 0 and 5, assigned to land cover types based

on the presumed relative ease of conversion to development (Table 2). The “land protection multiplier” variable was derived from back-dated versions of the Virginia Conservation Lands Database (VCLD; Virginia DCR Staff n.d.), calculated based on biodiversity management intent and legal protection status, as it was in the 2015 model edition (Hazler et al. 2016).

Soil suitability ratings for dwellings with basements, dwellings without basements, small commercial buildings, and local roads and streets were derived from the Gridded Soil Survey Geographic Database (Soil Survey Staff 2020) using the Soil Data Management Toolbox (NRCS, 2017). Additional steps were used to average the four suitability ratings, integerize and rescale the values, and fill in data gaps, resulting in values representing soil suitability for development that ranged from -10 (least suitable) to 10 (most suitable), except on already developed lands (11) and open water (-11).

Neighborhood Characteristics

Neighborhood characteristics reflect conditions surrounding a focal cell, and are calculated as statistical summaries of raster values within a specified surrounding neighborhood. Assuming that the influence of neighboring cells on a focal cell decays with distance, Meentemeyer et al. (2013) derived variations of a “development pressure” (p'_i) variable that is a function of neighboring developed cells and their distance from the focal cell:

$$p'_i = \sum_{k=1}^{n_i} State^k / d_{ik}^\gamma \quad \text{Equation 1}$$

where $State^k$ is a binary variable that indicates whether the k th neighboring cell is developed (1) or undeveloped (0), d_{ik} is the distance between the k th neighboring cell and the current cell i , γ is a coefficient that controls the influence of distance between neighboring cells and cell i , and n_i is the number of neighboring cells within a specific range with respect to cell i .

We extended the use of the above formula from Meentemeyer et al. (2013) to derive neighborhood variables quantifying the neighborhood influence of three different land cover types: water, wetlands, and impervious cover. However, in our formulation for impervious cover, $State^k$ was a value between 0 and 100 (i.e., percent impervious), rather than binary. We experimented with several values for the distance-decay parameter, gamma (γ), but settled on $\gamma = 0.5$ for the final model. For each of the three land cover types, we calculated influence using

two different neighborhoods: an annulus with 1-cell inner radius and 10-cell outer radius, and an annulus with 3-cell inner radius and 30-cell outer radius. For impervious surfaces, we also calculated percent coverage for a very small, 3-cell, unweighted rectangular neighborhood. Binary water and wetlands status (1 = presence, 0 = absence), as well as percent imperviousness values, were derived from NLCD land cover and impervious cover products.

Another neighborhood variable we included was a measure of “terrain roughness”. This variable was derived from the digital elevation model, and was calculated as the standard deviation of elevation within a 10-cell radius of the focal cell (similar, but not identical to the “terrain ruggedness index” of Riley et al. [1999]).

Travel Time or Distance to Development Attractors

Variables in this category reflect either straight-line distance or estimated travel time to development attractors including transportation, hydrographic, or land protection features; urban growth “hotspots”; and urban cores of various sizes. As a first step, we produced datasets representing locations of the various attractors. The processes used to delineate growth hotspots and urban cores are briefly described below. For other attractor features, it was a simple matter of reclassifying an existing raster dataset or selecting the appropriate vector features, depending on the source data (see Table 1). To produce distance variables, we applied a Euclidean distance function to measure distance to the nearest attractor. To produce travel-time variables, we had to first develop cost surfaces, then apply a cost-distance function to calculate travel time to the nearest attractor.

Cost surfaces

Roads data were taken primarily from NLCD Impervious Descriptor raster datasets, which identify primary, secondary, and tertiary road classes. To develop travel time cost surfaces, we supplemented the Impervious Descriptor with TIGER/Line roads data (U.S. Census Bureau 2021), which allowed us to distinguish between regular roads, limited access highways, and ramps. We used the U.S. Census Urban Areas dataset (U.S. Census Bureau 2021) to identify which roads were located in urban areas. We assigned travel speeds (mph) based on road class and location within or outside of an urban area (Table 3). Off-road travel speed was assumed to be 3 mph (i.e., walking speed). Cost surfaces were produced by converting travel speed to cost in minutes, i.e., the amount of time it would take to travel through a 30-m cell at the assigned travel

speed. We developed two cost surfaces: one for limited access roads, and another for all other (“local”) roads. To compute travel times, the two cost surfaces were used in conjunction with a custom script¹ which only allows connections between local roads and limited access roads at defined ramp locations.

Impervious growth hotspots

To produce a raster representing impervious growth hotspots, percent impervious cover data from two time periods were needed (Hazler et al. 2016). The source datasets were NLCD Percent Imperviousness products, with raster cell values ranging from 0 (no impervious cover) to 100 (completely impervious). We refer to the raster representing imperviousness at the nominal time period (e.g., 2006) as I_n and the raster representing imperviousness at the previous time period (e.g., 2001) as I_p . The general procedure for delineating hotspots was the same as for the previous edition of the Development Vulnerability Model (Hazler et al. 2016). First, I_p was subtracted from I_n to produce a difference raster. The difference raster was smoothed with a low-pass filter. Any raster cells in which the smoothed difference value was $\geq 20\%$ were identified as potential hotspot cells. Potential hotspot cells were grouped into contiguous regions, and regions less than a critical size threshold (S_{crit}) were eliminated from the final output. We experimented with several values of S_{crit} , but settled on a 5-ha minimum size for hotspots used as attractors in the final model.

Urban cores

We developed urban cores using a custom script implementing a methodology adapted loosely from U.S. Census methods for defining “Urban Areas” (U.S. Census Bureau 2011). The process began with the selection of all census blocks with a population density of at least 1000 persons per square mile (ppsm). Adjacent blocks with least 500 ppsm and/or meeting certain imperviousness and shape criteria were added iteratively and clustered to form urban cores. Urban cores were then categorized based on total population within the core: seed (1,000-2,500), small town (2,501-25,000), small city (25,001-250,000), large city (250,001-2,500,000), or major metropolitan area ($>2,500,000$).

¹ <https://github.com/VANatHeritage/ServiceAreas>

Sample data

To develop sample datasets for model training and testing, we first created raster datasets representing development status in two years: 2006 and 2016. Any raster cell with $\geq 1\%$ impervious cover, based on the NLCD Percent Imperviousness data, was coded 1 for development; all other cells were coded 0. We then created a raster representing the change in development status between the two years, with outcome class codes as follows:

- 0: not developed in either time period
- 1: transitioned from undeveloped to developed
- 2: developed in both time periods
- 3: transitioned from developed to undeveloped.

We created a sampling mask to cover lands not yet developed in 2006, but considered potentially developable. The sampling mask excluded cells meeting any of the following criteria:

- Already developed in 2006 (i.e., codes 2 or 3 in the development status change raster)
- Fully protected conservation lands (i.e., parcels assigned the highest rank for both biological management intent [BMI] and legal protection status [LPS] in the Virginia Conservation Lands Database)
- Very steep slopes ($\geq 70\%$ grade)
- Lands > 2 km from the nearest road

We applied the sampling mask to the development status change raster, and used the resulting binary raster as the sampling frame, where class 0 represents cells that remained undeveloped, and class 1 represents cells that transitioned from an undeveloped to a developed state. We stratified sampling spatially by dividing the state into a mesh of 3-mile diameter hexagons. Hexagons were randomly assigned to training or testing subsets (50% each), and points randomly generated within them as described below. Because of the rarity of cells that became developed, relative to cells that remained undeveloped, these two cases were sampled separately and with different protocols to ensure sufficient samples of the former.

To sample areas that transitioned from undeveloped to developed, the newly developed areas in the sampling frame raster were converted to polygons, and any polygons smaller than 9 raster cells (equivalent to 8100 m^2 or 2 acres) were discarded. Within the remaining polygons, up to 3 points were randomly generated, contingent on a 0.5-mi separation distance.

To sample areas that remained undeveloped, up to 10 points were randomly generated in each hexagon, contingent on a 0.5-mi separation distance. Generated points that fell on developed cells were discarded at this stage. The points from the two sampling procedures were combined into a single dataset attributed with development change status (outcome class 0 or 1), the hexagon ID, and values from each of the predictor rasters. The points were then split into fully independent training and testing datasets depending on the hexagons in which they fell. The training data were used to build and evaluate alternate models. The testing data, not used in training, were used to evaluate the final selected model. Sample sizes were as follows:

- Training, class 0: 19,925 points
- Training, class 1: 1,826 points
- Testing, class 0: 23,153 points
- Testing, class 1: 1,769 points

Modeling approach

Random Forests

We used the Random Forests (RF) machine-learning algorithm (Breiman 2001) to create a classification model predicting the probability of undeveloped lands becoming developed over a 10-year time period. RF models have been shown to perform particularly well with large datasets containing many predictor variables, even when the variables are highly correlated (Strobl et al. 2008, Hapfelmeier et al. 2014). The standard RF classification approach has become widely used in many fields, and in general the performance of RF has been found to compare favorably with that of logistic regression (Couronné et al. 2018).

An RF classification model is built by creating a large number (generally hundreds or more) of classification trees, where each tree is built from a random subset of the input sample points. In addition, a random subset of predictor variables are selected at each split in the tree. The final RF model is an ensemble of all trees. The model can then be applied to novel data (e.g. for new locations or a new time period), to predict class probabilities. Predictions derived from a random forest classifier are class probabilities, and are calculated as the proportion of votes (i.e., the proportion of trees in the ensemble) which assign a given data point to a particular class.

The modeling process was run using the statistical software environment R (R Core

Team 2021), and random forests models executed using the R package *randomForest* (Breiman et al. 2014). Unless otherwise noted, we used the default settings to build a classification model. To build each tree in the ensemble forest, we used a balanced sampling scheme, selecting the same number of samples ($n = 1826$), with replacement, from each outcome class (i.e., 0 = “remained undeveloped”, 1 = “transitioned to developed”).

Exploratory phase

We began with an exploratory, iterative phase, using the training dataset to build multiple alternate models with different sets of predictor variables (Table 1). We initially ran an RF model with 500 trees using a large number of variables, and calculated each variable’s importance using mean decrease in accuracy (MDA; Breiman et al., 2014). We used an unscaled (non-normalized) MDA, as suggested by Strobl et al. (2008). Variables with negative MDA scores were removed from further consideration.

For the remaining variables, we calculated pairwise squared Spearman correlation coefficients (ρ^2) in a hierarchical clustering analysis to identify groups of highly correlated variables, where $\rho^2 \geq 0.8$. In each correlation group, we selected only the variable with the highest importance score, and discarded the others.

The set of retained variables was used to build a final model with 1000 trees. We calculated MDA importance and generated a partial dependence plot for each variable in the final model. Guided by the hierarchical cluster analysis and MDA importance results of previous models, we ran several post hoc, reduced models, with various combinations of the less important variables removed. We also added some new predictor variables, not previously considered, in attempts to improve model performance. When new variables were added, we repeated the procedure of calculating variable importance, clustering variables into correlation groups, and dropping some variables.

To evaluate and compare various alternative models, we used a 10-fold cross-validation procedure (Fielding and Bell 1997), where samples in the training dataset were divided into 10 roughly equal-size validation groups. As previously described, sampling was stratified by hexagons. Validation groups were assigned at the hexagon level, with hexagons first ordered by the number of samples contained so that each group had a similar number of samples, roughly 10% of the total. Validation groups were assigned once, and did not vary between evaluations of

different models. During cross-validation, each one of the 10 groups was withheld and used once to test the predictions derived from a model built using the data in the other groups.

For each validation group of each alternate model under consideration, we derived receiver-operator characteristic (ROC) and precision-recall (PRC) curves (Saito and Rehmsmeier 2015), along with their associated area under the curve metrics (AUC-ROC, and AUC-PRC, respectively). For each model, we then combined the validation results from all 10 groups to derive composite curves and validation metrics. Both ROC and PRC curves, and their associated AUC values, are threshold-independent approaches to evaluating a model's classification power, but Saito & Rehmsmeier (2015) indicated that PRC curves provide a better indication of classification performance in strongly imbalanced datasets in which positive cases (e.g., the transition of land from an undeveloped to a developed state) are far outnumbered by negative cases. We also found that AUC-PRC varied more than AUC-ROC between alternate models, thus providing a stronger basis for model selection, which is consistent with the findings of Saito & Rehmsmeier (2015) given our strongly imbalanced dataset.

After comparing model cross-validation curves, performance metrics and mapped predictions, we selected what we considered to be the “best” model, balancing performance and parsimony. Table 1 is a comprehensive list of all predictor variables used in at least one alternate model, and indicates which variables were included in the final, selected model.

Model testing and prediction

We used the independent testing dataset to evaluate the final selected model that had been developed using the training data. The testing dataset, which was not used or examined at all during the exploratory phase, was run through the model to produce a prediction for each point (developed by 2016 or not), with the results used to derive ROC and PRC curves and associated AUC metrics, as we had done for cross-validation of the training data. Since the independent test results confirmed the predictive power suggested by the cross-validation results, we proceeded with the selected final model. Using that model, we developed a map of "raw vulnerability", representing the relative likelihood of development by 2029, based on predictors representing conditions in 2019 (Figure 2). Predictions range from 0 (least vulnerable to development) to 1.

To develop the final raster representing development vulnerability, we applied some adjustments to the raw vulnerability raster (Figure 3). First, we multiplied raw predictions by a

“BMI multiplier” which was derived from the Biodiversity Management Intent attribute in the Virginia Conservation Lands Database (VCLD), as shown in Table 4, under the assumption that BMI ranks are correlated with protection against development. Unprotected lands, and those with unknown status, were assigned a BMI Multiplier value of 100, meaning the raw vulnerability value was not reduced at all. The resulting raster had values ranging from 0 to 100. Final steps included setting cells with a BMI code of 1 (strongest protection) to a value of -1 (assumed undevelopable), setting already developed cells to a value of 101, setting open water cells to null, and clipping the raster to the border of Virginia. Visual interpretation of any thematic map is strongly dependent on how the numeric values are stretched or classified. Based on inspection of the histogram of final values, in combination with some subjective criteria, we classified vulnerability values as shown in Table 5.

Results

The final selected model included 20 of the 39 predictor variables considered for inclusion. The most influential variables, by far, were the two distance-weighted measures of the amount of impervious cover in the surroundings (Figure 4). As expected, the likelihood of development increased with increasing imperviousness in the surroundings, and decreased with increasing distance to a local road, travel time to a large city or major metro core, distance to an impervious growth hotspot, and terrain roughness (Figure 5).

Cross-validation results indicate that the model has reasonably good predictive power, and this was confirmed by validation metrics using the independent test dataset (Figure 6, Figure 7). The similarity between plots and AUC scores generated from cross-validation of training data versus independent test data is an indicator that the model did not overfit the training data. The AUC (ROC) scores were 0.941 and 0.944 for cross-validation and independent tests, respectively, well above the baseline of 0.5 for a random classifier. The AUC (PRC) scores were 0.599 (baseline: 0.084) and 0.564 (baseline: 0.071) for cross-validation and independent tests, respectively.

The final model output is a raster dataset in which undevelopable lands are coded -1, already developed lands are coded 101, and vulnerability values for potentially developable lands range from 0 to 100. In addition to a statewide static map (Map 1), we produced static

maps for each planning district (Maps 2-22). Data layers for interactive visualization are available on Natural Heritage Data Explorer¹ and in an ArcGIS Online map application², and raster data (30-m resolution) are available for download from the model web page³. These digital products will be available until the model is replaced with a new edition.

Discussion

Model interpretation

In the final model output, areas considered to be invulnerable to development (i.e., fully protected) are coded -1, and areas already developed are coded 101. The intervening values from 0 to 100 are a measure of development vulnerability, i.e., the relative likelihood of transitioning from an undeveloped to a developed state by the year 2029. It is important to note that vulnerability values are not probabilities. For example, a value of 30 does not mean that there is a 30% chance that the land will be developed. Actual probability of development will depend on a number of influences not explicitly included in this model, such as population trends, economic factors, local zoning, regional planning, and climate change. Instead of probabilities, model values should be interpreted as a relative measure of "development potential" (*sensu* Meentemeyer et al., 2013) or "attractiveness" (*sensu* Westervelt et al., 2011).

The creation of a true forecast model, in which output values can be regarded as the probability of development, goes beyond the scope of our modeling effort. Models used to forecast land use change typically identify several possible growth scenarios, and then conduct time-stepped simulations under the assumptions of each scenario to model urban growth over time, out to a specified future date (Irwin et al. 2003, Claggett et al. 2004, in prep., Sohl and Sayler 2008, Jantz et al. 2010, Westervelt et al. 2011, Meentemeyer et al. 2013, Terando et al. 2014, McGarigal et al. 2018). Ours is not a simulation model, but its output could serve as an input to a simulation exercise. Essentially, it produces a suitability surface upon which a

¹ <https://vanhde.org/content/map>

² <https://arcg.is/0qWej4>

³ <https://www.dcr.virginia.gov/natural-heritage/vaconvisvulnerable>

simulated scenario could be played out.

Limitations and opportunities

This model, like any other model, is limited by the data inputs as well as by the assumptions made and processes used in combining these inputs. Many of the inputs are raster datasets with a pixel size of 30-m, and this dictated the model's output resolution, which may or may not be sufficient for detailed planning at local scales. In addition, the model was trained on developed patches at least 2 acres in size, so it may or may not capture more dispersed development patterns such as single-family homes built on larger lots.

All input datasets unavoidably have some spatial and/or attribute errors, which propagate to the final output. Model performance depends on the predictor variables included, and it is possible that we missed some predictors that could have vastly improved the model's predictive power. That said, we developed a fairly extensive list of reasonable predictors to consider for inclusion, and the validation and testing statistics suggest that our model does have reasonably strong predictive power.

Our model relies heavily on how we define the transition from an undeveloped to a developed state. This definition is based on datasets representing percent impervious cover, and we defined pixels with $\geq 1\%$ imperviousness as developed. Had we chosen a different threshold (e.g., defining developed as $\geq 5\%$ impervious), our predictions and final outputs would have been at least somewhat different. Unlike some models (e.g., Claggett et al., in preparation), this model does not distinguish between different types of development, such as residential, commercial, and industrial, which may have very different growth patterns.

The maps presented in this report, and the underlying raster model used to produce them, should be considered as a snapshot in time, reflecting relative risk of development based on ground conditions in the year 2019 and development patterns inferred by the model based on training data from 2006-2016. Ground conditions as well as development pressures are constantly changing over time. National Land Cover Database data are now on a 2- to 3-year update cycle, and it makes sense to update the development vulnerability model whenever new land cover and imperviousness products become available. Because of the relative simplicity of the model (i.e., not requiring scenarios and simulations), reasonable data requirements, and the

fact that the model code has been archived on GitHub¹, we expect model updates in the future to be efficiently implemented on a regular basis.

Model comparisons

For comparison with the Virginia Development Vulnerability Model, we looked at mapped predictions from three other recent modeling efforts covering Virginia. For states with waters draining to the Chesapeake Bay, the “Current Zoning Vulnerability” 2035 scenario from the Chesapeake Bay Land Change Model (CBP 2020, Claggett et al. in prep.) maps the likelihood of development by 2035, under the assumption that current zoning policies persist. For Southeastern states, the Southeast Regional Assessment Project (Terando et al. 2014, SERAP 2016) applied the SLEUTH urban growth model to map urbanization probabilities at 10-year intervals, and we acquired the projections for 2030. Covering Northeastern states, we acquired a map of the probability of development by 2030 from the Nature’s Network project (NALCC 2017, McGarigal et al. 2018).

The methodology and mapped classes differed greatly between models, in particular because our model did not include a simulation component like the other models. Nonetheless, we recognized some general patterns distinguishing our vulnerability map from the other maps. In general, our model output shows relatively higher vulnerability at margins of large and medium metro areas, along large highways, and for most of the Eastern Shore. Among the four models, outputs from ours and the Northeast model were most similar. A visual assessment was corroborated with a simple Spearman rank correlation analysis, revealing correlations of 0.66, 0.34, and 0.21 between our model and the Northeastern, Southeastern, and Chesapeake Bay model outputs, respectively. There were stark differences between outputs from the Chesapeake Bay model and other models in certain localities due to the influence of current zoning maps. In some localities the Chesapeake Bay model suggested that certain areas are completely invulnerable to development, which is unlikely to be true since zoning can change over time and special use permits can be obtained. Models also differed in how protected lands were portrayed in their maps. Whereas only protected lands with the strongest biodiversity management

¹ https://github.com/VANatHeritage/ConsVision_DevVulnModel

protection are indicated as completely invulnerable to development in our model, it was evident that the Northeast and Chesapeake Bay models assigned their lowest vulnerability class to a broader set of protected lands.

Instead of distinguishing between different types of development as in the other models, our approach was more general, using the transition from greenspace to impervious surface to identify new development areas and train the model. Because of this, the model is influenced not only by urban or suburban development, but also the types of development which may be more common in rural areas (e.g. agricultural or energy infrastructure). For example, in recent years Virginia has seen rapid expansion of large, utility-scale solar installations in Virginia (SEIA 2022). While this type of development was not as common in the time period used to train the model (2006 – 2016), it will likely become more of a factor in future updates.

Model applications

The Virginia Development Vulnerability Model is intended to reflect the relative risk of losing greenspace to development across the state. We expect the model to be helpful to state and local governments, planning districts, environmental consultants, land trusts, and others involved in land use planning and strategic conservation. In most cases, this model should be used in conjunction with other relevant information and datasets, including other ConservationVision models, to help prioritize lands for conservation and to inform comprehensive planning efforts. The model can also serve as an input for simulating future land cover change and its consequences under different planning scenarios.

Tables

Table 1: Comprehensive list of predictor variables considered for use in the model.....	17
Table 2: Cost classes assigned to National Land Cover Database land cover classes.....	21
Table 3: Travel speed assignments by road class and urban status	21
Table 4: Biological Management Intent (BMI) and the BMI Multiplier.....	22
Table 5: Classification of raster values for the final development vulnerability map	22

Table 1: Comprehensive list of predictor variables considered for use in the model.

Predictor variables were developed as raster datasets with 30-m resolution. Variables in rows shaded grey were NOT included in the final model.

Variable Name	Description	Spatial focus	Units	Temporal Status	Data Source(s) ¹
elevcm	Elevation	Site characteristics	cm	Static	3DEP
slpx100	Slope	Site characteristics	%	Static	3DEP
soilSuit_Development	Soil suitability for development	Site characteristics	N/A	Static	SSURGO; NLCD
lc_cost	Land cover development cost	Site characteristics	N/A	Dynamic	NLCD
conslands_protMult	Land protection multiplier	Site characteristics	N/A	Dynamic	VCLD
imp_kRect3	Percent imperviousness, 3-cell rectangular neighborhood (in charts: “Imperviousness, 3-cell nbhd.”)	Neighborhood characteristics	%	Dynamic	NLCD
imp_wk10_025	Distance-weighted percent imperviousness (radius 1-10, gamma 0.25)	Neighborhood characteristics	%	Dynamic	NLCD
imp_wk10_050	Distance-weighted percent imperviousness (radius 1-10, gamma 0.50) (in charts: “Imperviousness, 10-cell nbhd.”)	Neighborhood characteristics	%	Dynamic	NLCD
imp_wk10_100	Distance-weighted percent imperviousness (radius 1-10, gamma 1.00)	Neighborhood characteristics	%	Dynamic	NLCD

¹ Data source abbreviations: 3DEP = 3D Elevation Program (U.S. Geological Survey (USGS) 2017); NHD = National Hydrography Dataset (U.S. Geological Survey (USGS) 2018); NLCD = National Land Cover Database (Yang et al. 2018, Dewitz and USGS 2021); PAD-US = Protected Areas Database of the U.S. (Prior-Magee et al. 2020); SSURGO = Soil Survey Geographic Database (Soil Survey Staff 2020); TIGER = Topologically Integrated Geographic Encoding and Referencing data (U.S. Census Bureau 2021); VCLD = Virginia Conservation Lands Database (Virginia DCR Staff 2021).

Variable Name	Description	Spatial focus	Units	Temporal Status	Data Source(s) ¹
imp_wk30_025	Distance-weighted percent imperviousness (radius 3-30, gamma 0.25)	Neighborhood characteristics	%	Dynamic	NLCD
imp_wk30_050	Distance-weighted percent imperviousness (radius 3-30, gamma 0.50) (in charts: "Imperviousness, 30-cell nbhd.")	Neighborhood characteristics	%	Dynamic	NLCD
imp_wk30_100	Distance-weighted percent imperviousness (radius 3-30, gamma 1.00)	Neighborhood characteristics	%	Dynamic	NLCD
water_wk10_050	Distance-weighted percent open water (radius 1-10, gamma 0.50) (in charts: "Open Water, 10-cell nbhd.")	Neighborhood characteristics	%	Dynamic	NLCD
water_wk30_050	Distance-weighted percent open water (radius 3-30, gamma 0.50) (in charts: "Open Water, 30-cell nbhd.")	Neighborhood characteristics	%	Dynamic	NLCD
wetland_wk10_050	Distance-weighted percent wetland (radius 1-10, gamma 0.50) (in charts: "Wetland, 10-cell nbhd.")	Neighborhood characteristics	%	Dynamic	NLCD
wetland_wk10_100	Distance-weighted percent wetland (radius 1-10, gamma 0.50)	Neighborhood characteristics	%	Dynamic	NLCD
wetland_wk30_050	Distance-weighted percent wetland (radius 3-30, gamma 0.50) (in charts: "Wetland, 30-cell nbhd.")	Neighborhood characteristics	%	Dynamic	NLCD
elevRough_10	Terrain roughness (standard deviation of elevation, 10-cell radius)	Neighborhood characteristics	N/A	Static	3DEP
edist_openwater	Euclidean distance to open water	Distance or time to attractors	m	Dynamic	NLCD
edist_inlandWater	Euclidean distance to river, lake, or reservoir (> 100 acres)	Distance or time to attractors	m	Static	NHD
edist_bayOcean	Euclidean distance to Chesapeake Bay or Atlantic Ocean	Distance or time to attractors	m	Static	NHD

Variable Name	Description	Spatial focus	Units	Temporal Status	Data Source(s) ¹
edist_conslands	Euclidean distance to conserved land	Distance or time to attractors	m	Dynamic	VCLD; PAD-US
edist_imphot_20_02	Euclidean distance to impervious growth hotspot (2-ha)	Distance or time to attractors	m	Dynamic, multi-temporal	NLCD
edist_imphot_20_05	Euclidean distance to impervious growth hotspot (5-ha) (in charts: “Dist. to Impervious Growth Hotspot.”)	Distance or time to attractors	m	Dynamic, multi-temporal	NLCD
edist_imphot_20_10	Euclidean distance to impervious growth hotspot (10-ha)	Distance or time to attractors	m	Dynamic, multi-temporal	NLCD
edist_road	Euclidean distance to road	Distance or time to attractors	m	Dynamic	NLCD
edist_newRoad	Euclidean distance to newly constructed road (in charts: “Dist. to New Road”)	Distance or time to attractors	m	Dynamic, multi-temporal	NLCD
edist_localRoad	Euclidean distance to local road (in charts: “Dist. to Local Road”)	Distance or time to attractors	m	Dynamic	TIGER; NLCD
edist_hwy	Euclidean distance to limited access highway	Distance or time to attractors	m	Dynamic	TIGER; NLCD
edist_ramp	Euclidean distance to limited access highway access ramp	Distance or time to attractors	m	Dynamic	TIGER; NLCD
ttRamps	Travel time to limited access highway ramps (in charts: “Travel Time to Highway Ramps”)	Distance or time to attractors	min	Dynamic	TIGER; NLCD
ttCore1	Travel time to seed core	Distance or time to attractors	min	Dynamic	TIGER; NLCD
ttCore2	Travel time to small town core	Distance or time to attractors	min	Dynamic	TIGER; NLCD

Variable Name	Description	Spatial focus	Units	Temporal Status	Data Source(s)¹
ttCore3	Travel time to small city core	Distance or time to attractors	min	Dynamic	TIGER; NLCD
ttCore4	Travel time to large city core	Distance or time to attractors	min	Dynamic	TIGER; NLCD
ttCore5	Travel time to major metropolitan core	Distance or time to attractors	min	Dynamic	TIGER; NLCD
ttCore2_3	Travel time to small town or small city core	Distance or time to attractors	min	Dynamic	TIGER; NLCD
ttCore2_5	Travel time to small town or larger core	Distance or time to attractors	min	Dynamic	TIGER; NLCD
ttCore4_5	Travel time to large city or major metro core	Distance or time to attractors	min	Dynamic	TIGER; NLCD

Table 2: Cost classes assigned to National Land Cover Database land cover classes

Cost Class	Description	NLCD Classes Included
0	Already developed	21 – Developed, Open Space 22 – Developed, Low Intensity 23 – Developed, Medium Intensity 24 – Developed, High Intensity
1	Barren	31 – Barren Land (Rock/Sand/Clay)
2	Herbaceous or agricultural	71 – Grassland/Herbaceous 81 – Pasture/Hay 82 – Cultivated Crops
3	Shrubland	52 – Shrub/Scrub
4	Forest or wetland	41 – Deciduous Forest 42 – Evergreen Forest 43 – Mixed Forest 90 – Woody Wetlands 95 – Emergent Herbaceous Wetlands
5	Open Water	11 – Open Water

Table 3: Travel speed assignments by road class and urban status

Road Class	Standard Speed (miles/hour)	Urban Speed (miles/hour)
Primary – limited access highway	70	60
Primary – not limited access highway	55	45
Secondary – limited access highway	60	50
Secondary – not limited access highway	45	35
Tertiary	35	25

Table 4: Biological Management Intent (BMI) and the BMI Multiplier

BMI Code ^a	Code Description	BMI Multiplier ^b
1	Specifically managed for the protection of plant and animal communities	0
2	Managed for the conservation of plant and animal communities with limited impacts permitted	20
3	Managed for general natural resource conservation	40
4	General open space conservation	60
5	No designation or management for conservation of natural conditions	80
U	Unknown status	100

^a Codes used in the Virginia Conservation Lands Database^b Multiplier value used to adjust raw vulnerability values to obtain final model output values**Table 5: Classification of raster values for the final development vulnerability map**

Raster Values	Map Class
-1	Undevelopable
0 - 5	Class I - Least Vulnerable
6 - 10	Class II - Somewhat Vulnerable
11 - 25	Class III - Moderately Vulnerable
26 - 50	Class IV - Highly Vulnerable
51 - 100	Class V - Most Vulnerable
101	Already Developed

Figures

Figure 1: Training the model	24
Figure 2: Predicting vulnerability using the trained model	25
Figure 3: Creating the final map of development vulnerability.....	26
Figure 4: Relative importance of predictor variables in the final selected model	27
Figure 5: Partial dependence plots for the most influential variables in the final selected model	28
Figure 6: Receiver operating characteristic (ROC) plots for cross-validation and independent tests	29
Figure 7: Precision-recall (PRC) plots for cross-validation and independent tests	30

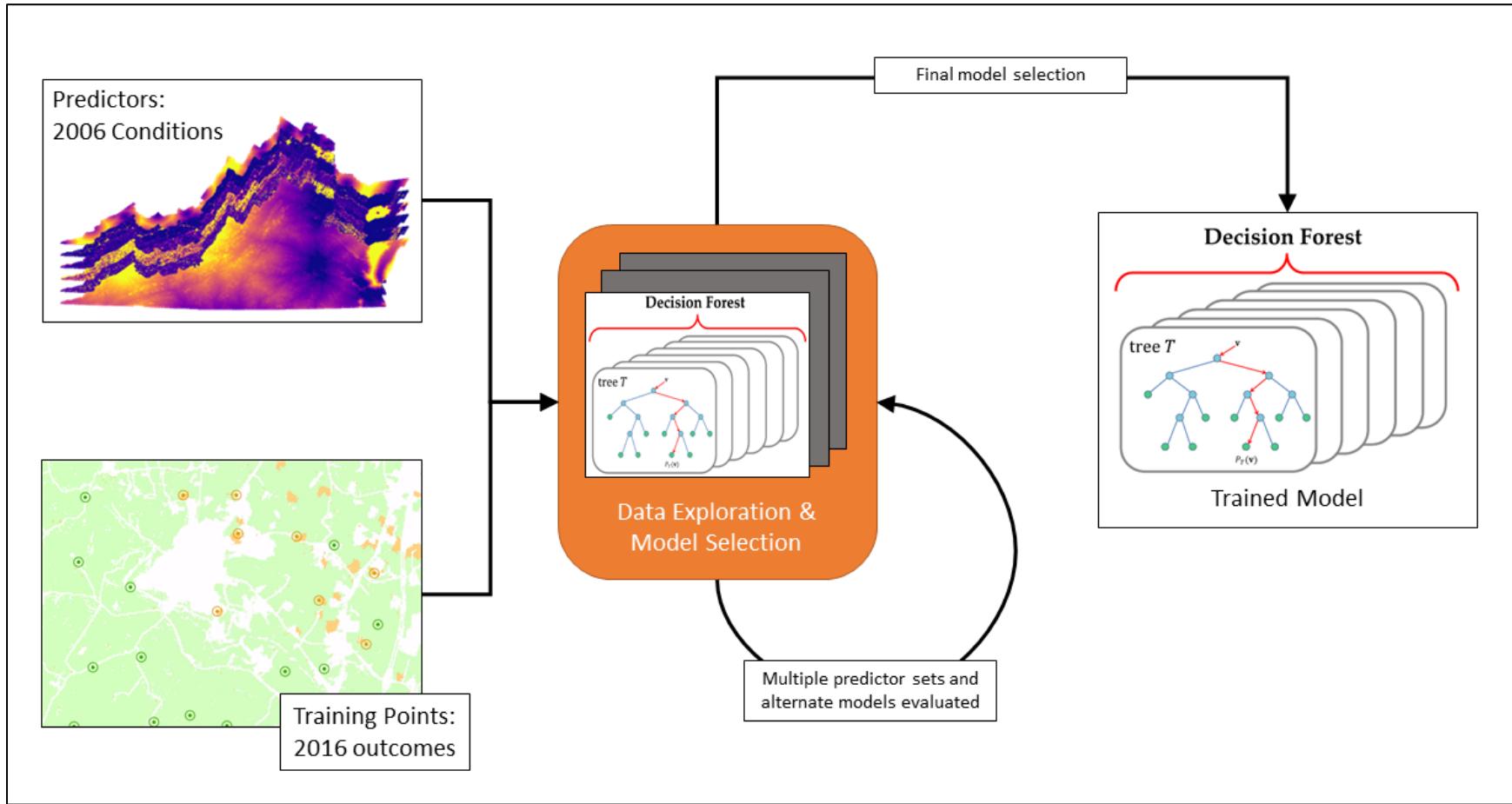


Figure 1: Training the model

A suite of 39 rasters representing conditions in 2006 were developed as potential predictor variables (upper left). Training points (lower left) were a random selection of sites that were undeveloped but developable in 2006, and either stayed undeveloped (green) or became developed (orange) by 2016. (Non-colored areas were not eligible for sampling based on criteria specified in methods.) The combination of 2006 predictor variables with 2016 outcomes at the training point locations were used to build alternate Random Forests models in an exploratory, iterative fashion. A final “best” model was selected, tested with independent sample data, and then used for prediction.

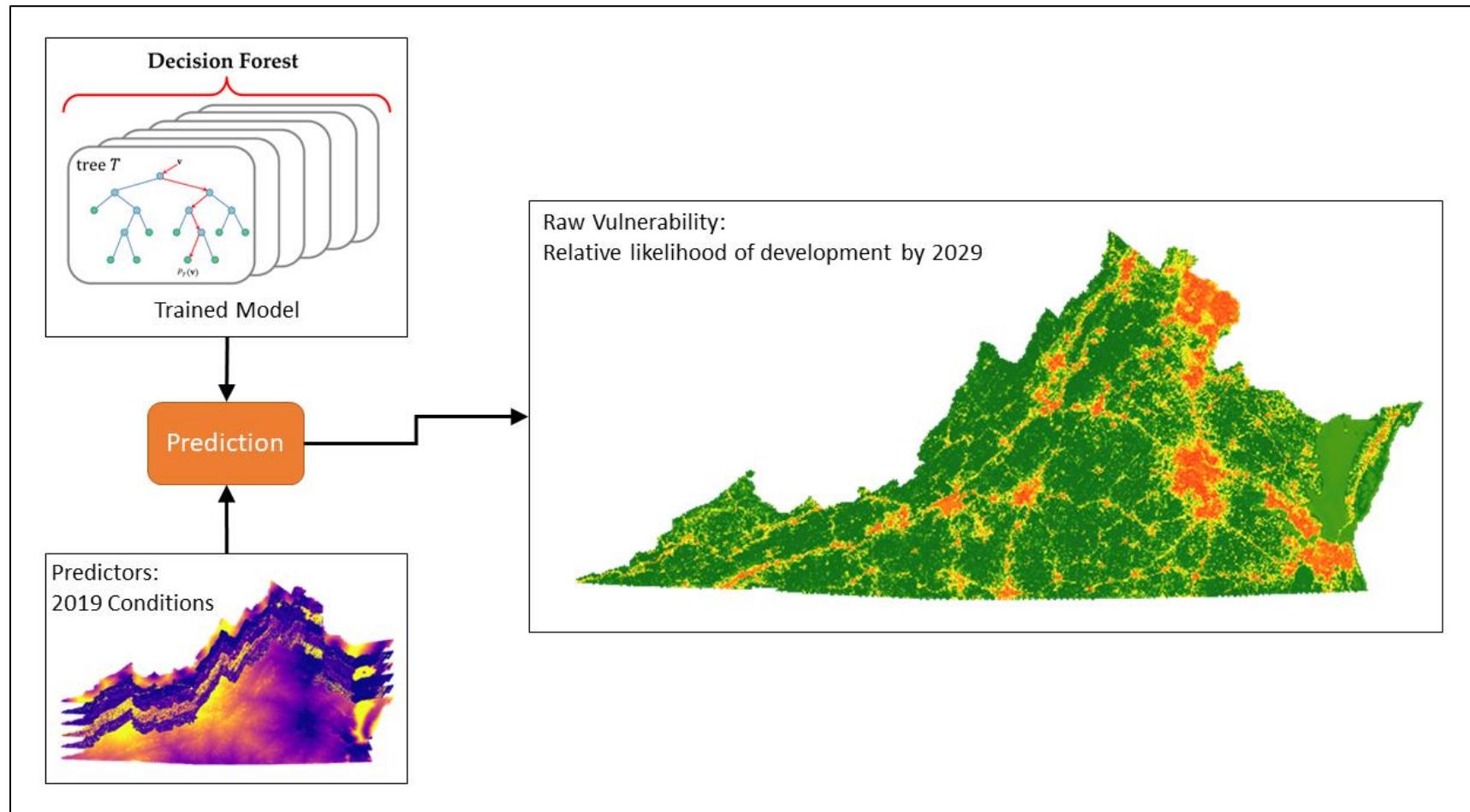


Figure 2: Predicting vulnerability using the trained model

The selected Random Forests model, trained on the dataset with 2006-era predictors and 2016 outcomes, was used to produce a raster dataset representing the relative likelihood of development by 2029, based on conditions in 2019.

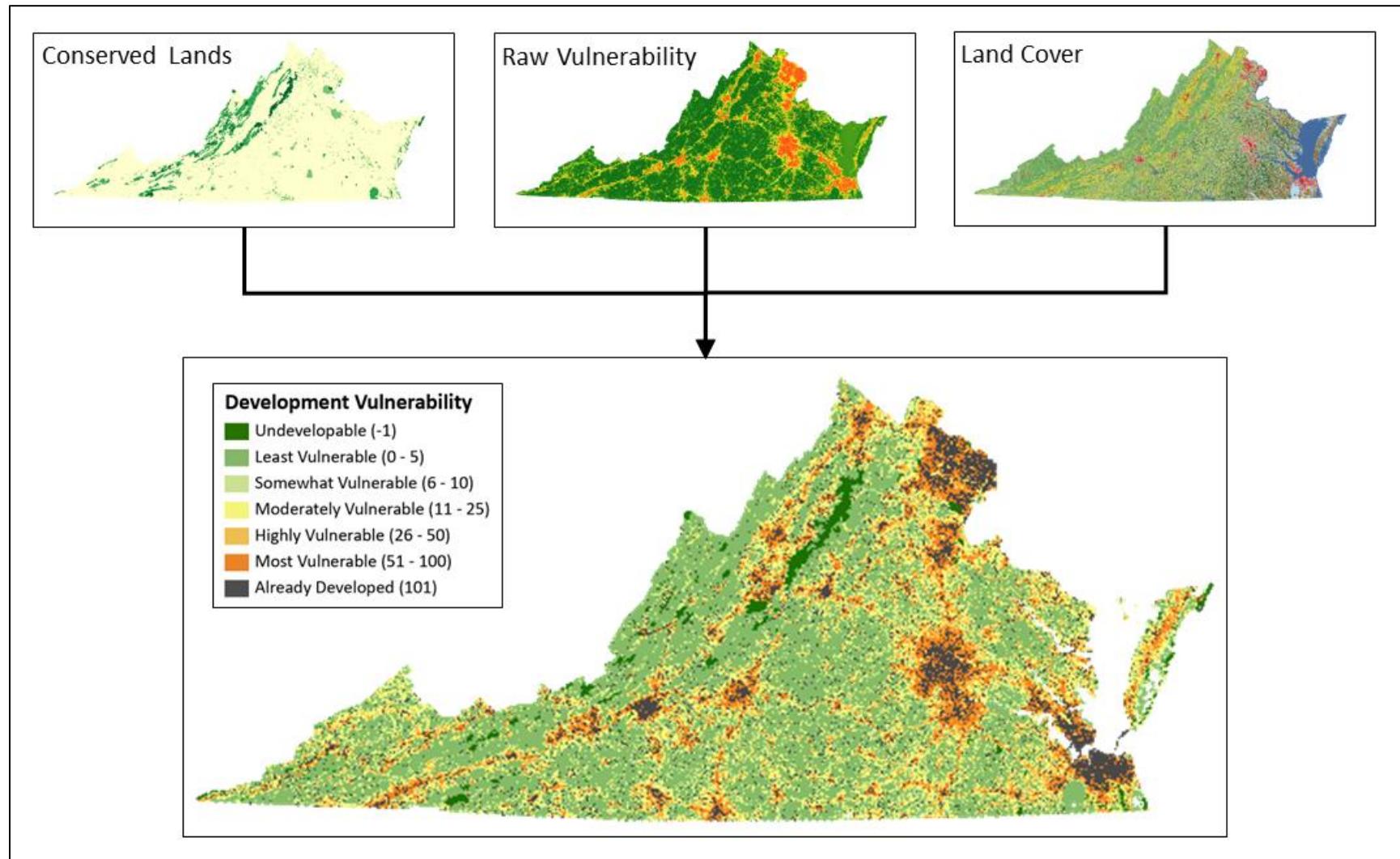


Figure 3: Creating the final map of development vulnerability

To produce the final map of development vulnerability, raw vulnerability values were adjusted based on land cover and conservation status.

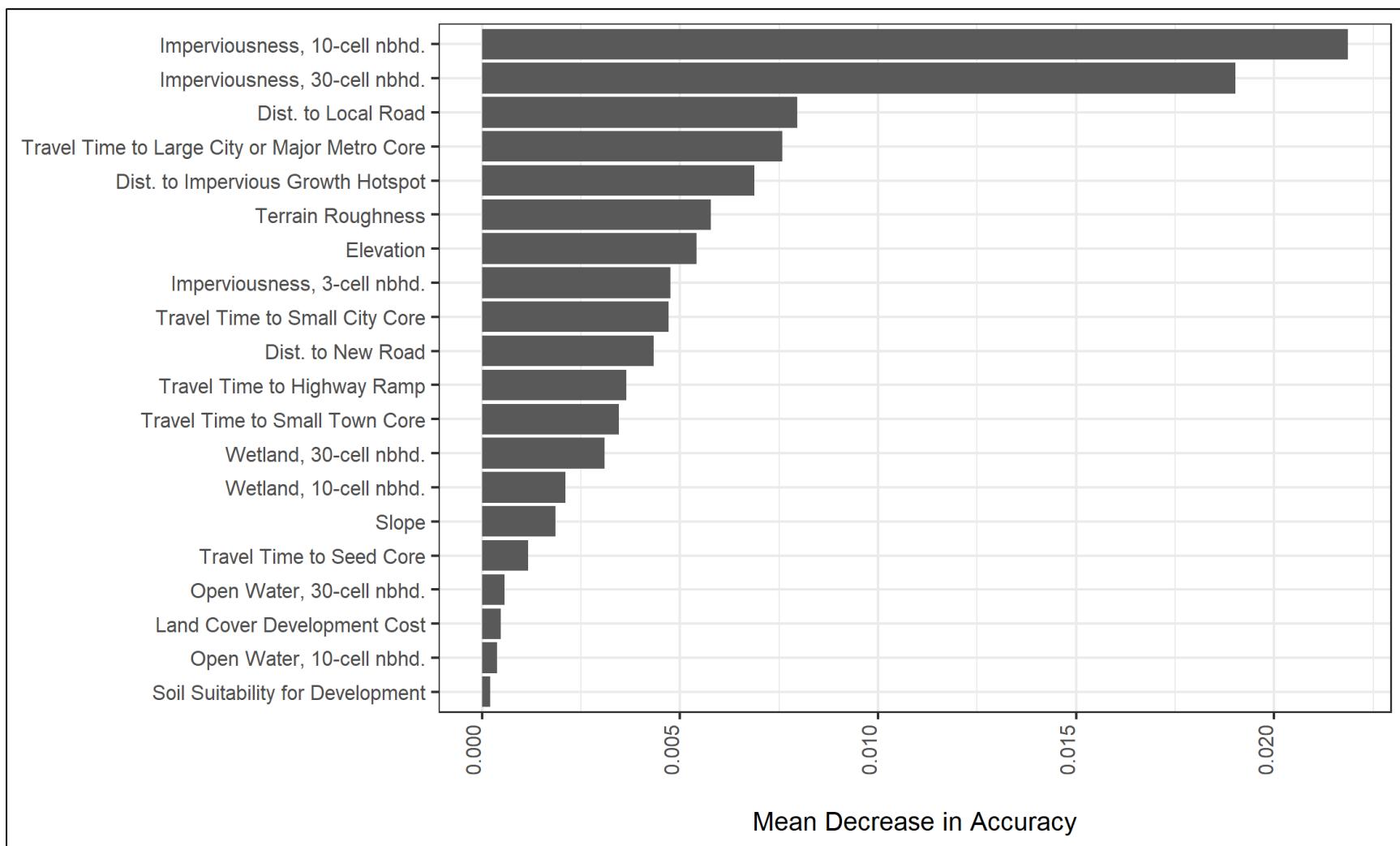


Figure 4: Relative importance of predictor variables in the final selected model

The relative importance of each predictor was measured by the mean decrease in accuracy after random permutation (Breiman et al. 2014). For more information about the predictor variables, see Table 1.

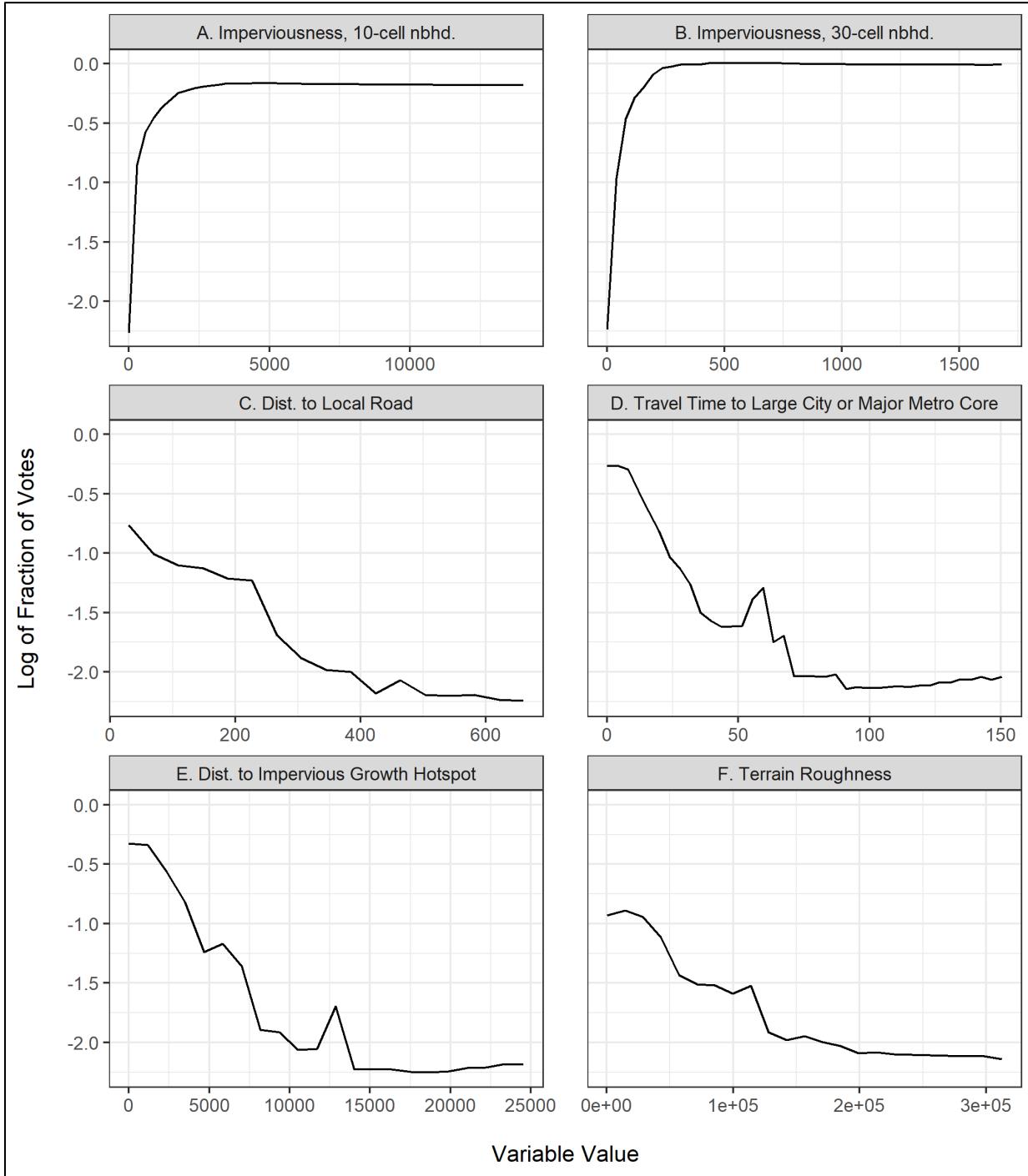


Figure 5: Partial dependence plots for the most influential variables in the final selected model

Each partial dependence plot shows the effect of the variable on the predicted relative likelihood of development with the effects of the other variables removed. The x-axis covers the range of values for the variable for points which became developed; the y-axis represents the model response. Curves generally decreasing from left to right indicate that the likelihood of development decreases with increasing values of the predictor, i.e., a negative relationship. Peaks in the curve indicate where the variable had the strongest influence on predicting relative likelihood of development.

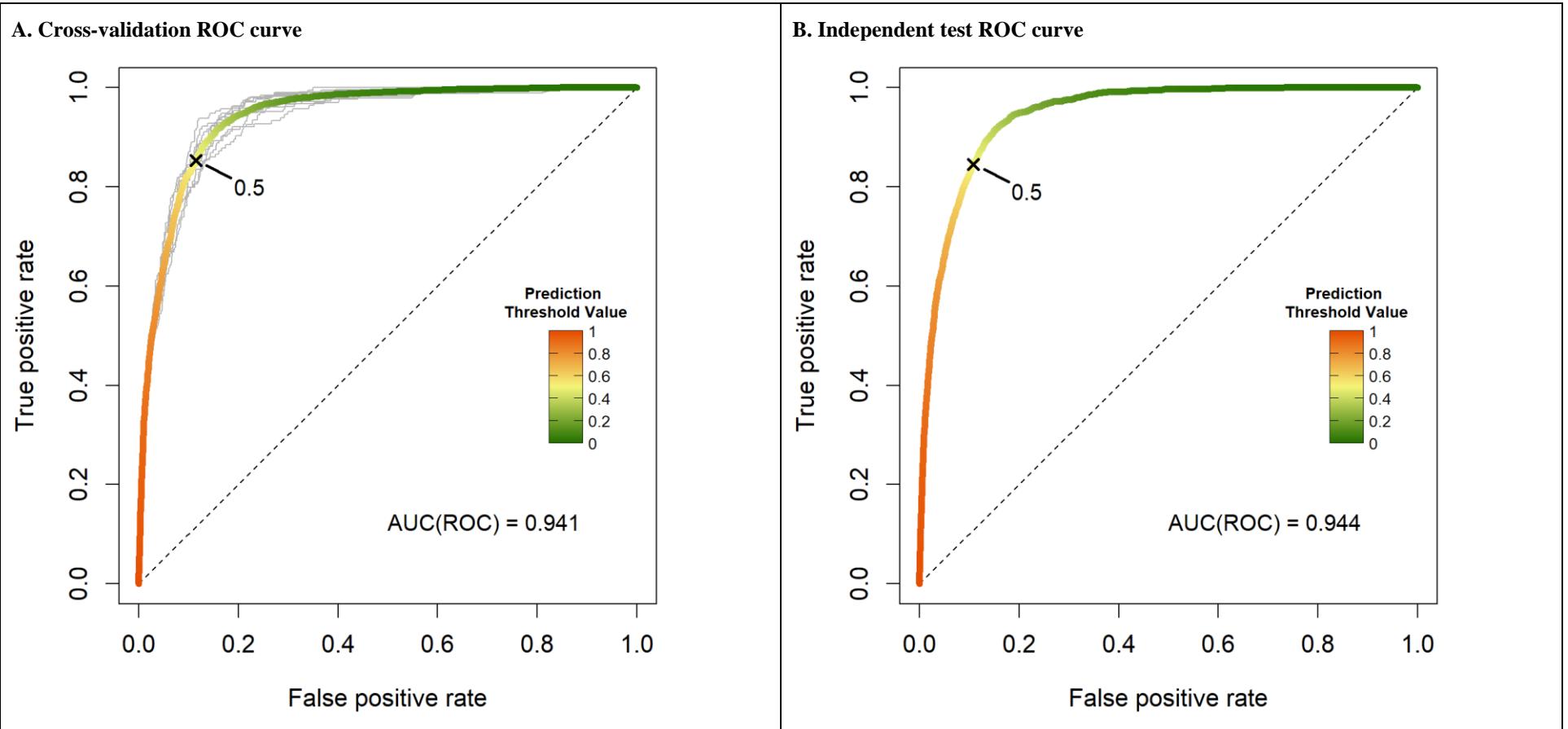


Figure 6: Receiver operating characteristic (ROC) plots for cross-validation and independent tests

The receiver operating characteristic (ROC) curve illustrates the tradeoff between the true and false positive rates as the threshold used to predict a positive outcome (i.e., development) is varied (Saito and Rehmsmeier 2015). The true positive rate (also known as sensitivity or recall) is the proportion of actual positive cases that the model predicts to be positive. The false positive rate is the proportion of actual negative cases predicted to be positive. If the prediction threshold is very low, almost all cases will be predicted positive, so both true and false positive rates will be very high. At the other extreme, both will be very low. The more the ROC curve bends toward the upper left corner, away from the diagonal dotted line, the greater the area under the curve (AUC) and the better the classification power of the model. AUC(ROC) is 1.0 for perfect classifiers, and 0.5 (area under the dashed diagonal line) for classifiers that perform no better than random (Saito and Rehmsmeier 2015).

Panel A depicts the results from the 10-fold cross-validation, using the training data, whereas Panel B shows the results from a test using independent data that were not used to train the model.

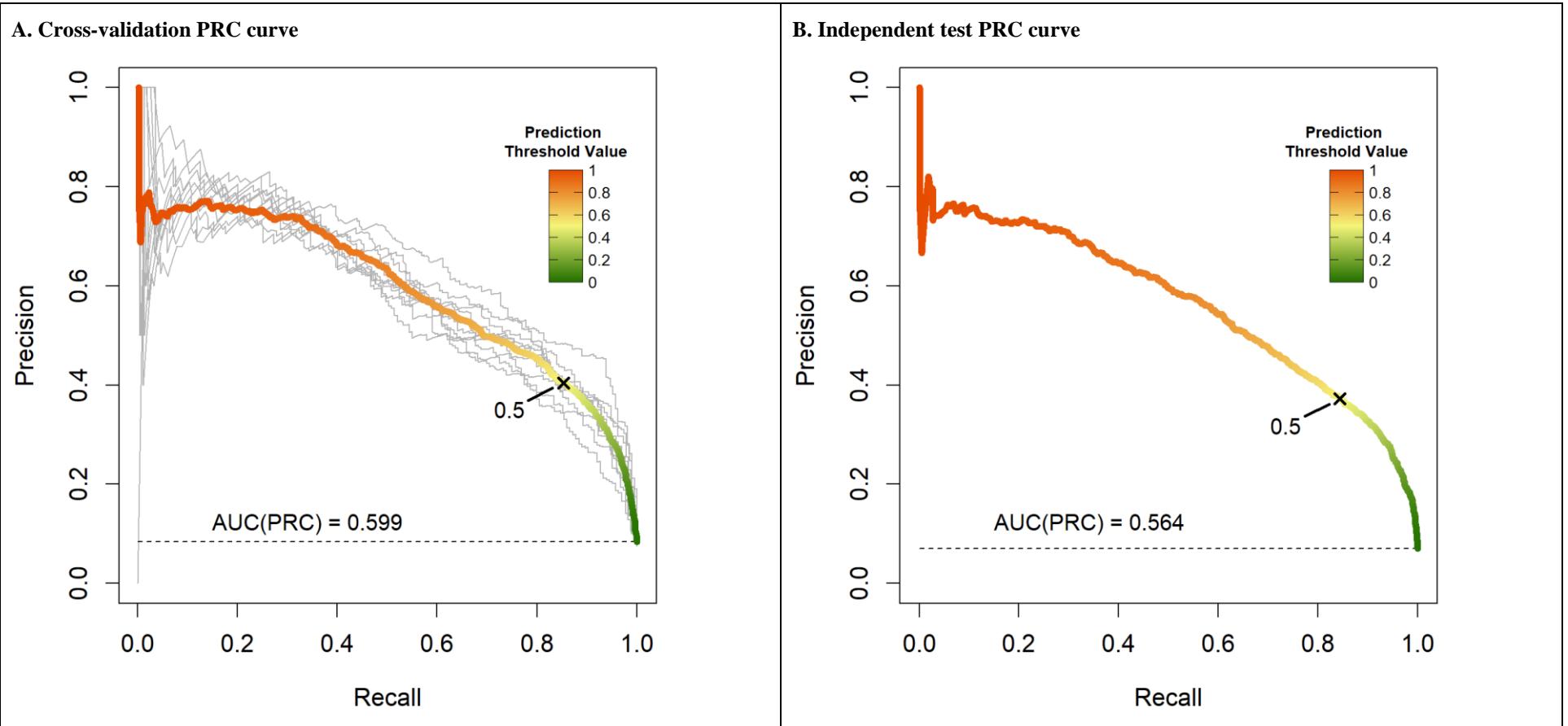


Figure 7: Precision-recall (PRC) plots for cross-validation and independent tests

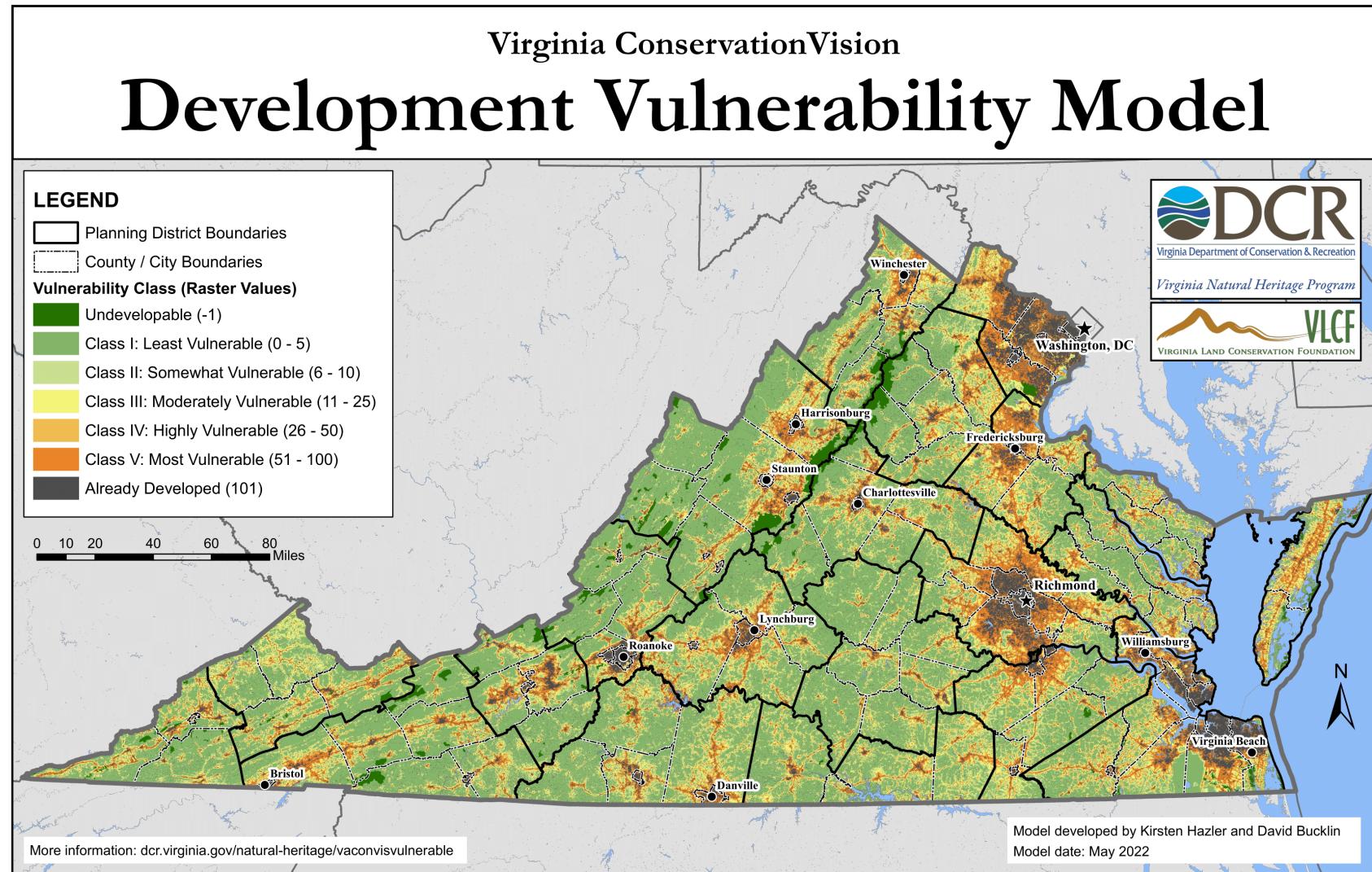
The precision-recall curve illustrates the tradeoff between precision and recall as the threshold used to predict a positive outcome is varied (Saito and Rehmsmeier 2015). Precision, also known as positive predictive value, is the proportion of predicted positive cases that are, in fact, positive. Perfect precision would mean there are no false positives. Recall, also known as sensitivity or the true positive rate, is the proportion of actual positive cases that the model predicts to be positive. Perfect recall would mean there are no false negatives. The baseline (horizontal dashed line) against which the PRC plot should be evaluated is determined by the relative proportions of positives and negatives in the dataset (Saito and Rehmsmeier 2015).

Panel A depicts the results from the 10-fold cross-validation, using the training data, whereas Panel B shows the results from a test using independent data that were not used to train the model.

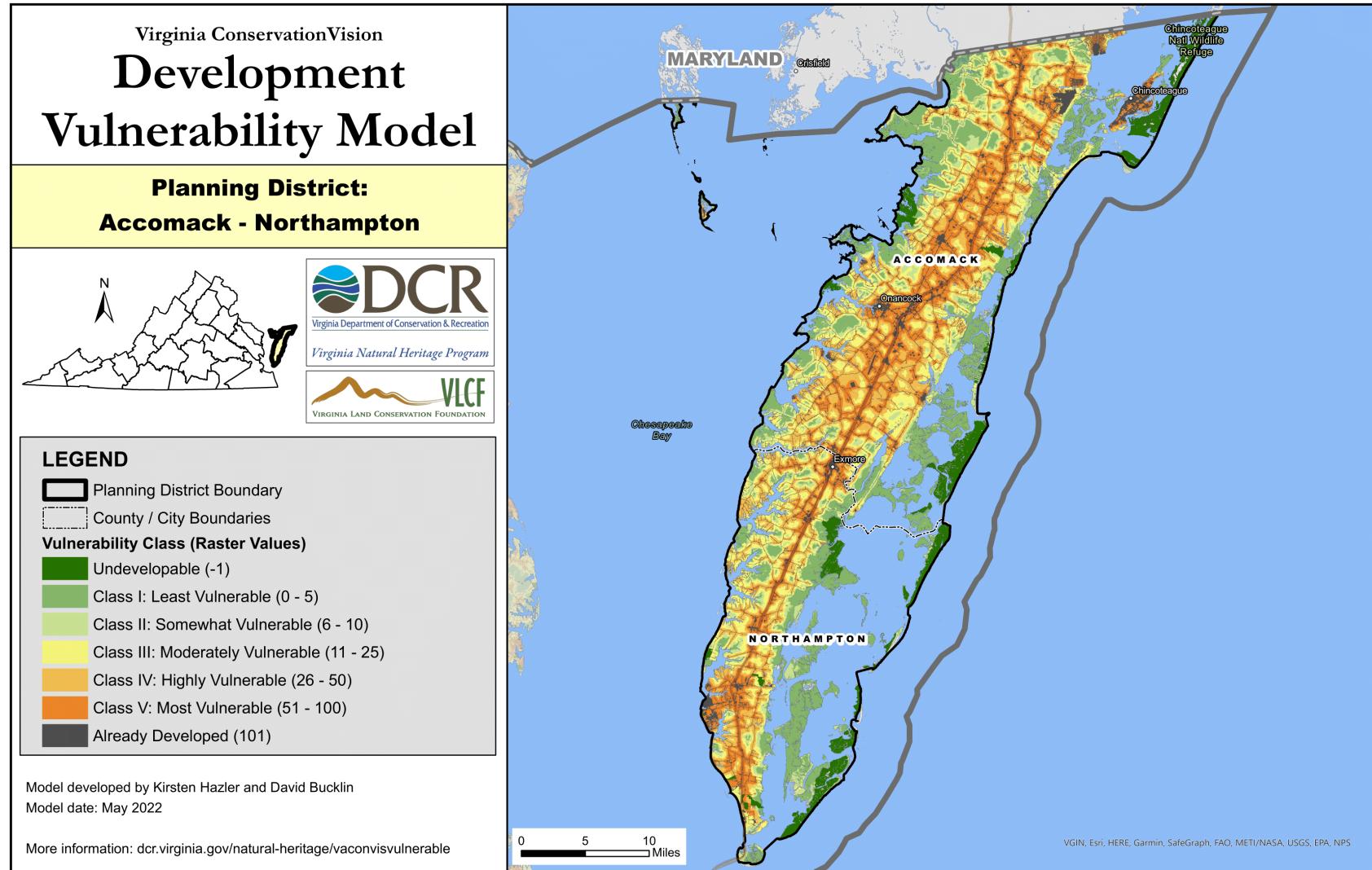
Maps

Map 1: Statewide development vulnerability	32
Map 2: Accomack-Northampton Planning District	33
Map 3: Central Shenandoah Planning District.....	34
Map 4: Commonwealth Regional Council Planning District	35
Map 5: Crater Planning District.....	36
Map 6: Cumberland Plateau Planning District	37
Map 7: George Washington Planning District.....	38
Map 8: Hampton Roads Planning District	39
Map 9: LENOWISCO Planning District	40
Map 10: Middle Peninsula Planning District.....	41
Map 11: Mount Rogers Planning District.....	42
Map 12: New River Valley Planning District.....	43
Map 13: Northern Neck Planning District	44
Map 14: Northern Shenandoah Valley Planning District	45
Map 15: Northern Virginia Planning District	46
Map 16: Rappahannock - Rapidan Planning District	47
Map 17: Region 2000 Planning District	48
Map 18: Richmond Regional Planning District.....	49
Map 19: Roanoke Valley - Alleghany Planning District	50
Map 20: Southside Planning District	51
Map 21: Thomas Jefferson Planning District	52
Map 22: West Piedmont Planning District	53

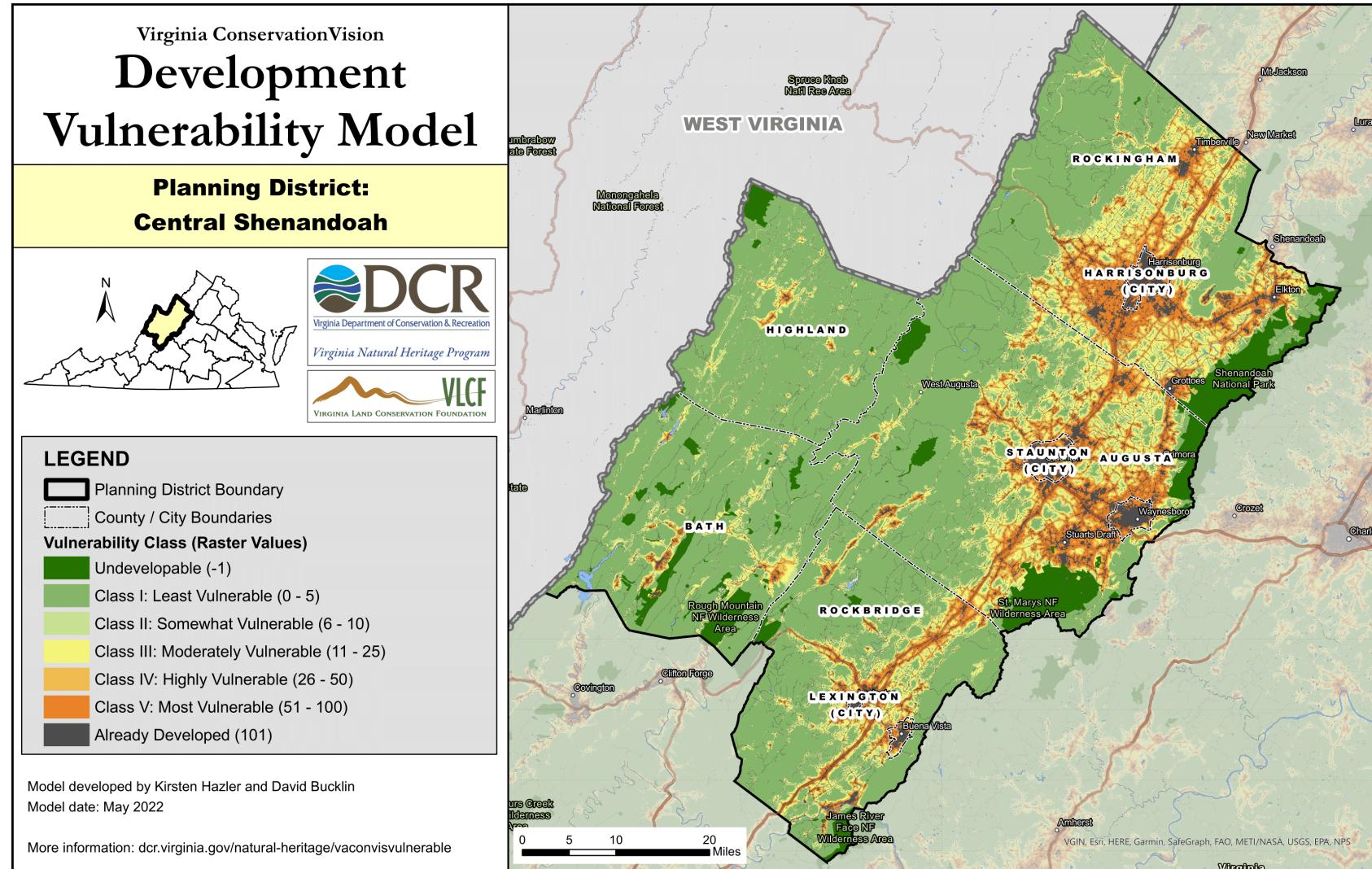
Map 1: Statewide development vulnerability



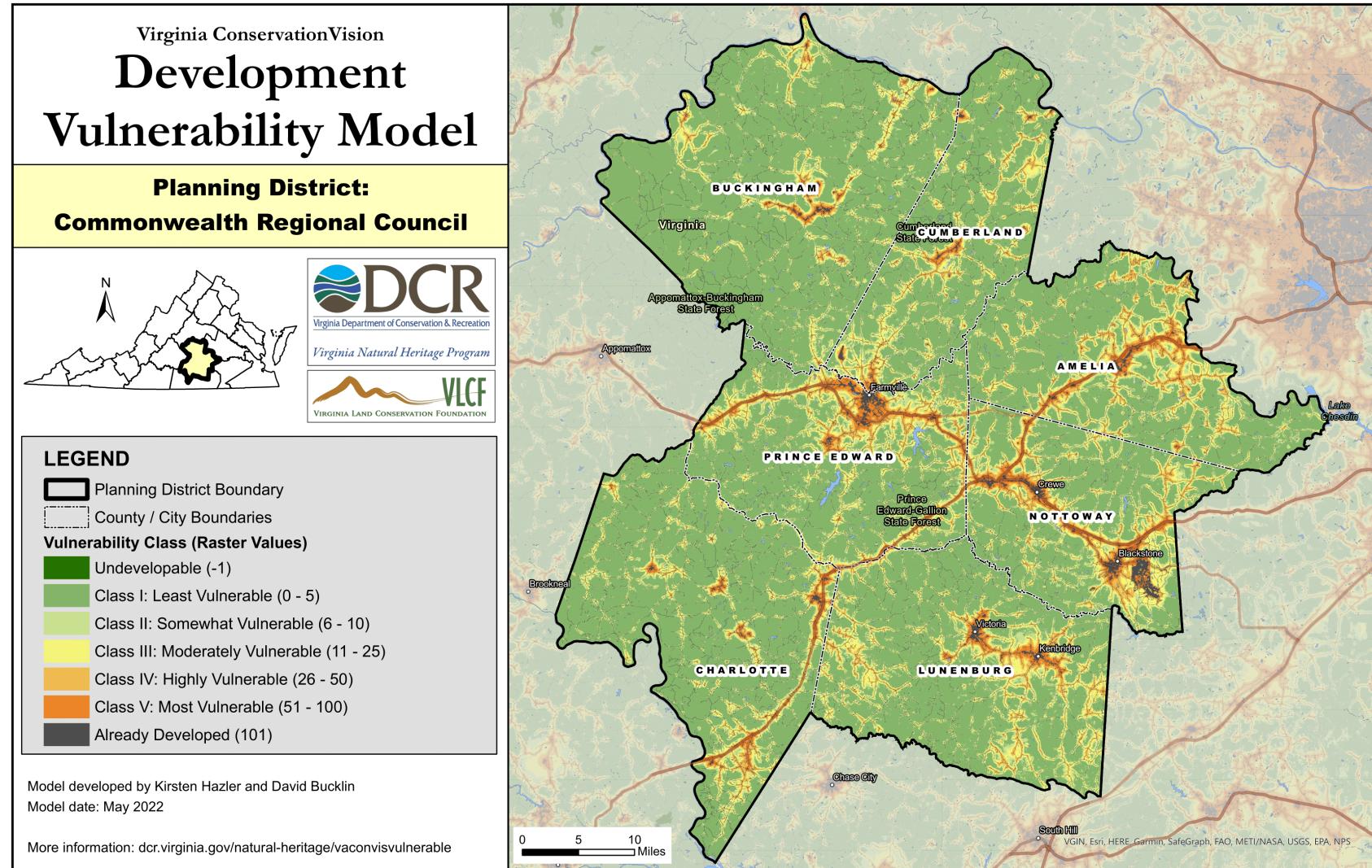
Map 2: Accomack-Northampton Planning District



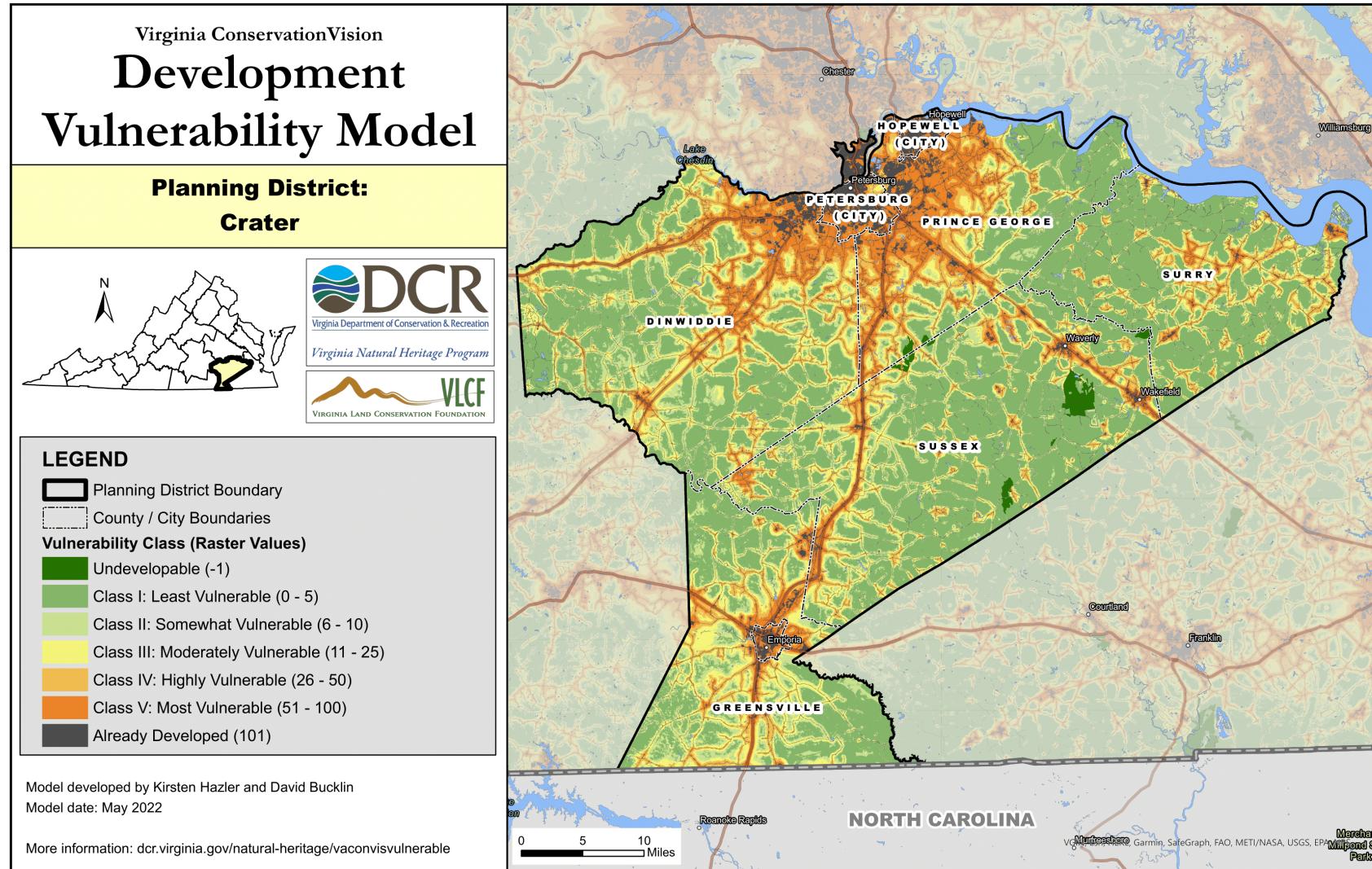
Map 3: Central Shenandoah Planning District



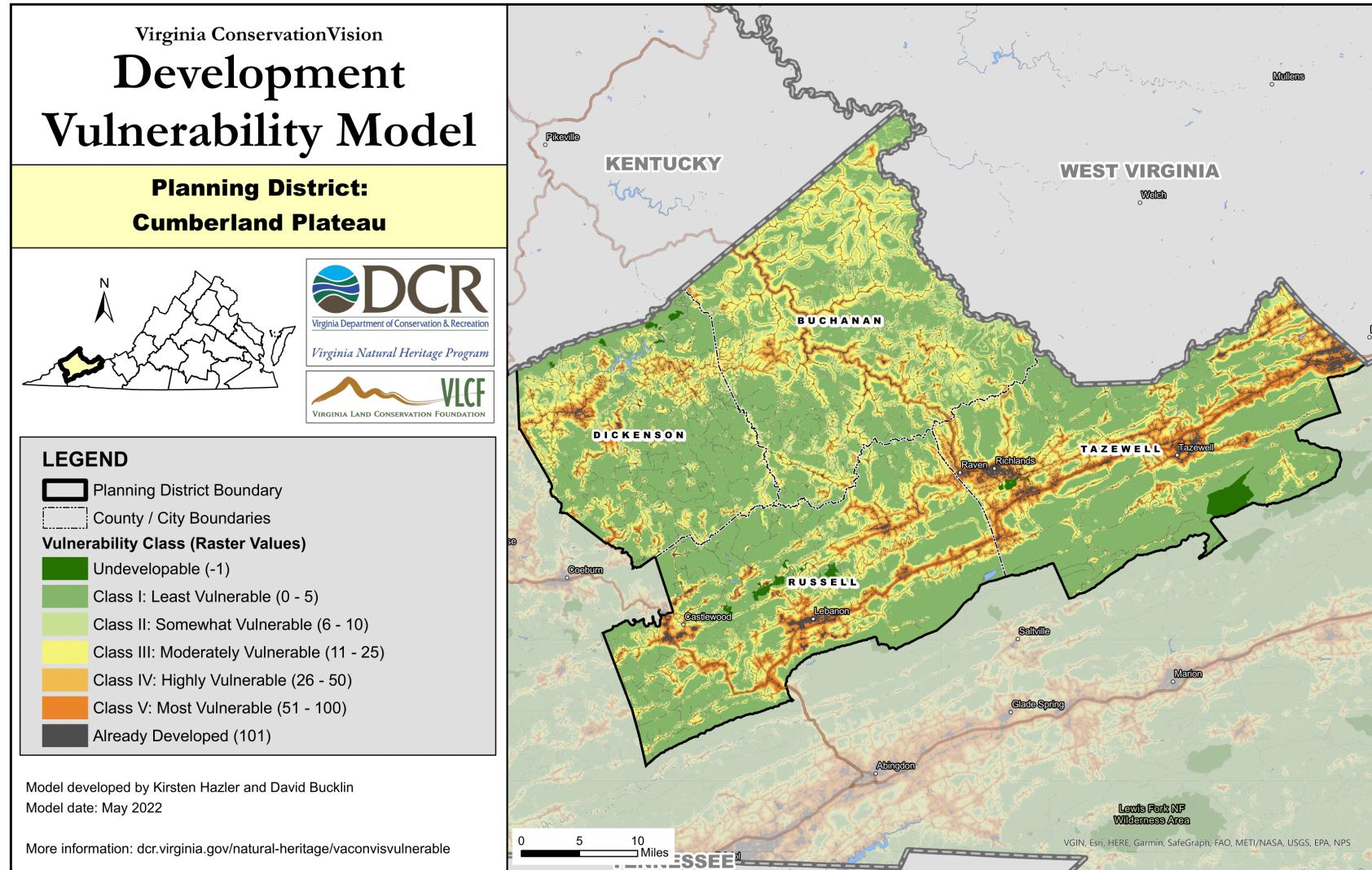
Map 4: Commonwealth Regional Council Planning District



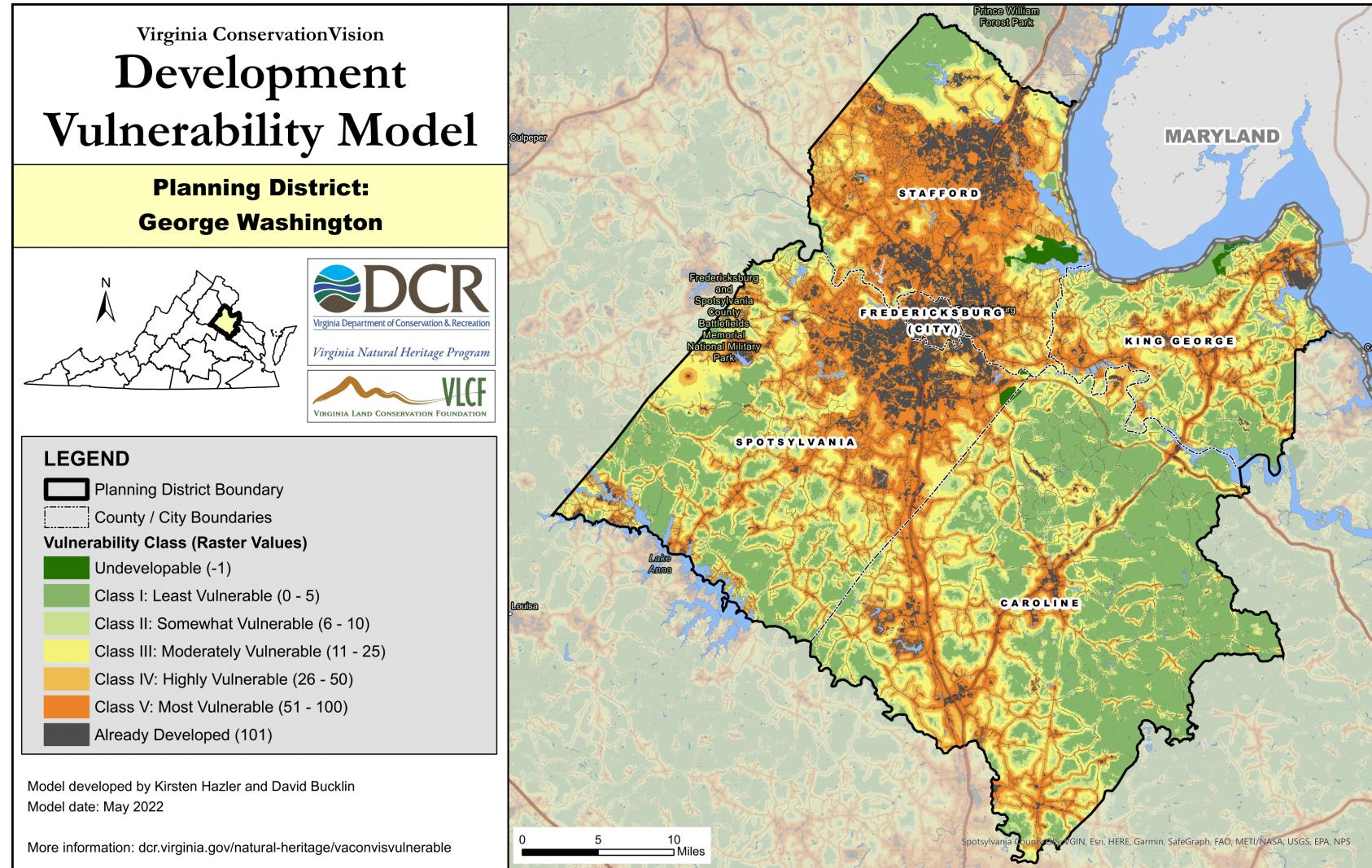
Map 5: Crater Planning District



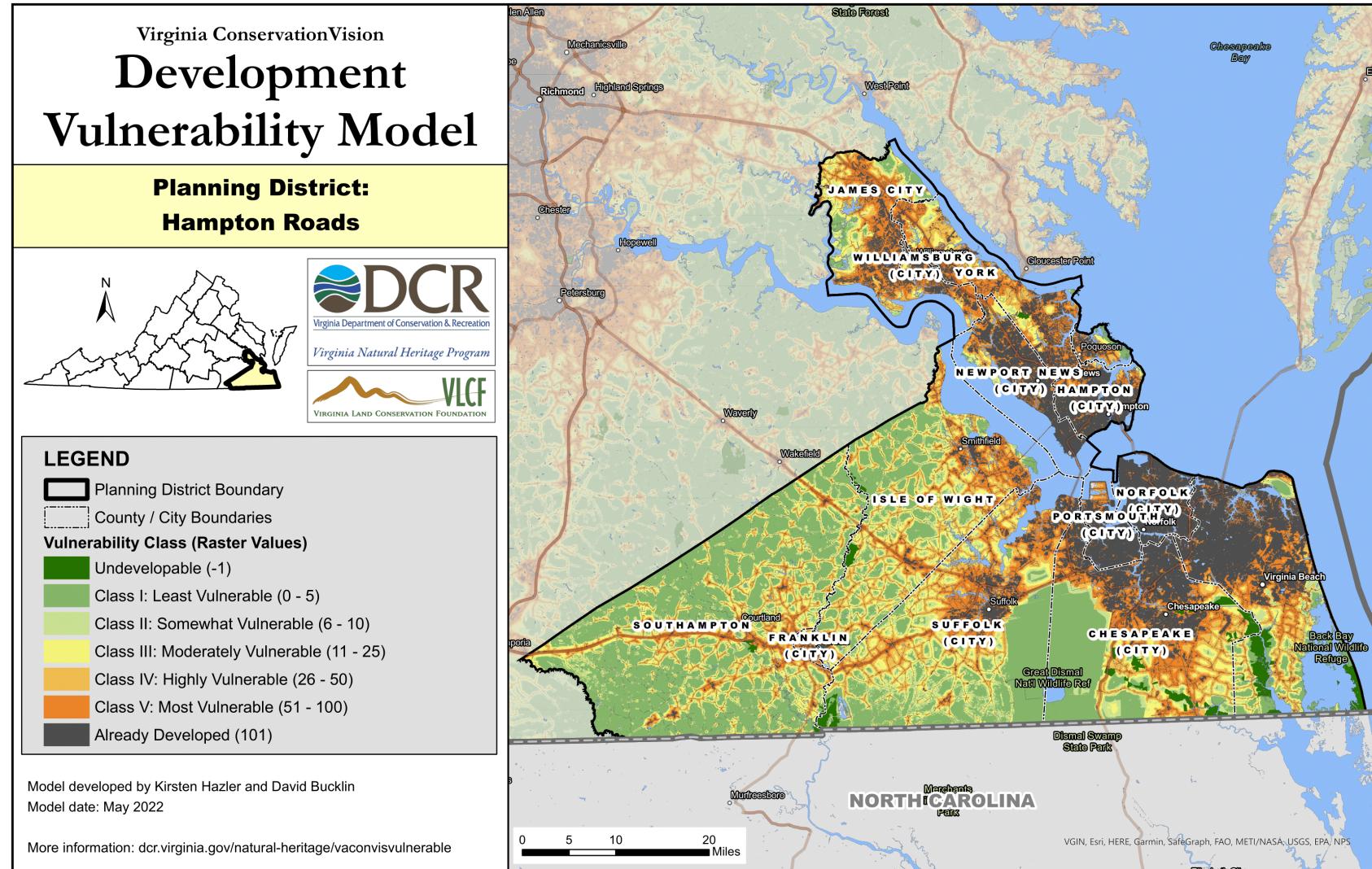
Map 6: Cumberland Plateau Planning District



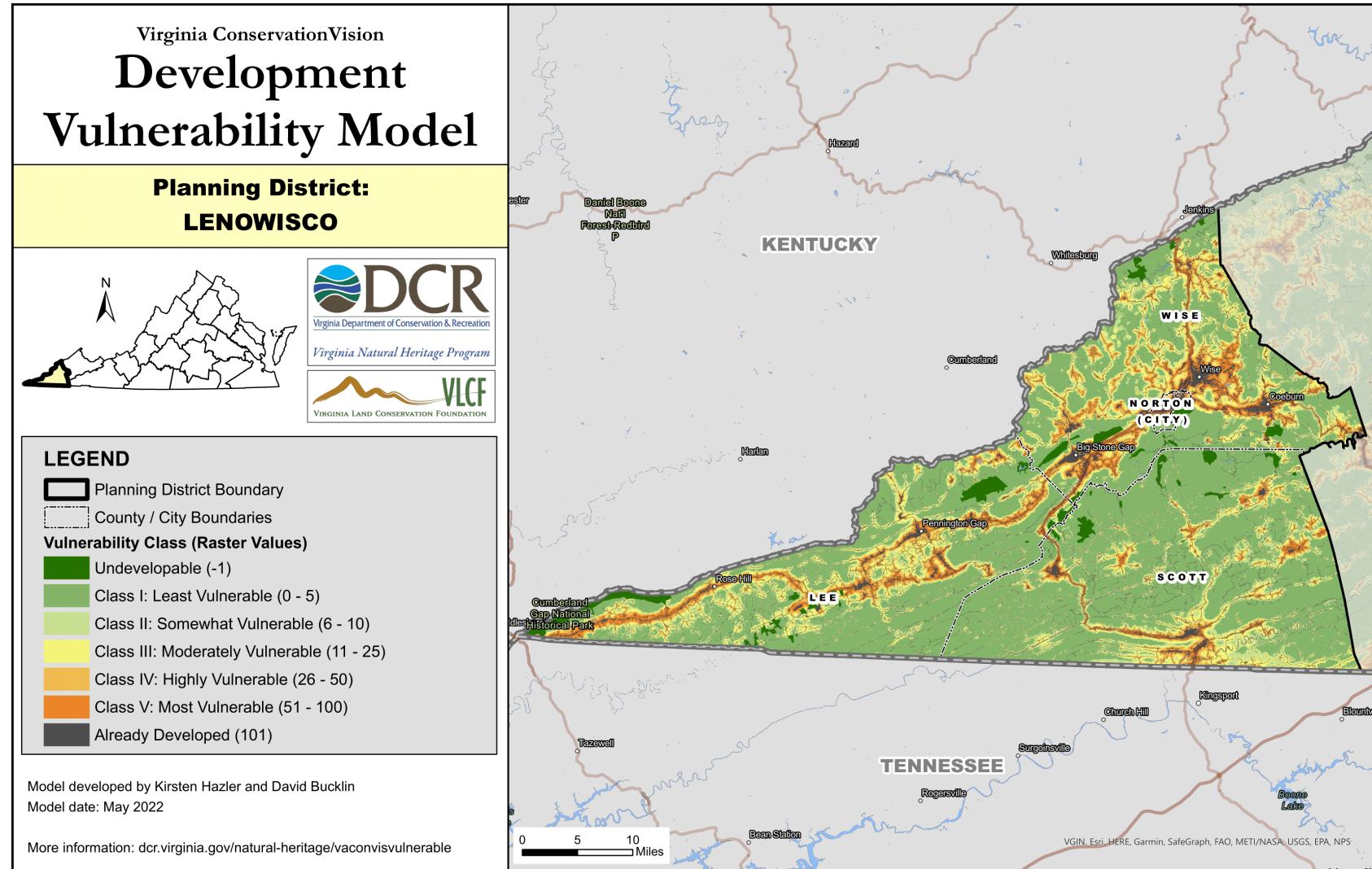
Map 7: George Washington Planning District



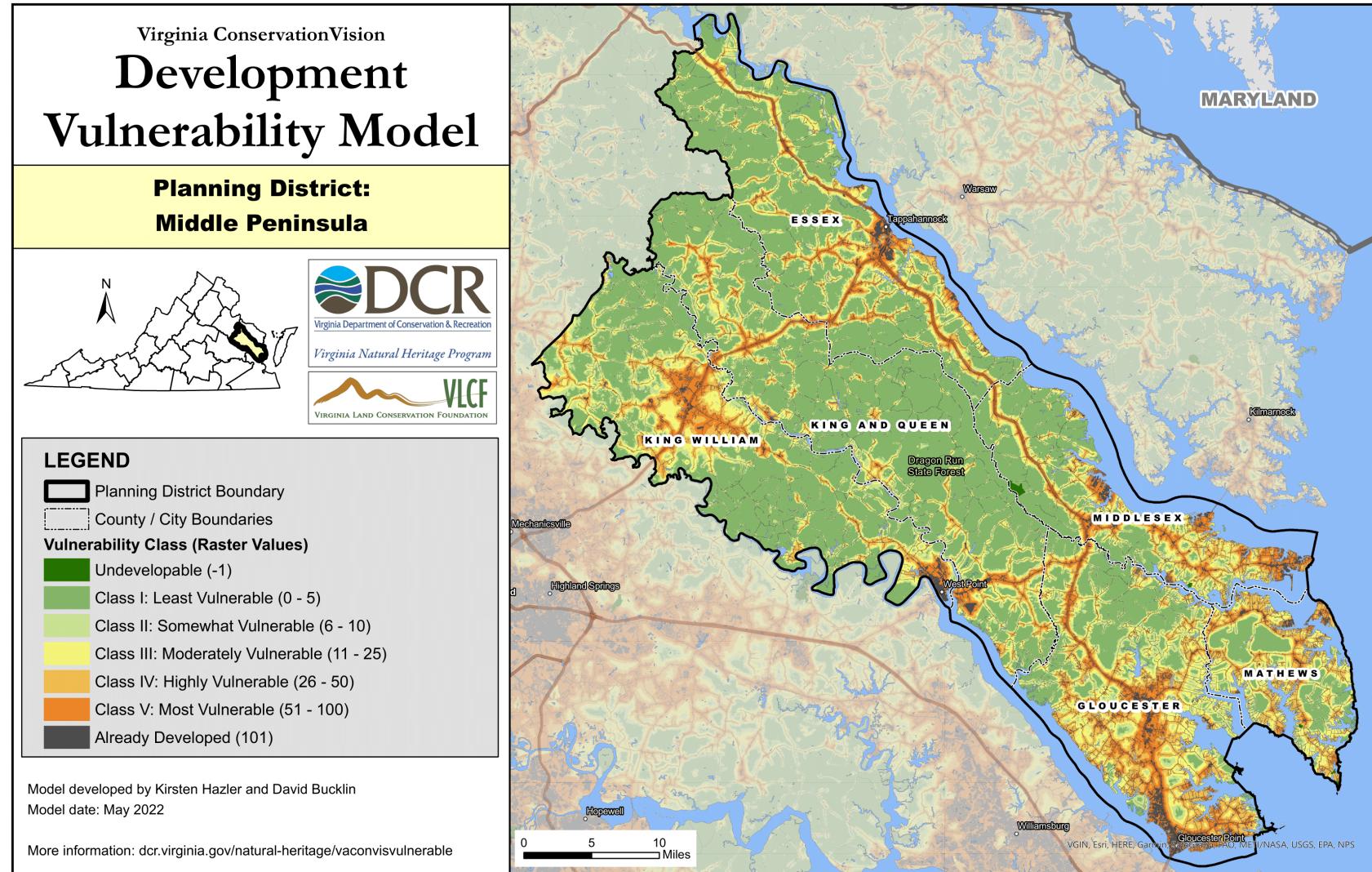
Map 8: Hampton Roads Planning District



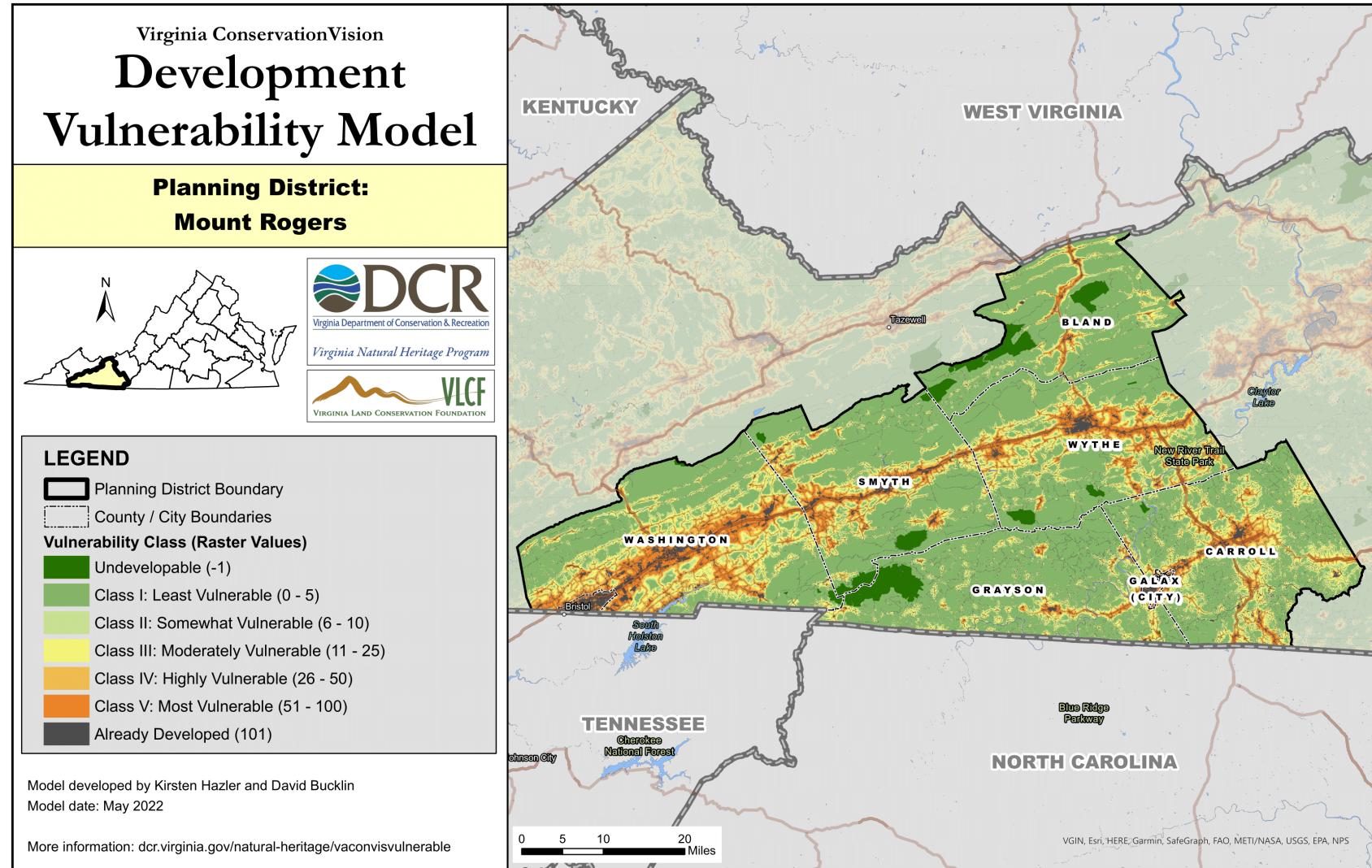
Map 9: LENOWISCO Planning District



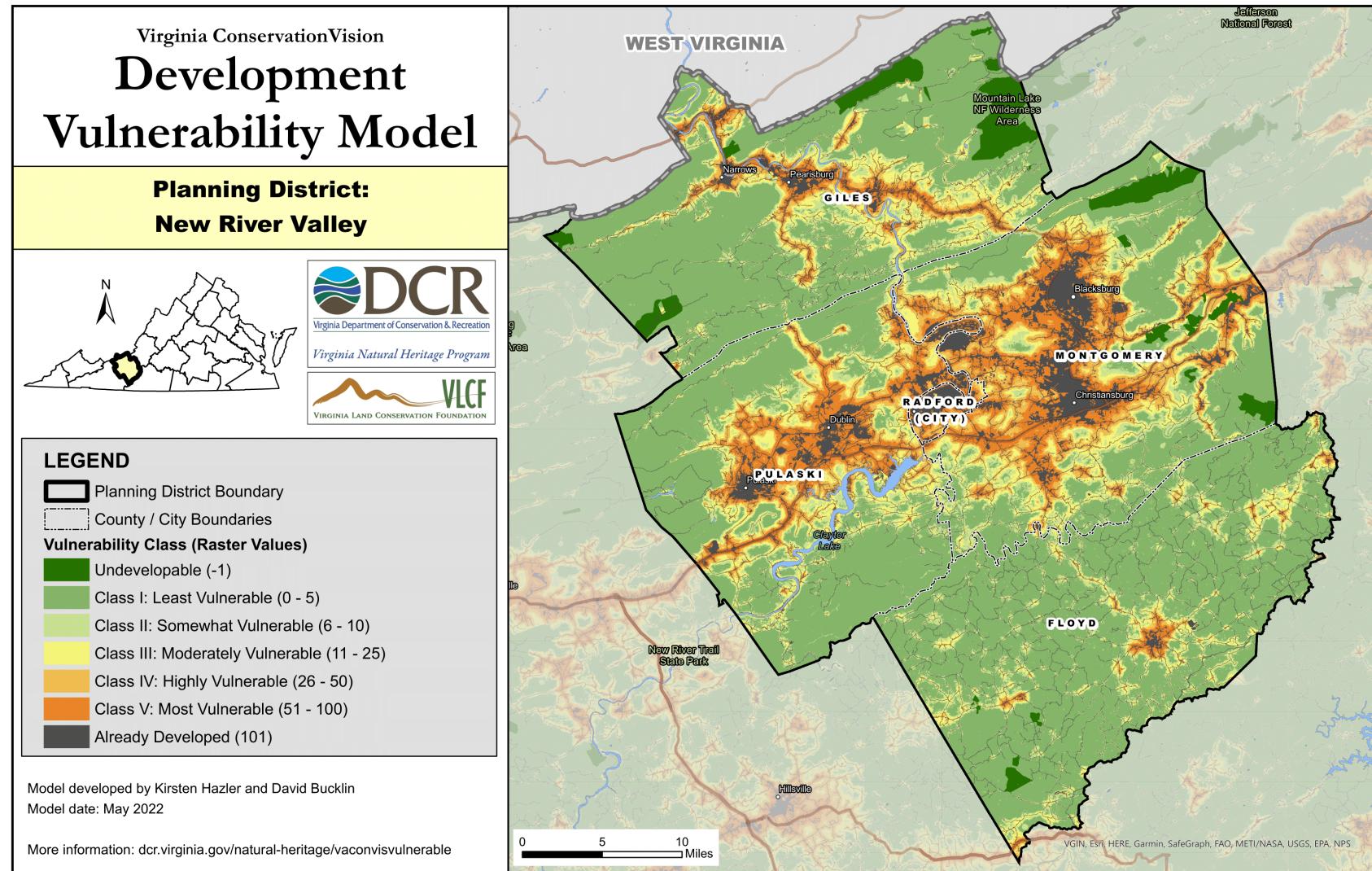
Map 10: Middle Peninsula Planning District



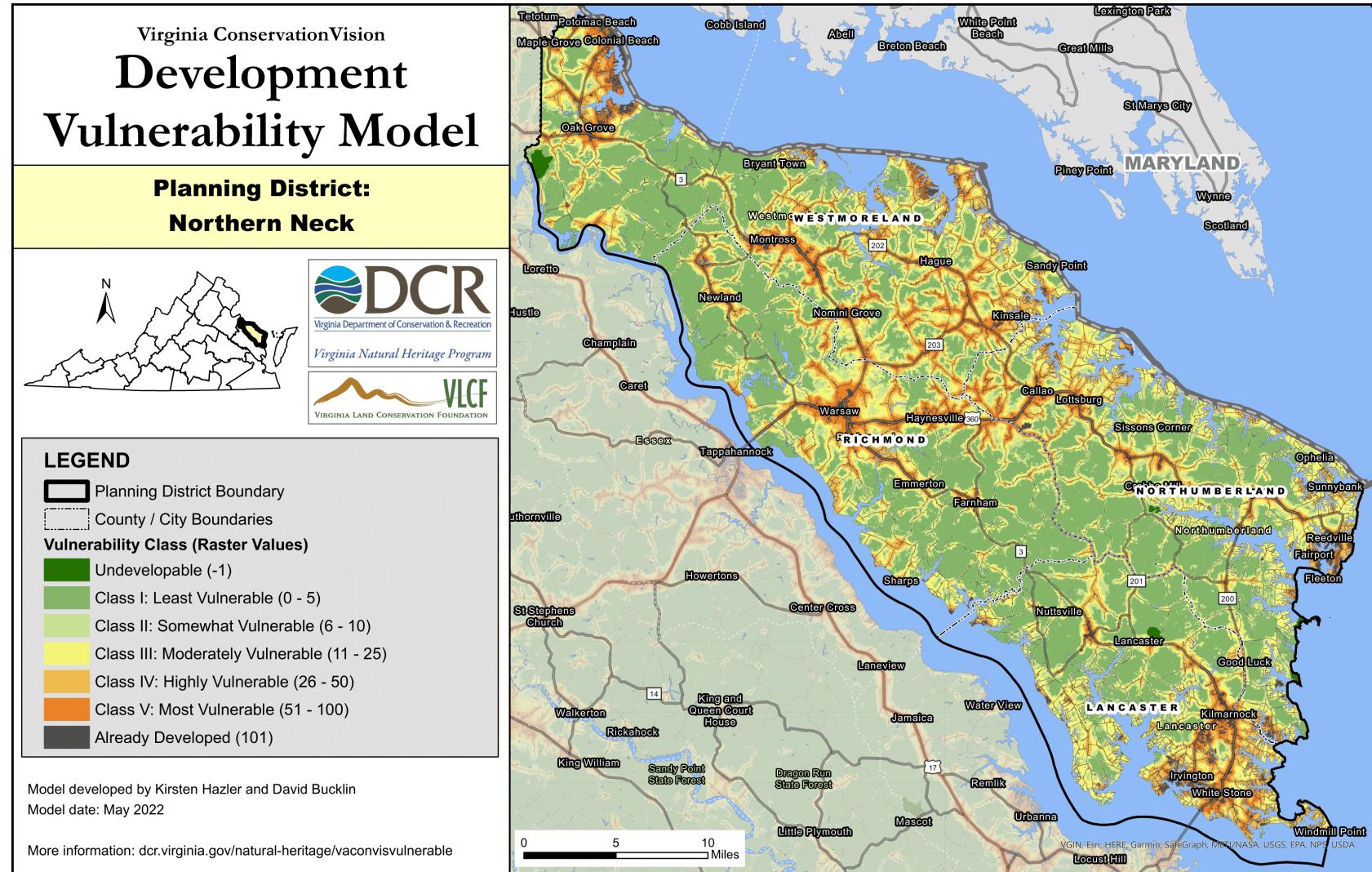
Map 11: Mount Rogers Planning District



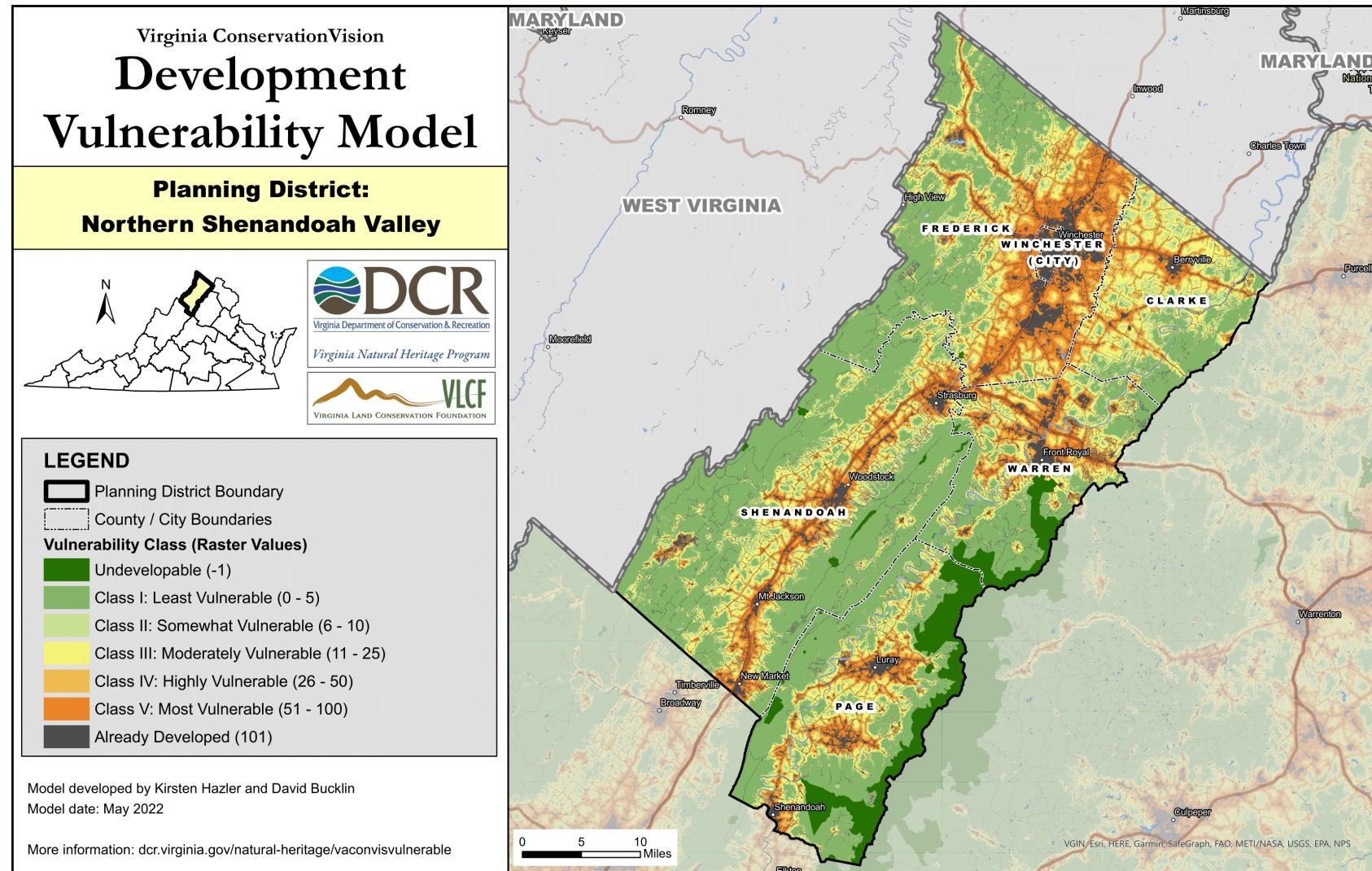
Map 12: New River Valley Planning District



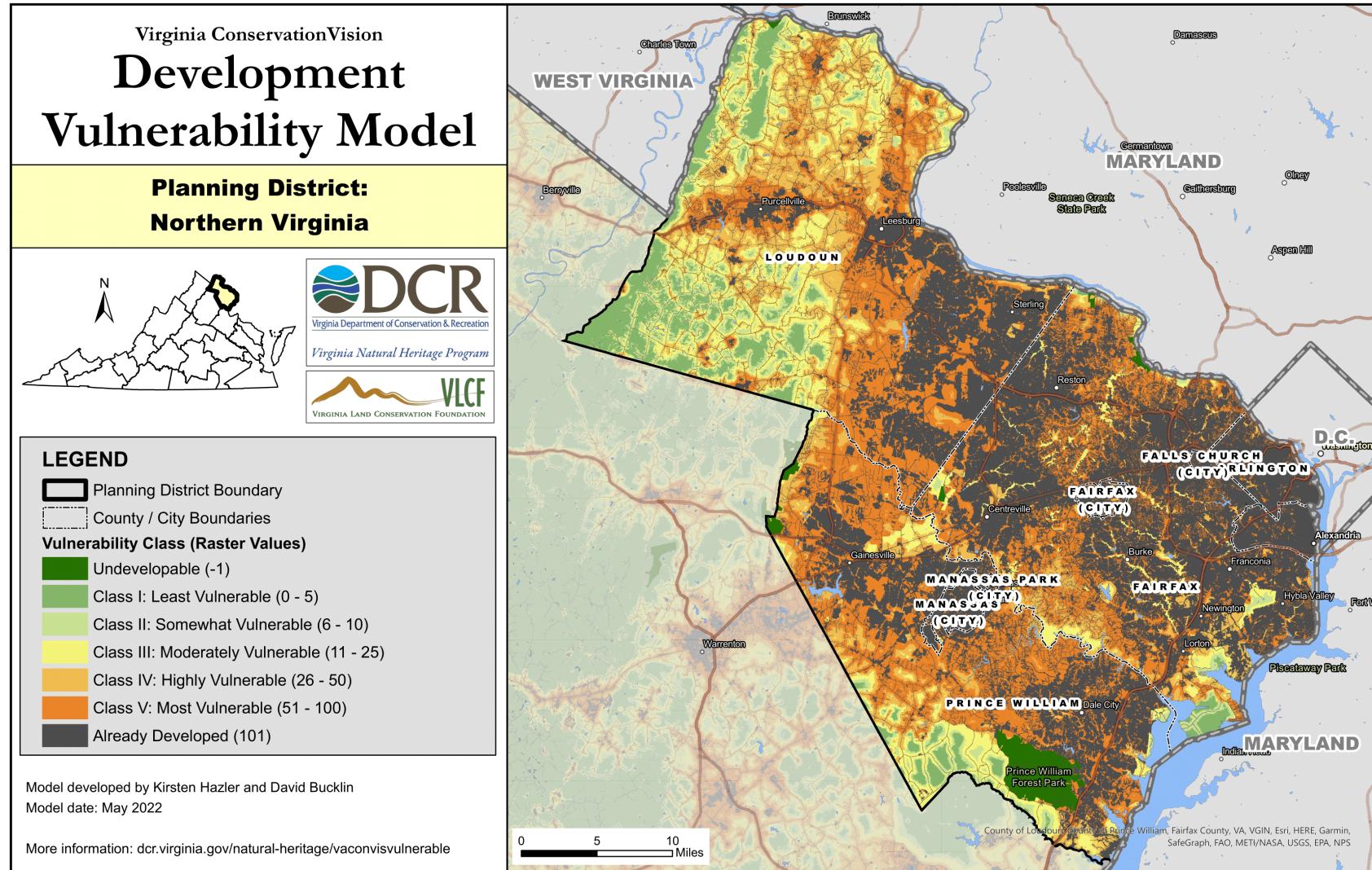
Map 13: Northern Neck Planning District



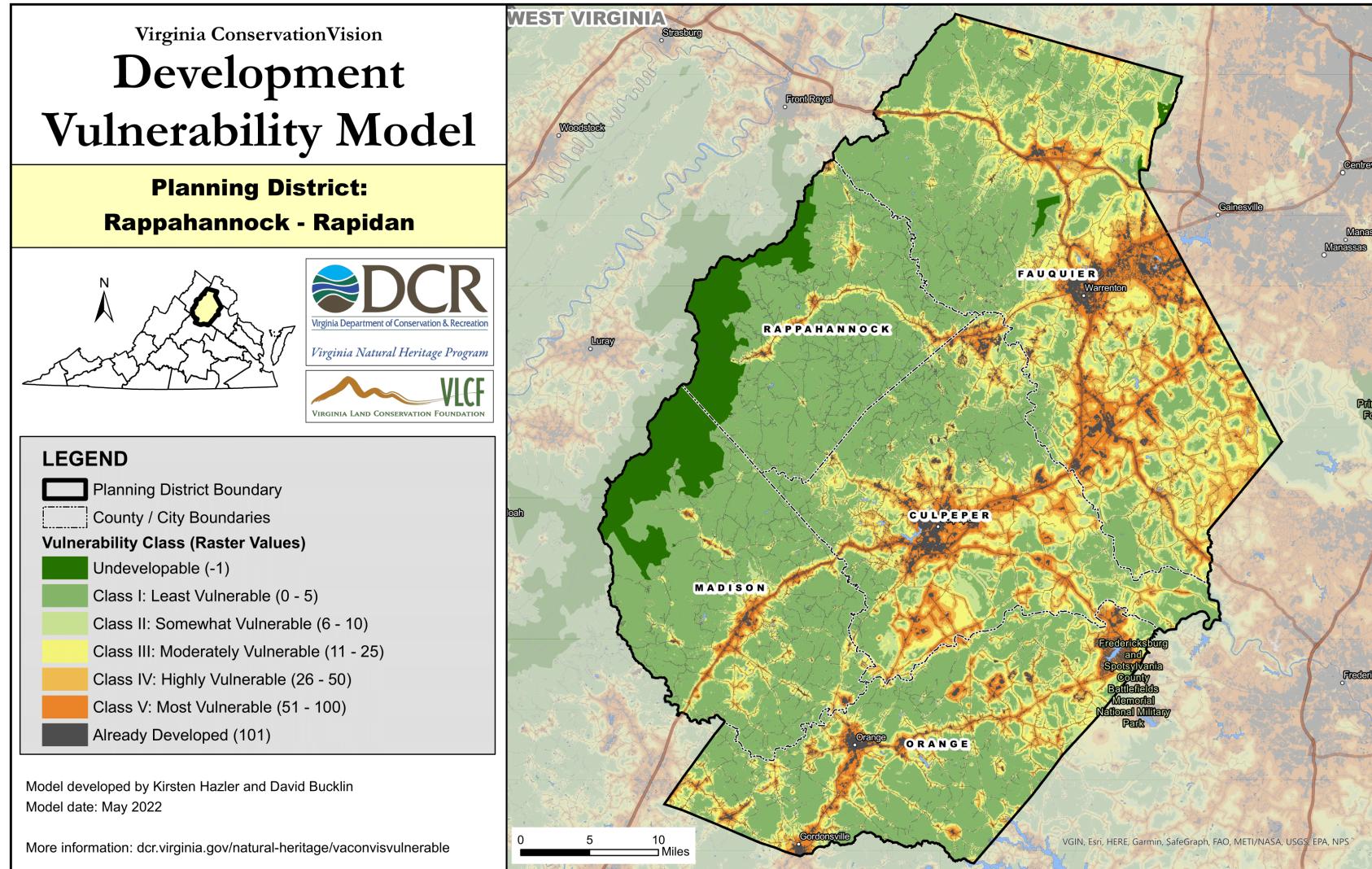
Map 14: Northern Shenandoah Valley Planning District



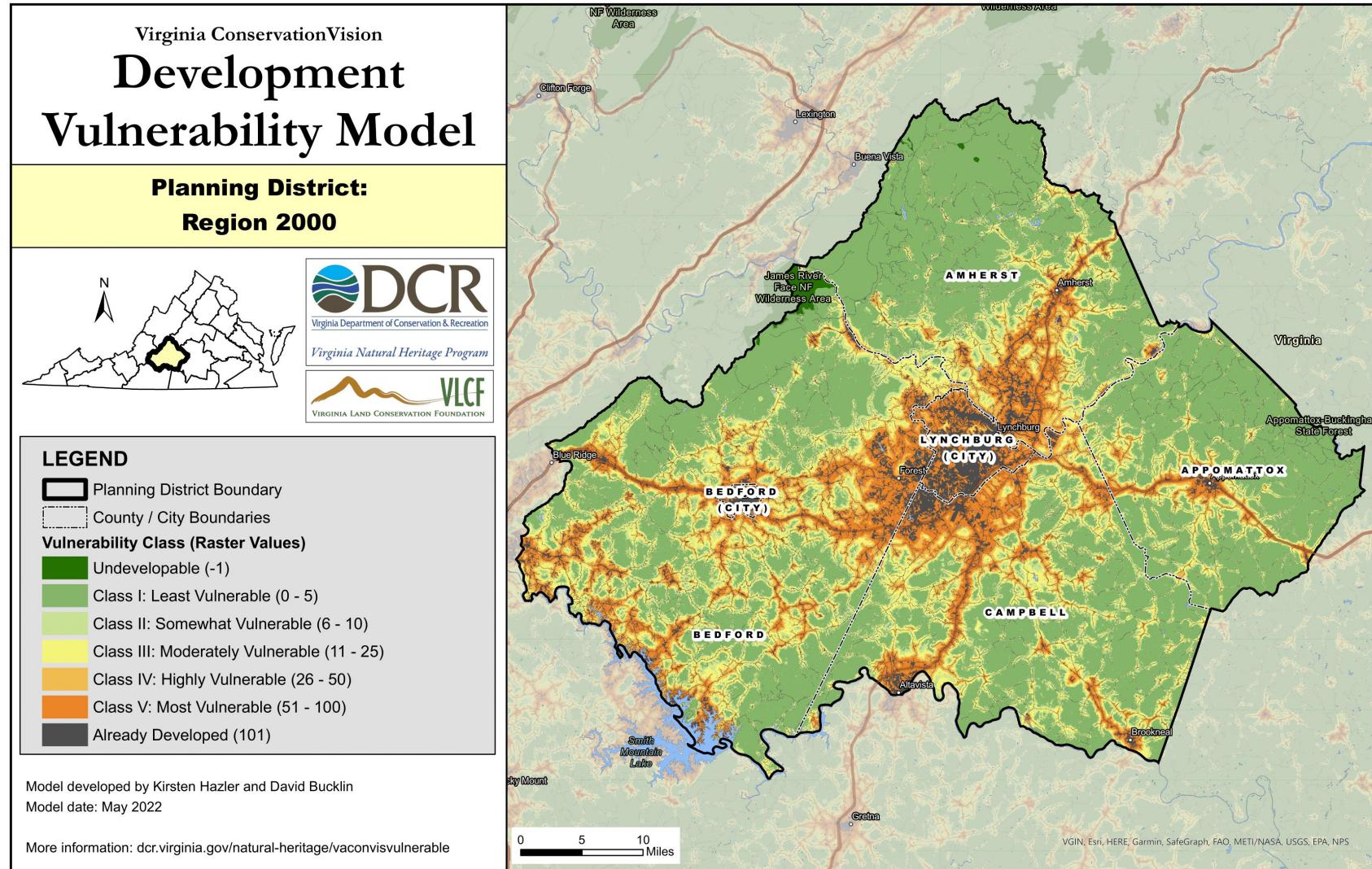
Map 15: Northern Virginia Planning District



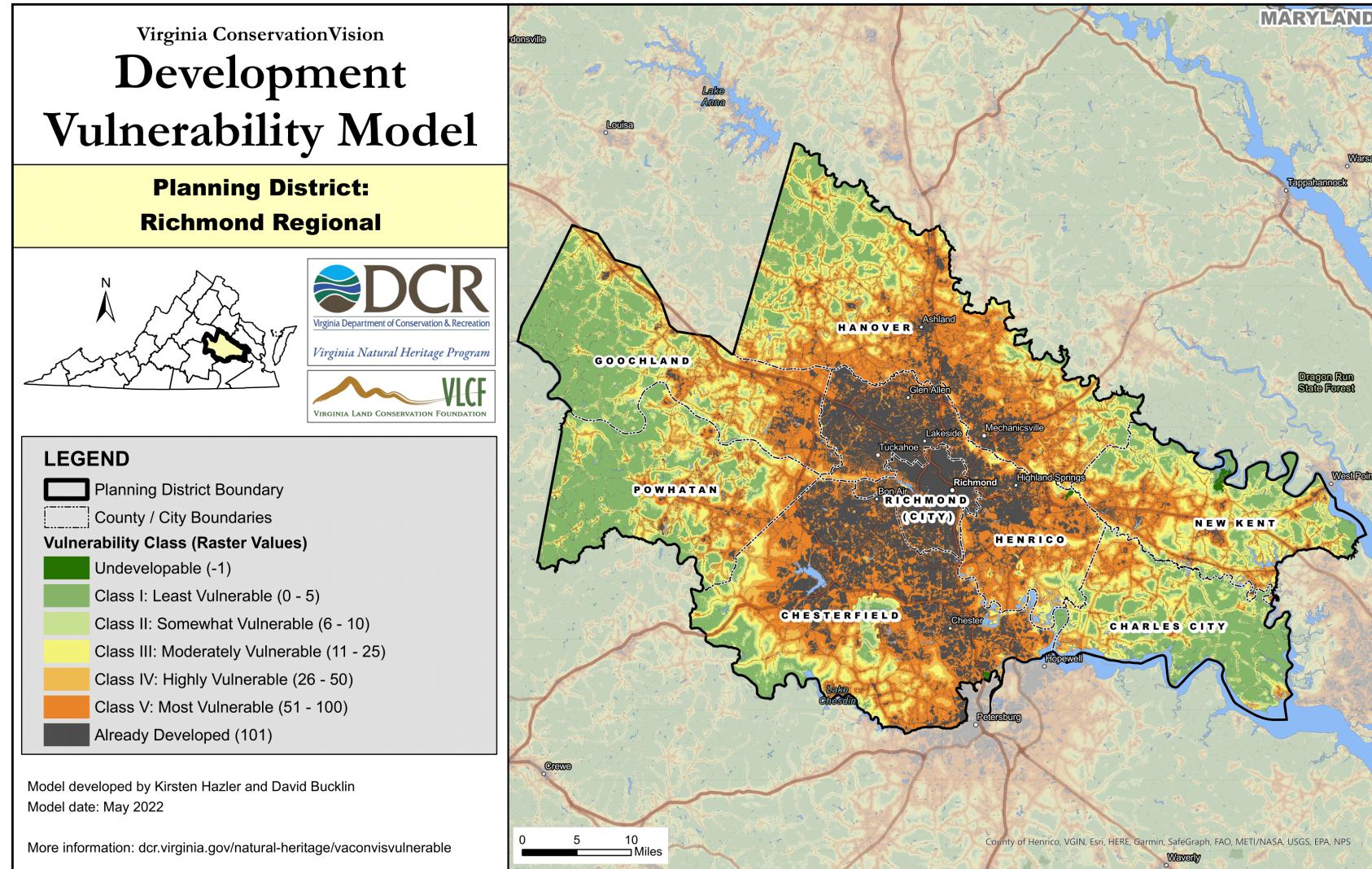
Map 16: Rappahannock - Rapidan Planning District



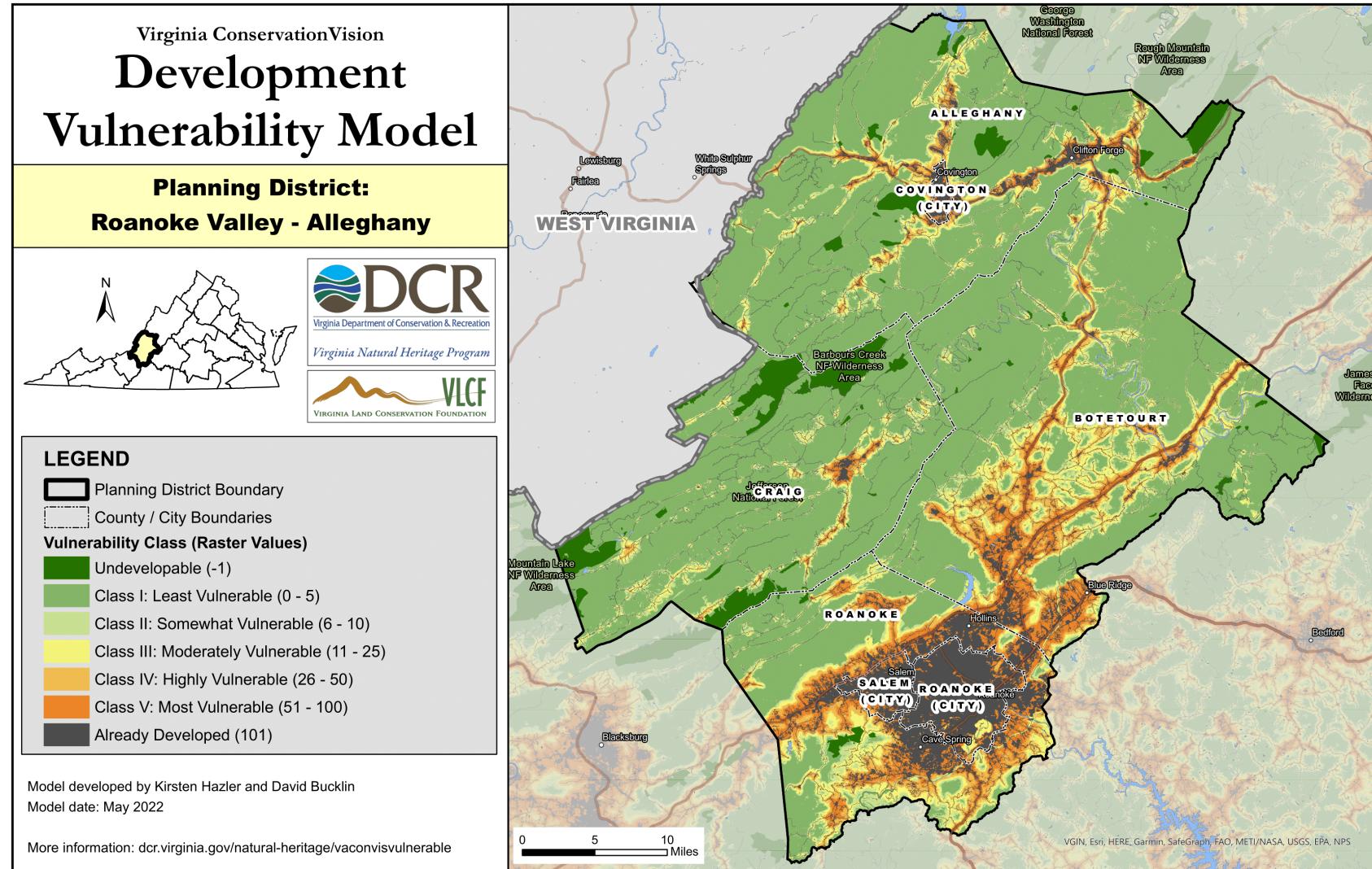
Map 17: Region 2000 Planning District



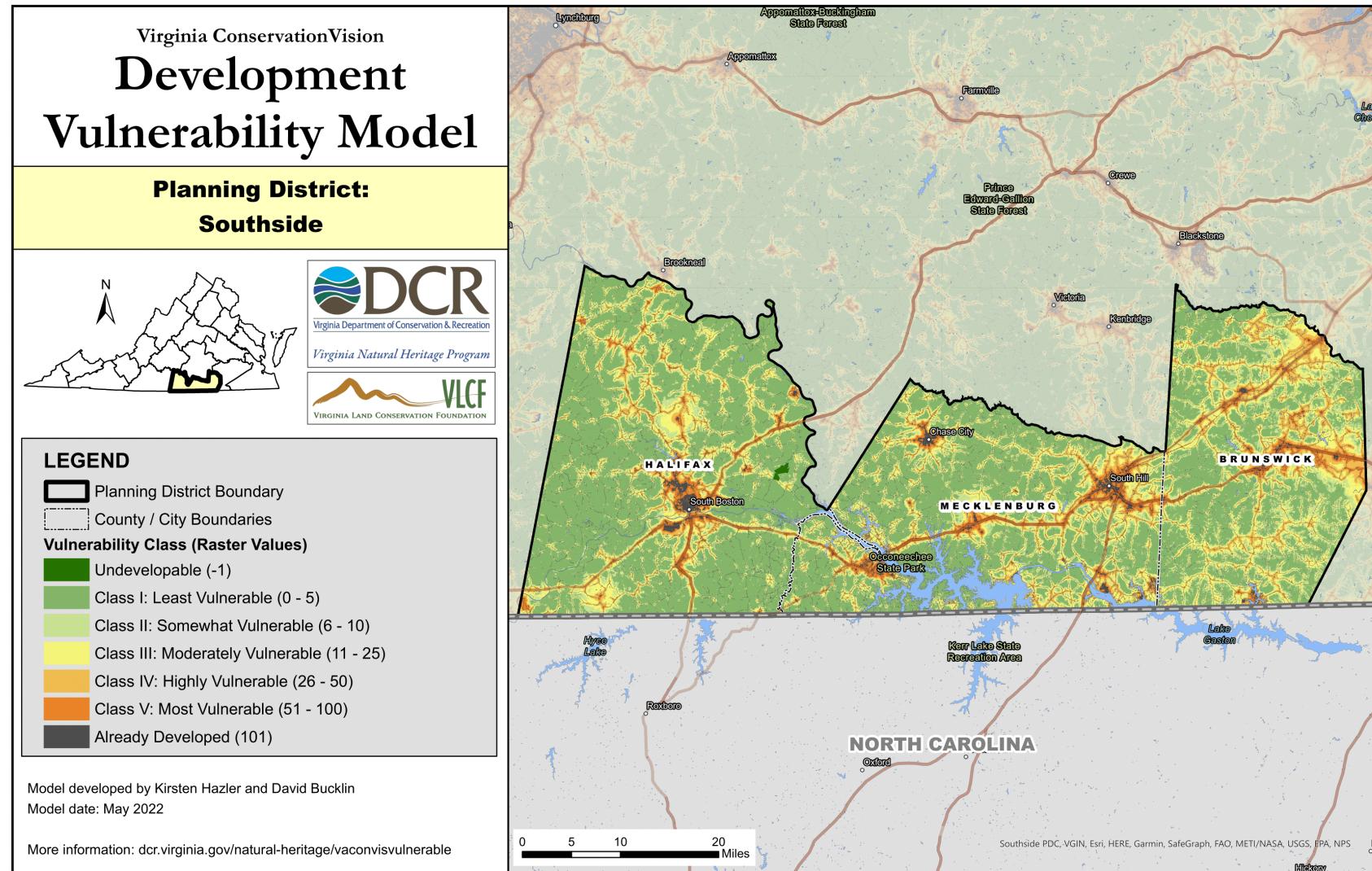
Map 18: Richmond Regional Planning District



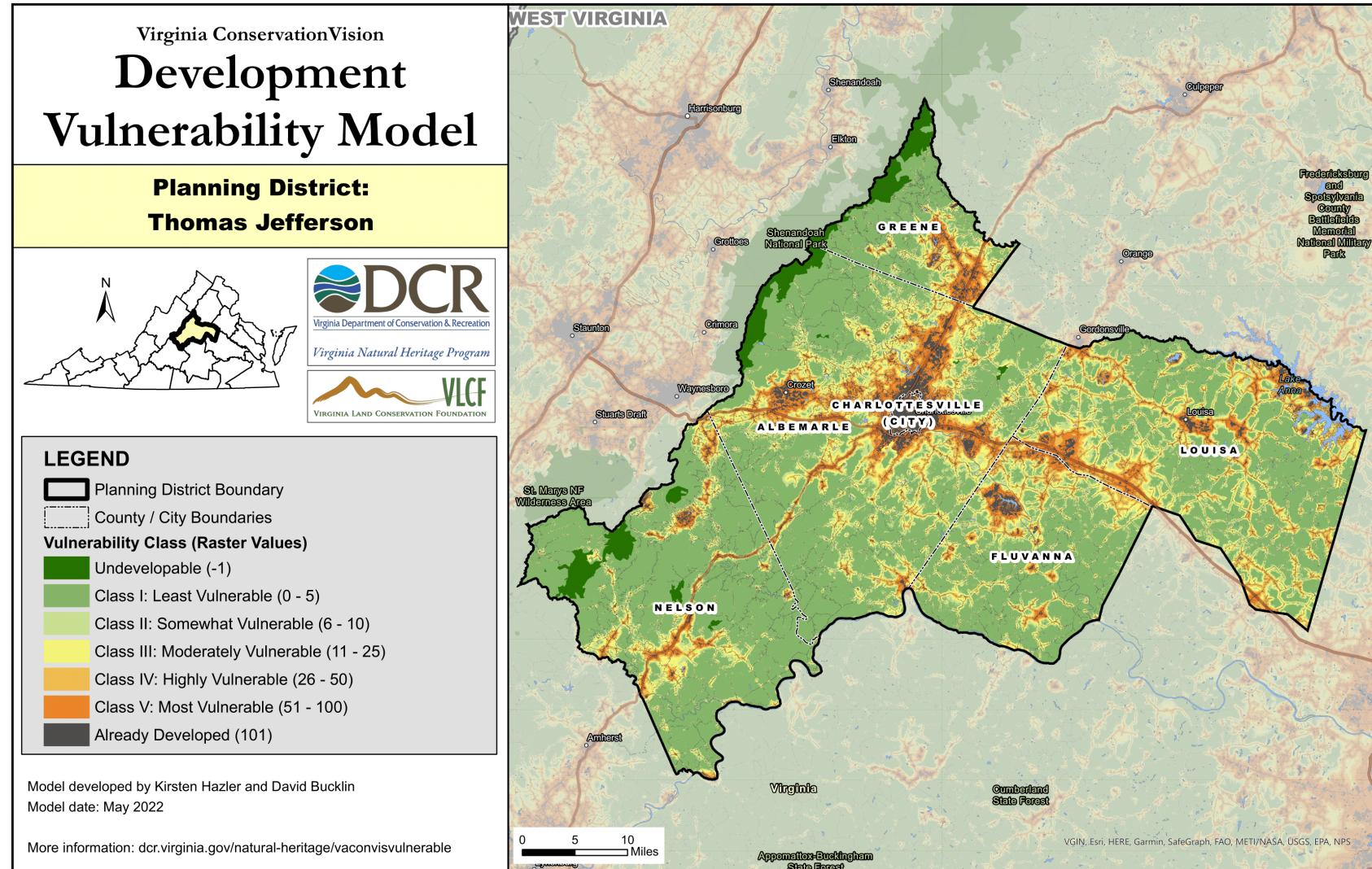
Map 19: Roanoke Valley - Alleghany Planning District



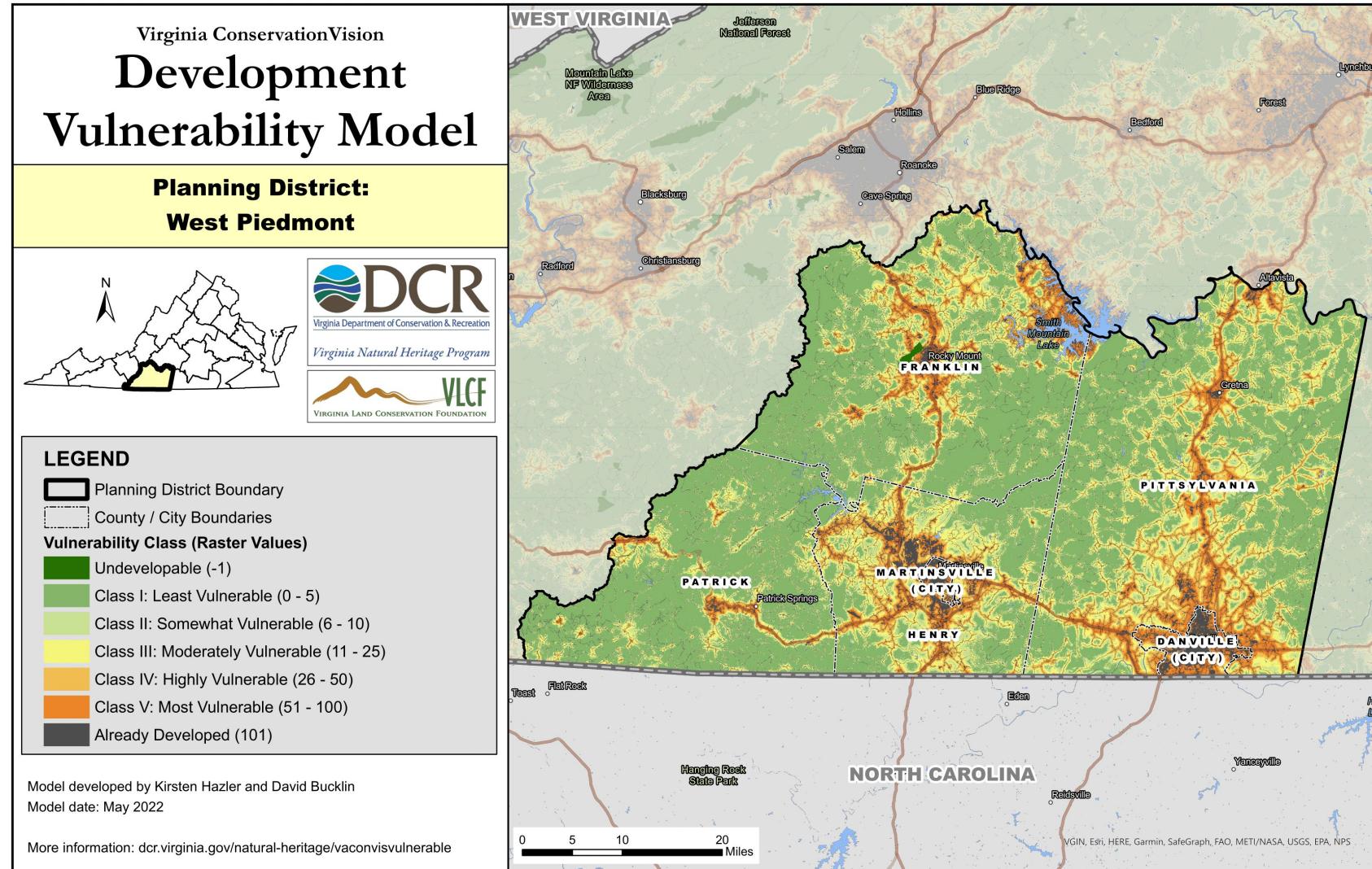
Map 20: Southside Planning District



Map 21: Thomas Jefferson Planning District



Map 22: West Piedmont Planning District



Acknowledgements

We thank Peter Claggett, Doug Shoemaker, and Jim Westervelt for fruitful discussions and insights on vulnerability modeling and David Boyd for the provision of data from the Virginia Conservation Lands Database.

References

- Breiman, L. 2001. Random Forests. *Machine Learning* 45:5–32.
- Breiman, L., A. Cutler, A. Liaw, and M. Wiener. 2014. Package “randomForest” [software].
- Chaudhuri, G., and K. Clarke. 2013. The SLEUTH Land Use Change Model: A Review. *Environmental Resources Research* 1:88–105.
- Chesapeake Bay Program (CBP). 2008. Vulnerability - Resource Lands Assessment for the Chesapeake Bay Watershed [map].
- Chesapeake Bay Program (CBP). 2020. Chesapeake Bay Land Change Model Current Zoning Vulnerability [raster dataset].
- Claggett, P. R., L. Ahmed, F. M. Irani, S. McDonald, and R. L. Thompson. In preparation. The Chesapeake Bay Land Change Model (CBLCM): A stochastic model for simulating land use change related to water quality.
- Claggett, P. R., C. A. Jantz, S. J. Goetz, and C. Bisland. 2004. Assessing development pressure in the Chesapeake Bay watershed: an evaluation of two land-use change models. *Environmental monitoring and assessment* 94:129–146.
- Couronné, R., P. Probst, and A.-L. Boulesteix. 2018. Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinformatics* 19:270.
- Dewitz, J. and USGS. 2021. National Land Cover Database (NLCD) 2019 Products. U.S. Geological Survey.
- ESRI. 2021. ArcGIS Pro (Version 2.8+) [software]. Environmental Systems Resource Institute, Redlands, California.
- Fielding, A. H., and J. F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24:38–49.
- Hapfelmeier, A., T. Hothorn, K. Ulm, and C. Strobl. 2014. A new variable importance measure for random forests with missing data. *Statistics and Computing* 24:21–34.

- Hazler, K., T. Tien, and R. Gilb. 2016. Virginia ConservationVision: Development Vulnerability Model, 2015 Interim Edition. Natural Heritage Technical Report 16-14, Dept. of Conservation and Recreation, Division of Natural Heritage, Richmond, Virginia.
- Irwin, E. G., K. P. Bell, and J. Geoghegan. 2003. Modeling and managing urban growth at the rural-urban fringe: A parcel-level model of residential land use change. *Agricultural and Resource Economics Review* 32.
- Jantz, C. A., S. J. Goetz, D. Donato, and P. Claggett. 2010. Designing and implementing a regional urban modeling system using the SLEUTH cellular urban model. *Computers, Environment, and Urban Systems* 34:1–16.
- McGarigal, K., E. B. Plunkett, L. L. Willey, B. W. Compton, W. V. DeLuca, and J. Grand. 2018. Modeling non-stationary urban growth: The SPRawl model and the ecological impacts of development. *Landscape and Urban Planning* 177:178–190.
- Meentemeyer, R. K., W. Tang, M. A. Dorning, J. B. Vogler, N. J. Cunniffe, and D. A. Shoemaker. 2013. FUTURES: Multilevel simulations of emerging urban–rural landscape structure using a stochastic patch-growing algorithm. *Annals of the Association of American Geographers* 103:785–807.
- North Atlantic Landscape Conservation Cooperative (NALCC). 2017. Probability of Development, 2030, Version 3.1, Northeast U.S. [raster dataset].
- Prior-Magee, J. S., L. J. Johnson, M. J. Croft, M. L. Case, C. M. Belyea, and M. L. Voge. 2020. Protected Areas Database of the United States (PAD-US) 2.1 (Provisional Release) [vector dataset]. U.S. Geological Survey.
- R Core Team. 2021. R: A language and environment for statistical computing [software]. R Foundation for Statistical Computing, Vienna, Austria.
- Riley, S., S. Degloria, and S. D. Elliot. 1999. A terrain ruggedness index that quantifies topographic heterogeneity. *Internation Journal of Science* 5:23–27.
- Saito, T., and M. Rehmsmeier. 2015. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLoS ONE* 10:e0118432.
- Sohl, T., and K. Sayler. 2008. Using the FORE-SCE model to project land-cover change in the southeastern United States. *Ecological Modelling* 219:49–65.
- Soil Survey Staff. 2020. Gridded Soil Survey Geographic (gSSURGO) Database for Virginia, Washington DC, Delaware, Kentucky, Maryland, North Carolina, Pennsylvania, Tennessee, and West Virginia (FY 2020 Release). United States Department of Agriculture, Natural Resources Conservation Service.

- Solar Energy Industries Association (SEIA). 2022. State Solar Spotlight: Virginia [factsheet].
- Southeast Regional Assessment Project (SERAP). 2016. SLEUTH Projected Urban Growth [raster datasets].
- Strobl, C., A.-L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis. 2008. Conditional variable importance for random forests. *BMC Bioinformatics* 9:307.
- Terando, A. J., J. Costanza, C. Belyea, R. R. Dunn, A. McKerrow, and J. A. Collazo. 2014. The southern megalopolis: Using the past to predict the future of urban sprawl in the southeast U.S. *PLoS ONE* 9:e102261.
- U.S. Census Bureau. 2011. 76 FR 53029: Urban Area Criteria for the 2010 Census. *Federal Register* 76:53029–53043.
- U.S. Census Bureau. 2021. TIGER/Line Shapefiles Technical Documentation.
- U.S. Dept. of Agriculture, Natural Resources Conservation Service (NRCS). 2017. Soil Data Management Toolbox for ArcGIS, version 5.0: User Guide. National Soil Survey Center, National Geospatial Center of Excellence.
- U.S. Geological Survey (USGS). 2017. 1/3 arc-second Digital Elevation Models (DEMs) - USGS National Map 3DEP Downloadable Data Collection for Virginia [raster dataset].
- U.S. Geological Survey (USGS). 2018. USGS National Hydrography Dataset Plus High Resolution (NHDPlus HR) for 4-digit Hydrologic Units in Virginia [geodatabase].
- Virginia DCR Staff. 2008. Virginia Conservation Lands Needs Assessment: Virginia Vulnerability Model. Unpublished technical report, Dept. of Conservation and Recreation, Division of Natural Heritage, Richmond, Virginia.
- Virginia DCR Staff. 2016. Virginia Department of Conservation and Recreation Strategic Plan.
- Virginia DCR Staff. 2021. Virginia Conservation Lands Database [vector dataset].
- Virginia DCR Staff. (n.d.). Virginia Land Conservation Foundation [web site].
<http://www.dcr.virginia.gov/virginia-land-conservation-foundation/>.
- Westervelt, J., T. BenDor, and J. Sexton. 2011. A technique for rapidly forecasting regional urban growth. *Environment and Planning B: Planning and Design* 38:61–81.
- Yang, L., S. Jin, P. Danielson, C. Homer, L. Gass, S. M. Bender, A. Case, C. Costello, J. Dewitz, J. Fry, M. Funk, B. Granneman, G. C. Liknes, M. Rigge, and G. Xian. 2018. A new generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies. *ISPRS Journal of Photogrammetry and Remote Sensing* 146:108–123.