Supporting Information

McKee et al. 10.1073/pnas.1405713112

SI Materials and Methods

Cohort-Component Method. The cohort-component method is rooted in the basic demographic equation for change over time in a population:

$$P_{x+n} = P_x + \left(B_{x,x+n} - D_{x,x+n}\right) + \left(IM_{x,x+n} - OM_{x,x+n}\right),$$

where P = population, B = births, D = deaths, IM/OM = in/out-migration, x = current time, and $x + n$ = future time. In other words, future population equals current population plus births, less deaths, plus net migration. (The counts for births, deaths, and migration are the number of each that occurred in the x to $x + n$ interval.) Age- and sex-specific numbers and rates for births and deaths were used for the birth and death components of the equation. However, age- or sex-specific migration rates were not available, so the same migration rate was assumed for all age and sex groups. Rates from the most recent year available are assumed to stay the same throughout the projection period.

All of the data were from the US Census 2010 (1, 2), the National Center for Health Statistics (NCHS) (3, 4), and the Internal Revenue Service (IRS) (5). The base-year population count for 2010 was from the US Census 2010 (6). These data were for population count for each county by 5-y cohorts from ages 0–4 to 85–89 with 90+ grouped together. The birth data (3) are from NCHS and are for 2009, the most recent year available. The number of live births was provided by 5-y age cohorts of the mother from ages 10–14 to 50–54. Mortality data (4) were from NCHS as well. Because mortality data were to be used for each sex for each 5-y cohort up to age 90+, data for years 2005–2009 were grouped together because counties with smaller populations would have had more unreliable rates otherwise. These data had the number of deaths by 5-y age cohorts from age 5–9 to 95–99 with 100+ grouped together and <1 separated from 1 to 4; to match with the US Census 2010 population age cohorts, age cohorts 90–94, 95–99, and 100+ were summed together for ages 90+ and <1 and 1–4 were combined to form age cohort 0–4. The migration data (5) are from the IRS and are for 2009–2010, the most recent year available. These data were based on year-to-year address changes as reported on tax returns filed with the IRS. The numbers of inflow, outflow, and nonmigrants for each county are provided as the number of tax returns (i.e., the approximate number of households) and the number of exemptions (i.e., approximate number of individuals). Therefore, because we are interested in the migration rate of individuals, the number of exemptions was used for the inflow, outflow, and nonmigrant figures.

Fertility. Birth data from NCHS were combined with population count data from the Census 2010 to calculate age-specific fertility rates (ASFRs). The ASFR provides the rate at which babies are born to women of specific age cohorts and is helpful in differentiating the varying fertility rates of women at different ages. It is calculated as follows:

$$ASFR_{x,x+n} = \frac{B_{x,x+n}}{P_{F_{x,x+n}}},$$

where x = the starting year of an age cohort, n = the length of the age cohort, B = number of births, P = population count, and F = the female population. It should be noted that the birth data are from 2009 whereas the population count data are from 2010, so the ASFRs might be slightly lower due to the general notion that 2010 population is higher than 2009 population. Birth data were also only available for counties with a population of 100,000 or more; the number of births was aggregated for all small counties for each state. Therefore, for the purposes of projecting population for these smaller counties, the age-specific fertility rates were assumed to be the same for counties with less than 100,000 people within each state.

Mortality. Death data from NCHS were partially combined with population count data from the Census 2010 to calculate age-specific survivability rates (ASSRs) for each sex. The ASSR provides the rate at which people within one age interval survive to the next age interval. For example, the ASSR for most ages will be upwards of 99%, whereas the ASSR for new infants will tend to be slightly lower and the ASSR for the older ages will tend to drop off steadily as the age increases. It is calculated as follows:

$$ASSR_{x,x+n} = \frac{L_{x+n}}{L_x},$$

where $ASSR_{x,x+n}$ = probability of a member of the age cohort surviving from time x to $x + n$, L_{x+n} = number of persons alive at end of period $x + n$, L_x = number of persons alive at beginning of age interval x, n = the length of the time period (in units of years). In this projection, the formula was slightly rearranged as follows:

$$ASSR_{x,x+n} = 1 - \frac{D_{x+n}}{L_x},$$

where D_{x+n} is the number of deaths within each age cohort. The rate from each formula is mathematically the same. In addition to the count of deaths, the NCHS data had population counts (of the alive population) for each age cohort up through ages 80–84. From age 85 onward, the population was "not applicable," so Census 2010 data were used, likely causing the survivability rates at these older ages to be slightly higher than they actually were. For very small counties that might not have had deaths in some age cohorts in the 2005–2009 interval, the survivability rate was assumed to be the average of the adjacent age cohort above and below; the survivability rate was assumed to be zero if there were no data in the oldest age cohort.

Migration. Migration data from the IRS were used to calculate the migration rate. These data included the numbers for total inflow migration, outflow migration, and nonmigration by county for both US and foreign people who filed tax returns. Therefore, low-income people, very-high-income people, noncitizens, and unauthorized immigrants were likely to be underrepresented in these numbers. The migration rate reflects the net change in migration over a specific time period, which in this case is 2009–2010. The migration rate by county was calculated as follows:

$$MR = \frac{IF - OF}{NM + OF},$$

where MR = the migration rate, IF = total number of inflow migrants, OF = total number of outflow migrants, and NM = total number of nonmigrants. $IF-OF$ represents the net migration number of people, whereas $NM+OF$ represents the number of people who lived in the county at the beginning of the time interval. Age-specific migration rates were not available, so the same migration rate was assumed to be the same for all age and sex groups.

Projecting Population from Fertility, Mortality, and Migration Rates. After calculating all of the rates affecting population change, the population can be projected using the assumptions that people will be born, die, and move at the same rates as the current time. Projections were calculated for every 5 y from 2010 to 2050. The method to project to 2015 is described below. This same procedure was repeated for each 5-y increment out to 2050.

First, the base-year population from the Census 2010 was inserted for the 2010 population. Because the age cohorts are in 5-y intervals, we can only project population for every 5 y because we know that everyone in each 5-y cohort will age into the next age cohort in 5 y. Similarly, a population projection for each successive year can be done if there are sufficient data for each 1-y age cohort. The procedures to project the youngest, oldest, and middle cohorts slightly differ. To project the middle age cohorts, we use the following formula:

$$\textit{For } x = \{0, 5, 10, \ldots, 80\} \textit{ at time } t,$$

$$P_{t+n_{x+n,x+2n}} = \left(P_{t_{x,x+n}} \times MR\right) \times ASSR_{x,x+n},$$

where x represents the beginning age of each age cohort at time t, $P_{t_{x,x+n}}$ represents the population count of one of the middle-age cohorts at time t, $P_{t+n_{x+n,x+2n}}$ represents the population count of that same cohort aged to time $t + n$, MR = migration rate, and $ASSR$ is the age-specific survivability rate. Therefore, for example, to project the age 5–9 cohort for 2015, the MR was applied to the age 0–4 cohort for 2010 and the result is multiplied by the ASSR. This is repeated for all age cohorts except the eldest one. In the case of the eldest cohort, we use the following formula:

$$\textit{For } x = \{85\} \textit{ at time } t,$$

$$P_{t+n_{90+}} = \left[(P_{t_{85-89}} \times MR) \times ASSR_{85-89}\right] + \left[(P_{t_{90+}} \times MR) \times ASSR_{90+}\right],$$

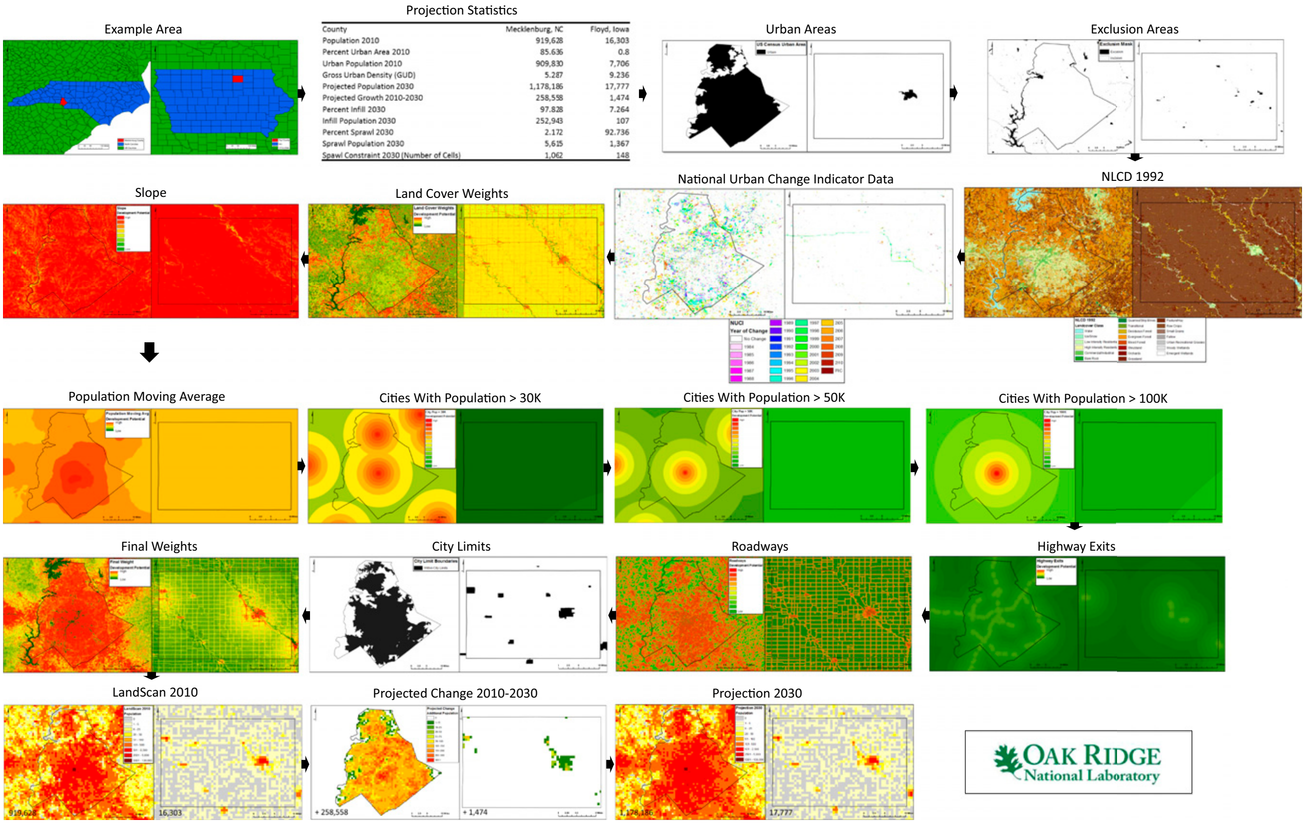
where $P_{t_{90+}}$ represents the eldest age cohort at time t, $P_{t_{85-89}}$ represents the second-eldest age cohort at time t that will age into the eldest age cohort at time $t + n$, and $P_{t+n_{90+}}$ represents the eldest age cohort at time $t + n$. So, for example, to project the age 90+ cohort for 2015, we do the same thing for the middle-age cohorts (i.e., apply the MR and ASSR for the age cohort adjacently younger in 2010) as well as adding the population that migrated and survived from the age 90+ cohort from 2010. To project new births at time $t + n$, we use the following formula:

$$\textit{For } x = \{0\} \textit{ at time } t+n,$$

$$P_{t+n_{0-4}} = 5 \times \sum_{x=10}^{50} \left[ASFR_{x,x+n} \times P_{F_{t_{x,x+n}}}\right],$$

where $x = \{10, 15, 20, \ldots, 50\}$ *and refers to the age of the female population at time* t, where $ASFR$ = age-specific fertility rate, P_F = the female population count of a given age cohort, and $P_{t+n_{0-4}}$ represents the number of new people at time $t + n$. It is necessary to multiply everything by 5 because the summation only provides us with the number of new people for 1 y. Also, note that in this formula, x refers to the age of just the female population at time t. Therefore, for example, to determine the age 0–4 cohort for 2015, the 2010 population counts of females from age cohorts 10–14 through 50–54 are multiplied by their respective ASFRs and summed and multiplied by 5. However, because this count is the number of total new people (births) and not sex-specific, this number for each projection year is then distributed proportionally among males and females to match the age 0–4 cohort's sex ratio in 2010. The entirety of this same procedure for calculating new people was repeated for each 5-y increment out to 2050. When our county projections were summed to a national total and compared with the official US Census projections for 2030 and 2050, our method itself had a national overprojection of 9.8% for 2030 and 14.7% for 2050. The county totals were adjusted proportionately to match the US Census's official population projections for those years.

1. US Census Bureau (2004) State Interim Population Projections by Age and Sex: 2004–2030. Available at www.census.gov/population/www/projections/projectionsagesex.html. Accessed July 14, 2012.
2. US Census Bureau (2009) National Population Projections. Available at www.census.gov/population/www/projections/natproj.html. Accessed July 14, 2012.
3. National Center for Health Statistics (2014) National Vital Statistics System–Birth Data. Available at www.cdc.gov/nchs/births.htm. Accessed July 14, 2012.
4. National Center for Health Statistics (2014) National Vital Statistics System–Mortality Data. Available at www.cdc.gov/nchs/deaths.htm. Accessed July 14, 2012.
5. Internal Revenue Service (2014) SOI Tax Stats–Migration Data. Available at www.irs.gov/uac/SOI-Tax-Stats-Migration-Data. Accessed July 14, 2012.
6. US Census Bureau (2010) 2010 US Census. Available at http://www.census.gov/2010census/data/. Accessed July 14, 2012.



County	Mecklenburg, NC	Floyd, Iowa
Population 2010	919,628	16,303
Percent Urban Area 2010	85.636	0.8
Urban Population 2010	909,830	7,706
Gross Urban Density (GUD)	5.287	9.236
Projected Population 2030	1,178,186	17,777
Projected Growth 2010-2030	258,558	1,474
Percent Infill 2030	97.828	7.264
Infill Population 2030	252,943	107
Percent Sprawl 2030	2.172	92.736
Sprawl Population 2030	5,615	1,367
Spawl Constraint 2030 (Number of Cells)	1,062	148

Fig. S1. Stepwise process illustrating the application of the locally adaptive weighting scheme for two contrasting counties: Mecklenburg, NC, which is predominately urban and more populated compared with Floyd, IA, which is significantly less populated and more rural.

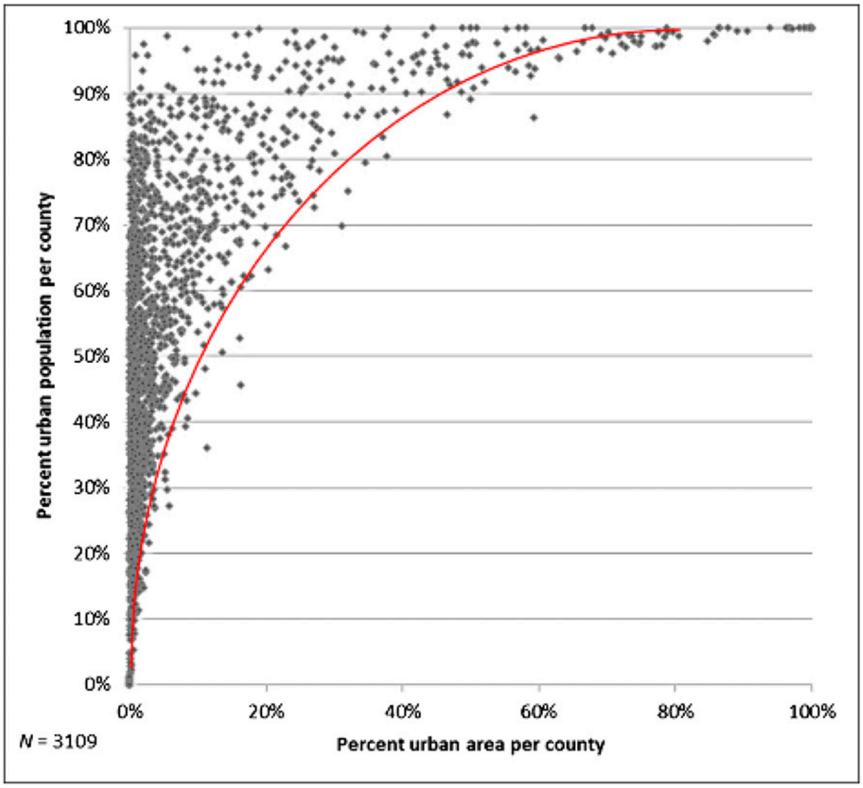


Fig. S2. Percentage of urban land area and urban population for all counties in the contiguous United States. The red line signifies the logarithmic function used to calculate infill and sprawl rates.

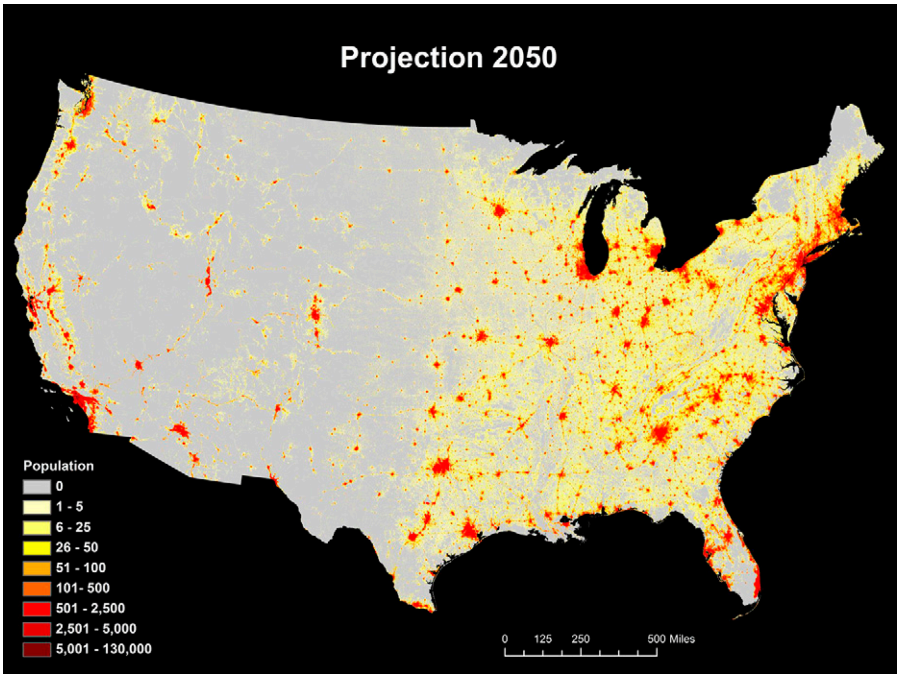


Fig. S3. Gridded (30 arc-second) projected population distribution for the contiguous United States in 2050.

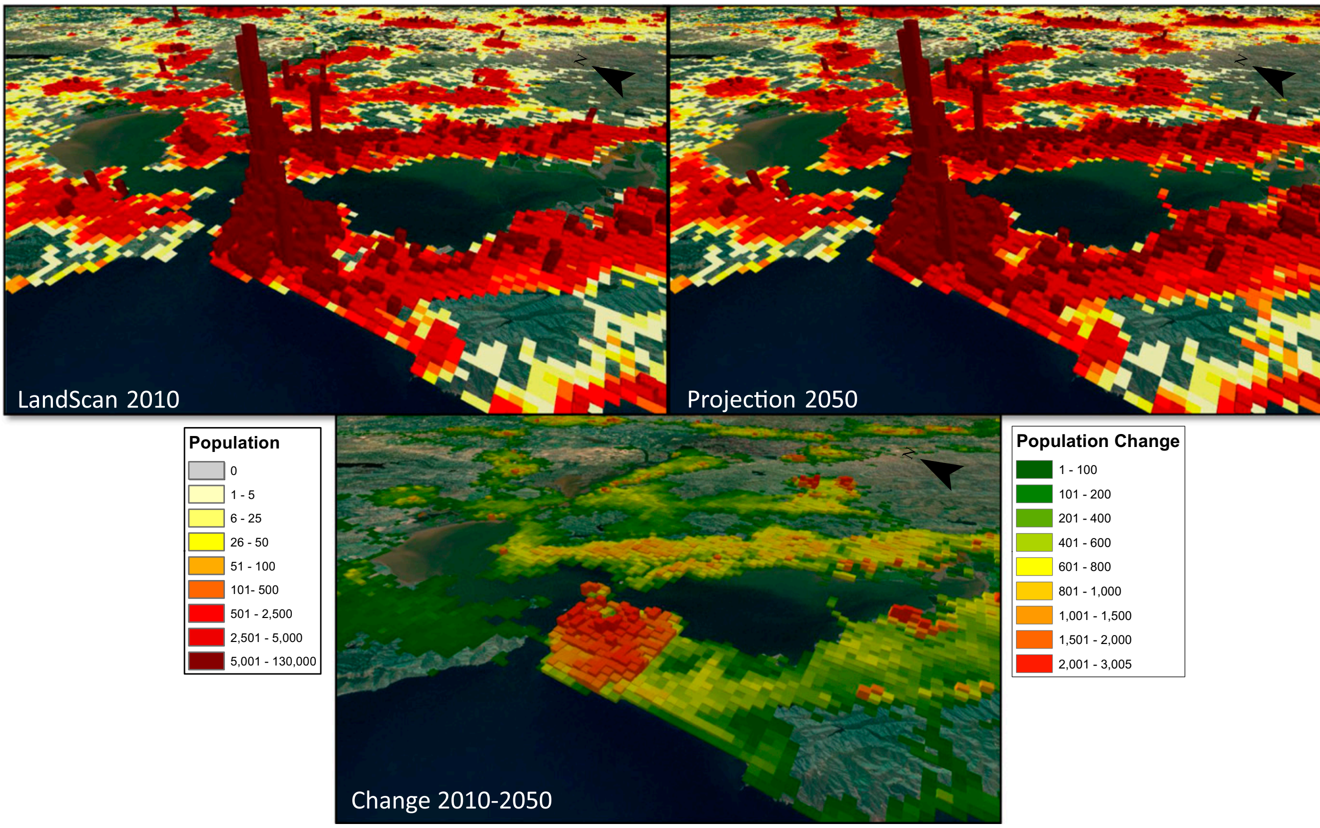


Fig. S4. Three-dimensional visualization of the San Francisco Bay area from the southwest, displaying LandScan 2010, Projection 2050, and the population change projected to occur between 2010 and 2050. The spatial resolution of each cell is 30 arc-seconds (~ 1 km). In LandScan 2010 and Projection 2050, cells with higher extrusions indicate larger quantities of population, whereas these cells indicate greater increases in population for the projected change 2010–2050. For the San Francisco Bay area, no population loss is projected to occur during this study period.

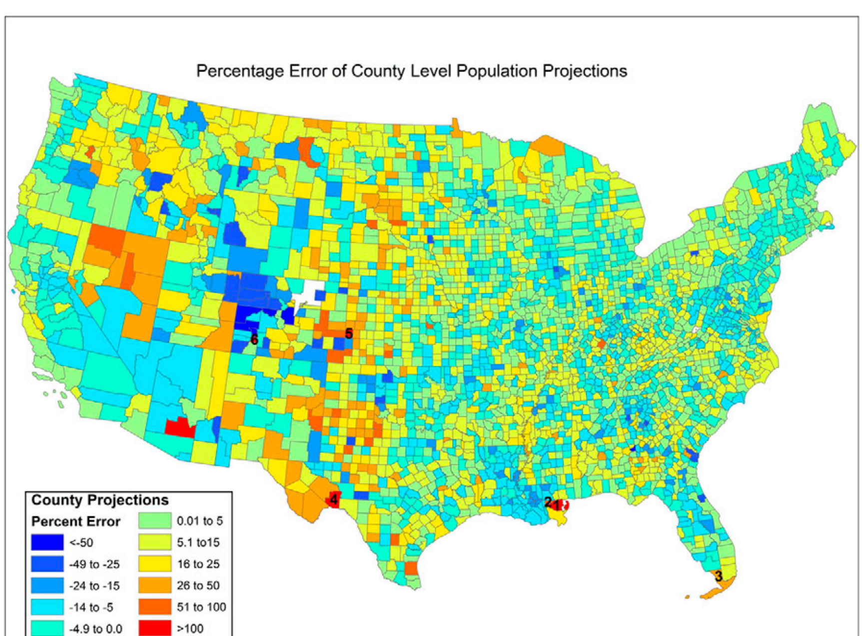


Fig. S5. Choropleth map illustrating the projection error associated with each county. Negative numbers indicate underestimation whereas positive numbers indicate overestimation. Labels represent the following counties: 1. Bernard, LA; 2. Orleans, LA; 3. Monroe, FL; 4. Terrell, TX; 5. Greely, KS; 6. San Juan, CO.

Table S1. Ordinal rank of variables from greatest to least significance

Variable	Rank
Land cover	1
Population	2
City with population ≥100 K	3
City with population ≥50 K	4
City with population ≥30 K	5
Slope	6
Highway exits	7
Roadways	8
City limits	9

Table S2. Areas excluded from future development

Exclusion type	Source
Airport boundaries	HSIP 2012
Federal defense sites	HSIP 2012
National parks	HSIP 2012
National monuments	HSIP 2012
National forests	HSIP 2012
Wildlife refuges	HSIP 2012
State/county/city parks	HSIP 2012
Golf courses	HSIP 2012
Cemeteries	HSIP 2012
Water	NLCD 2006
Perennial ice	NLCD 2006
Wetlands	NLCD 2006
High-intensity urban*	NLCD 2006

*High-intensity urban areas were also excluded under the assumption that these areas had reached maximum capacity and have therefore exhausted all potential resources for future growth.