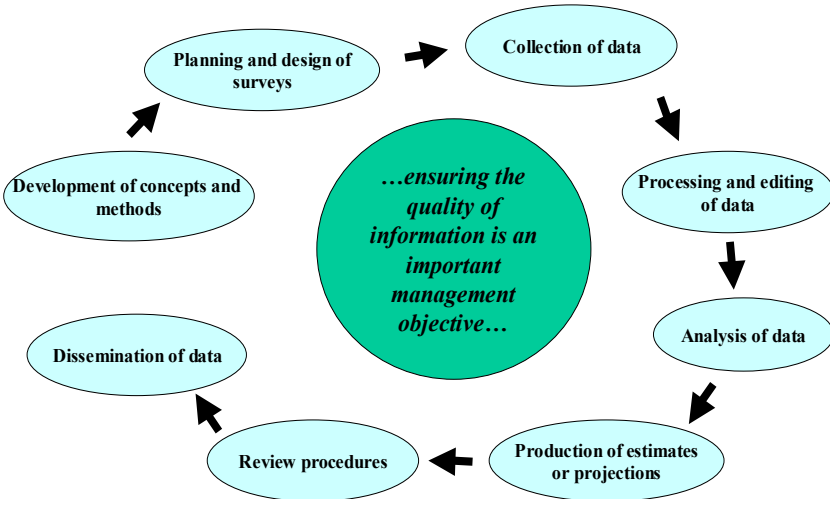
POPULATION ESTIMATION

Process and quality requirements

The Census Bureau is the largest statistical agency of the Federal Government. While best known for the decennial census of population and housing, it conducts other surveys and censuses that measure changing individual and household demographics and the economic condition of the Nation. The Census Bureau is responsible for quinquennial censuses of manufactures, retail trade, wholesale trade, service industries, finance, insurance, real estate, transportation, communication, utilities, mining, and governments. The Census Bureau also conducts approximately 200 surveys per year. It is the source of household demographic surveys sponsored by other Federal agencies, as well as by the Census Bureau. It is also the source of the country's official population estimates and projections that are used as the basis for allocating federal funds each year. Census Bureau economic surveys provide a majority of the information the Bureau of Economic Analysis (BEA) uses to update Gross Domestic Product estimates, the data used by the Bureau of Labor Statistics (BLS) in reporting Producer Price changes, and the data used by the Federal Reserve Board (FRB) as input to indices of industrial production and capacity utilization. However, most of the methods used rely on very spatially and temporally limited information, which consume great quantities of time and expense. We have investigated the utility of remote sensing imagery as a very helpful, easier, and cheaper way to provide input on change and rate of change to the population estimation DSS.

Process Life Cycle



Uses/Impacts of Census Bureau Data

- Input to the BEA to update GDP estimates
- Input to the BLS to report Producer price changes
- Used by the FRB as input to indices of industrial production
- Determine the appointment of Congressional seats
- Used to distribute hundreds of billions of $$ in federal funding

- Inform about education, income, health insurance coverage
- Used by National, state, and local governments to formulate policy
- Used by large corporations and local businesses to devise their business plans

The Office of Management and Budget (OMB), in its February 22, 2002 issuance of government-wide information quality guidelines, recognizes that Federal statistical organizations provide a substantial variety of data. Accordingly, while the Census Bureau is part of a joint Federal statistical agency notice on information quality guidelines, it presents its specific response to the OMB directive on quality, including utility, objectivity, and integrity.

The Census Bureau's Implementation of the OMB Information Quality Guidelines:
(Federal register, Vol.67, No.36, February 22,2002)

1) The Census Bureau shall ensure that information disseminated to the public shall be useful to its intended users. The requirements of utility are ongoing for a Federal statistical agency like the Census Bureau, which must be engaged in the continual development of more useful data.
2) The Census Bureau maintains ongoing contact with a broad spectrum of users to ensure that its information continues to remain relevant. Information collected by the Census Bureau is designed to provide measures that are relevant. These measures are released to the public as official statistics. Relevance is the degree to which information products provide useful information for both current needs and anticipated future needs.
3) The Census Bureau disseminates statistical information products to the public in a timely manner. Timeliness encompasses frequency of data dissemination, as well as the closeness of the release to the data's reference period. Efforts are made to collect and publish data in a time interval that allows high quality data to be disseminated to the public and also ensures that the information is usable.
4) The Census Bureau disseminates statistical information products to the public in a manner that allows them to be accessible to a broad range of data users with different requirements for data availability and understandability. Accessibility is the ease of access or effort needed for customers to acquire statistical data, products, or services. The Census Bureau conducts usability tests to ensure that its statistical products are accessible and understandable to its data users.
The Census Bureau strives for ongoing improvements to meet our customers' expectations for ease of access, quick turnaround times, simple interface mechanisms, and comparability among different data sources. We also continually enhance the quality of our products and services through greater functionality in data collection instruments as we migrate to e-commerce and computer-assisted technologies.
5) The Census Bureau shall provide information that is accurate, reliable and unbiased and shall ensure that its information products are presented in an accurate, clear, complete and unbiased manner. This objectivity is achieved by using reliable data sources and sound analytical techniques and by using highly qualified people to prepare data products that are carefully reviewed. In the area of statistical information, objectivity also requires acknowledging that errors in statistical estimates are unavoidable. These areas generally fall under the categories of "sampling" and "nonsampling" errors. Sampling errors result when estimates are based on a sample and not a complete canvass of the population of interest (as in a census).
6) The Census Bureau bases its information products on reliable, accurate data that have been validated. The Census Bureau assumes responsibility for determining sources of data (including administrative records and other data sources), measurement methods, and methods of data collection and processing for its censuses and surveys while minimizing respondent burden.

Statistical products are accompanied by descriptions of, or references to descriptions of, the methods and procedures used in their development, and other information about the data that may affect its use. The information on methodology provided or referenced permits the user to determine whether the data adequately approximate what they wish to measure, and whether the estimates they wish to use were produced with tolerances acceptable for their intended purpose.

Objectivity in analytic results is achieved by ensuring disclosure of the specific quantitative methods and assumptions that have been employed, and the disclosure of error sources affecting data quality. Statistical information products disseminated to the public by the Census Bureau must be transparent and reproducible following prescribed methodology. Reproducibility means that there is the capability to use the documented methods on the same data set to achieve a consistent result.

In summary, the quality criteria for complying with the OMB Guidelines appear to be as follows:

- Utility
- Relevance
- Timeliness
- Accessibility (Ease of access, Quick turnaround, Friendly interface)
- Objectivity (Accurate, Complete, Unbiased, Acknowledging errors, Use of reliable data sources, Informing users of data quality and methodology, and Transparency and reproducibility)
- Minimizing respondent burden

Description of the DSS – There are several DSS being used within the FSCPE to generate the estimate of population for a given geography. The two to which our effort is targeted are the State/County system as well as the Sub-County system. It should be pointed out that one of the inputs into the Sub-County system is the estimate from the State/County system. The inputs to the current State/County population estimates decision support system are listed below along with the current source for that information.

- Births and deaths (NCHS, FSCPE)
- Population (Census 2000)
- International Migration (INS, State Department)
- Internal Migration (IRS)
- Internal Migration (SSA NUMIDENT)
- Medicare (CMS)
- Group Quarters Updates (FSCPE)

Outputs from the State/County DSS go to support the development of: analytical tables, media and government briefing tables, denominators, monthly survey controls, and dissemination products of the U.S. Census Bureau and FSCPE.

The Sub-County DSS is used to produce estimates for 40,630 cities, towns, villages, and township governments. The estimates have been produced for the following historic period: 1992, 1994, 1996, 1998, 1999, and 2002. It should be noted that the distributive housing unit method replaced administrative records method in 1996. The inputs to the current Sub-County population estimates DSS are listed below along with the current source for that information.

- Housing Units and Population (Census 2000)
- Building Permits (Census)
- BAS Data (Census)
- Vacancy / PPH (Census 2000)
- Group Quarters Updates (FSCPE)
- Special Census Outputs
- County Population Estimates (State/County System)

Outputs from the Sub-County DSS go to support the development of: analytical tables, media and government briefing tables, MSA designation updates, HUD funds allocation, and various other dissemination products.

DSS product support - This effort is evaluating the addition of a remote sensing based parameter of growth that is currently not in these DSS's. Remote sensing data and information products (e.g., change detection products) can benefit the production of population estimates by addressing the following needs:

- Increase robustness of estimates
- Have more reliable empirical data
- Have additional unbiased/objective data source
- Enhance usefulness/utility
- Reduce error (MAPE) in estimates
- Minimize respondent burden
- Improve reproducibility/consistency
- Data attributes: *timeliness; accuracy; accessibility; ease of use; usability.*

Objectives

The population estimation products are in response to the requirements of the U.S. Census Bureau's Federal-State Cooperative Program for Population Estimates (FSCPE). The current chair of the research component of the FSCPE is the Missouri State Demographer. The FSCPE has wanted to evaluate new approaches to the determination of parameters for use in their population estimation tasks.

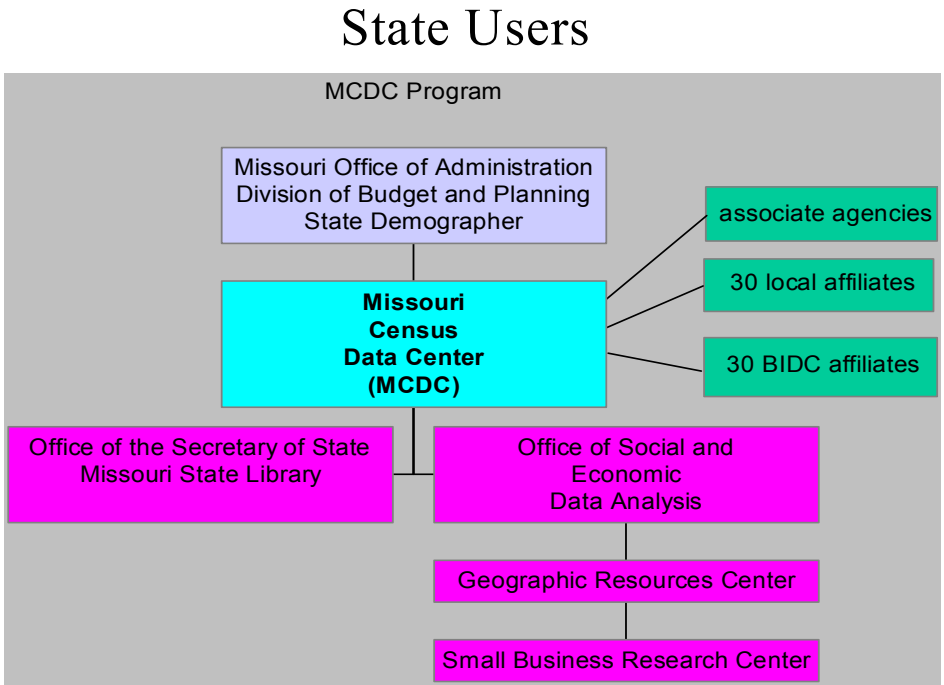
The research arena of the population estimation was to develop a methodology for local governments that would allow them to more easily update population data using remote sensing. The results of this research will benefit the decision-making for natural resources management, natural environment monitoring and social economy developing etc.

The research-based objectives of this project were:

- Create the multitemporal land-cover database and change-detection products.
- To build a model for population estimation based on residential information available in remote sensed imagery.
- To compare result from different remote sensing image and select the most suitable remote sensing data for the estimation.
- Integrate remote sensing to Census Bureau's traditional population estimation model to increase the accuracy of estimation.

User Community of Focus

Missouri State Office of Administration is the user community for this product. This product helps the planners to understand the population growth and land cover changes in Missouri State.



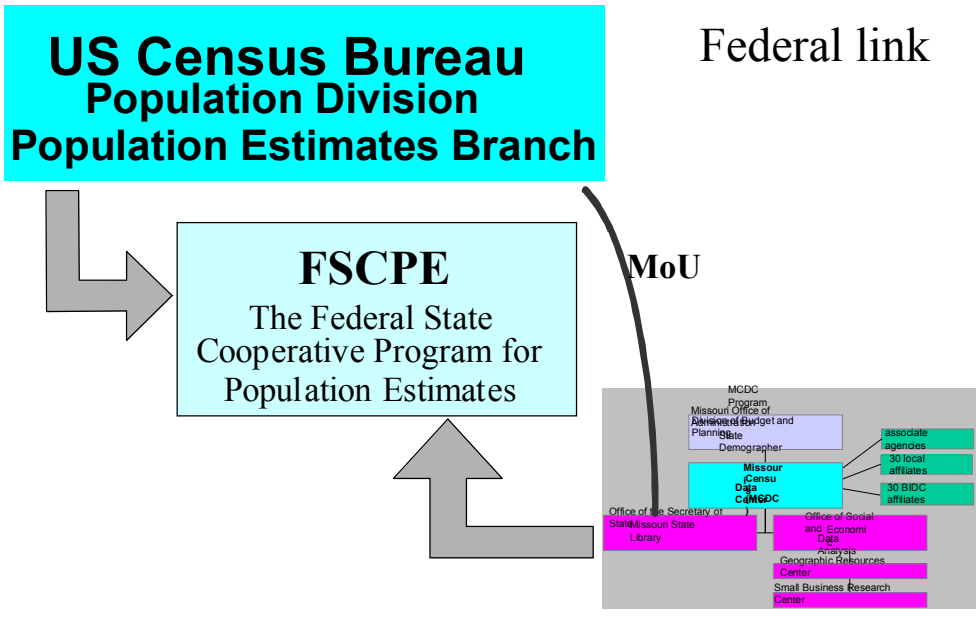
The Missouri Census Data Center (MCDC) Program is a cooperative program operating under a memorandum of understanding between the Office of the Secretary of State and the U.S. Bureau of the Census. The Missouri State Library in the Office of the Secretary of State is the agency responsible for the program. MCDC holdings consists of more than 2,000 public machine readable data files and numerous printed reports.

Coordinating Partners. The State Library works in conjunction with a coordinating group of partners comprised of the Missouri Office of Administration's Division of Budget and Planning and offices at the University of Missouri that provide census support services under a contractual arrangement with the Secretary of State's Office. From the university side, the Office of Social and Economic Data Analysis leads a group of other coordinating partners that include the Geographic Resources Center, and the Small Business Research Center to provide census support services such as Internet application development, mapping and geographic information systems (GIS) services, special studies.

Local Affiliates. In addition to the lead and coordinating agencies, the Census Bureau allows a maximum of 30 local affiliates in the MCDC Program to receive free Census Bureau products in exchange for outreach, training, and consulting services to data users in their respective geographic or service areas.

Business and Industry Development Affiliates. The Census Bureau also allows a maximum of 30 local affiliates in the state to be a part of its Business and Industry Development Center (BIDC) Program.

Local Associate Members. Rounding out the MCDC Program are many associate agencies, comprised of federal depository libraries, state agencies, not-for profit agencies, and other organizations that use and disseminate data.



State FSCPE agencies, designated by their respective governors, work in cooperation with the Population Estimates Branch (PEB) to produce subnational population estimates.

PEB begins the process of preparing population estimates by updating population information from the most recent census with information found in the annual administrative records of Federal and

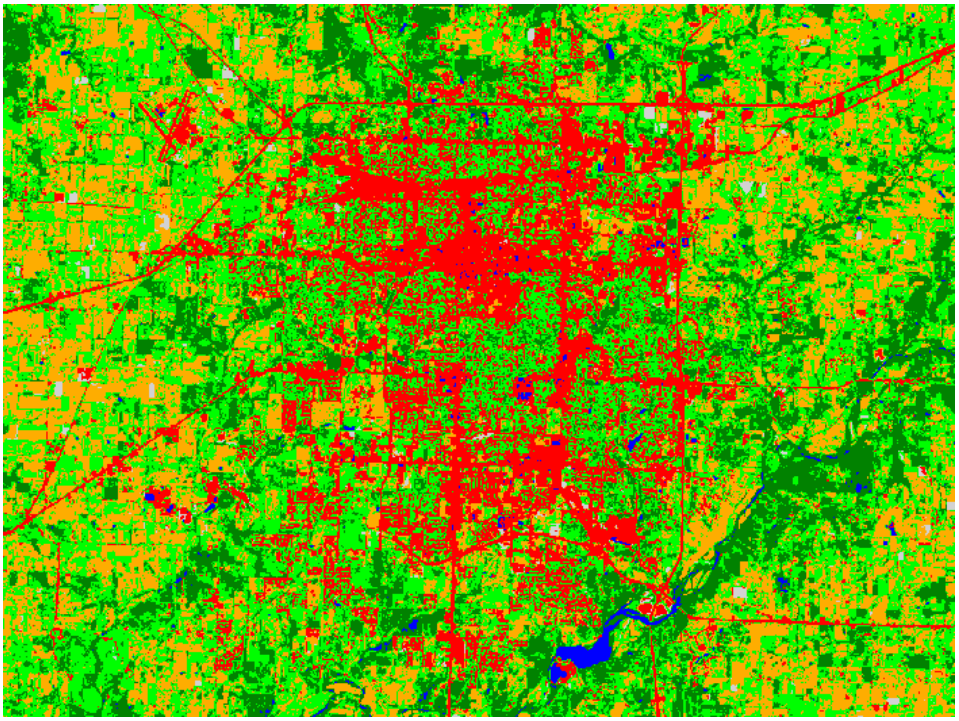
state agencies. The Federal agencies provide tax records, Medicare records and some vital statistics information. The FSCPE agencies supply state school enrollments, vital statistics, and information about group quarters like college dorms or prisons. The Census Bureau and FSCPE members use statistical models that combine the census and administrative records information to produce current population estimates consistent with the last decennial census counts. After PEB produces estimates, they are sent to the FSCPE agencies for review.

Objectives identified by the FSCPE members:

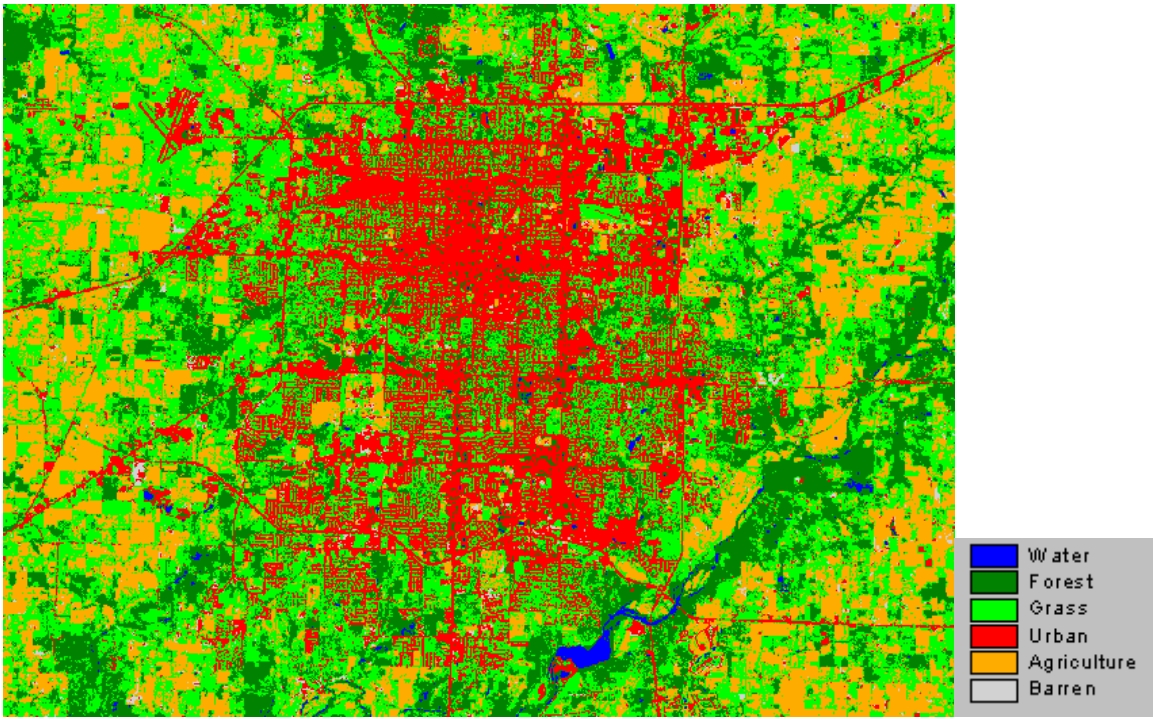
- promotion of cooperation between the states and the US Census Bureau;
- preparation of a set of consistent and jointly prepared county and subcounty estimates with complete state coverage;
- assurance of highest quality estimates through the use of established methods, comprehensive data review and thorough testing;
- reduction of duplication in the production of population estimates;
- improvement of communication among the groups compiling population figures;
- improvement and advancement of techniques and methodologies and the encouragement of joint research efforts;
- enhancement of the recognition of local demographic work.

Product Development

All of the Missouri pilot test sites (St. Louis, Kansas City, Springfield, and Columbia) were classified using dual-season TM/ETM imagery into 6 classes (water, forest, grass, urban, barren, agriculture) using an unsupervised classification method. The land-cover products were created for 1990, 2000, and 2002, and the corresponding change matrices were created for 1990-2000 & 2000-2002. The change detection analyses were completed except for some aspects of the statistical analysis of population change and its relationship to the observed landscape / land cover changes. ASTER, MODIS, Landsat ETM+, and in some cases IKONOS image datasets were acquired and assessed to determine the scalability of the change detection results for the various resolutions represented by the sensor suite. Related activity in support of this research included: 1) Conducting MODIS unsupervised & supervised classification, 2) Developing a fuzzy classification methodology to improve accuracy, 3) Comparison of several change detection algorithms (change vector analysis, post-classification comparison, simple vegetative indexes, etc.), 4) Evaluation of traditional point-based and polygon-based accuracy assessments, and 5) Conducting correlation analysis between land cover and population, land cover change and population change, and population estimation change.



TM 1992 Classification for City of Springfield

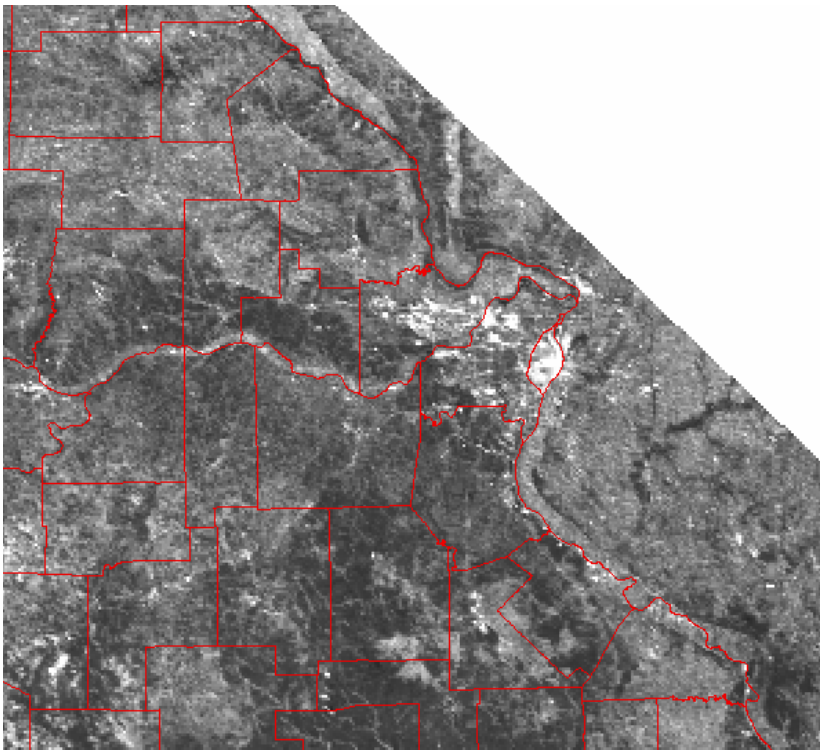


ETM 2000 Classification for City of Springfield

The classification results were transformed into vector format. Then, overlaid with city, & county boundary data. Therefore, the land-cover data is linked to population data at MSA, city & county level. A series of scatter-maps of urban area vs. population were plotted to visualize the correlation between land-cover and population at different administrative level. The plots showed a strong correlation between total urban areas vs. total population. Linear regression models were used to estimate population from TM derived urban area.

The MOD43B4 Nadir BRDF-Adjusted Reflectance (NBAR) Product (resolution is 1KM) is computed for each of the MODIS spectral bands (1-7) at the mean solar zenith angle of each 16 day period. Since the view angle effects will have been removed from the directional reflectances, this will result in a more stable and consistent product.

We selected the best available MOD43B4 data for our study areas. The urban areas in 2000 & 2002 for the 4 MSAs were detected using unsupervised classification. The preliminary results are quite good for the 1km resolution data. The population estimation was produced at MSA and county level.



2000 MOD43B4 Nadir BRDF-Adjusted Reflectance for St. Louis Area

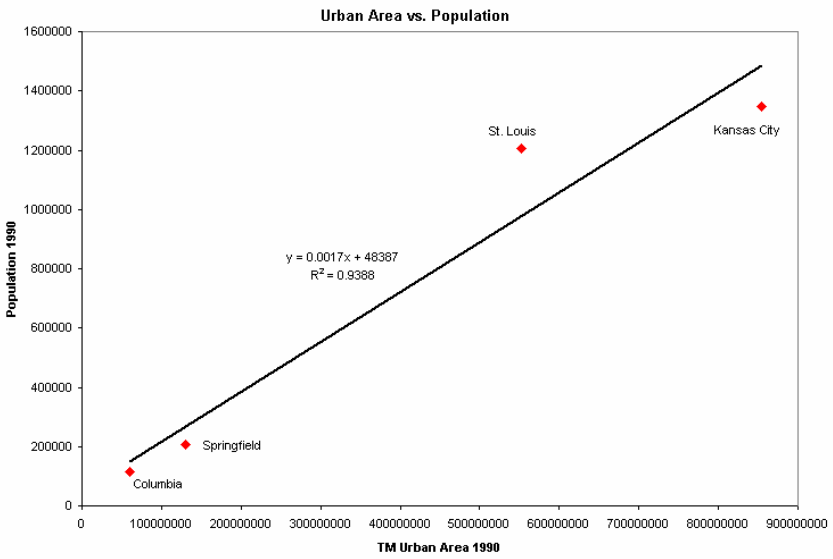
The results of the index for population estimation at these different levels using regression analysis were good. The strength of relationship is very high. The relative strength index for regression (r^2) is listed in following Table 1 (1990, 2000 population are from census, 2002 population is from population estimates produced by the U.S. Census Bureau).

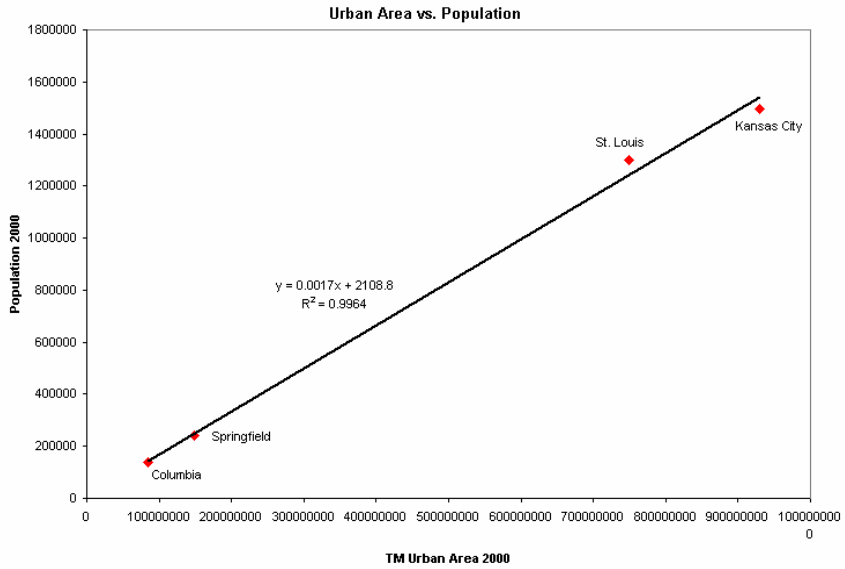
Coefficient of determination for Population Estimation from Urban Remotely Sensed land Cover

R^2	MSA	County	City
TM1990	0.9388	0.9281	0.9544
TM2000	0.9964	0.9569	0.9434
MODIS2000	0.8936	0.8134	N/A
MODIS2002	0.8921	0.8216	N/A

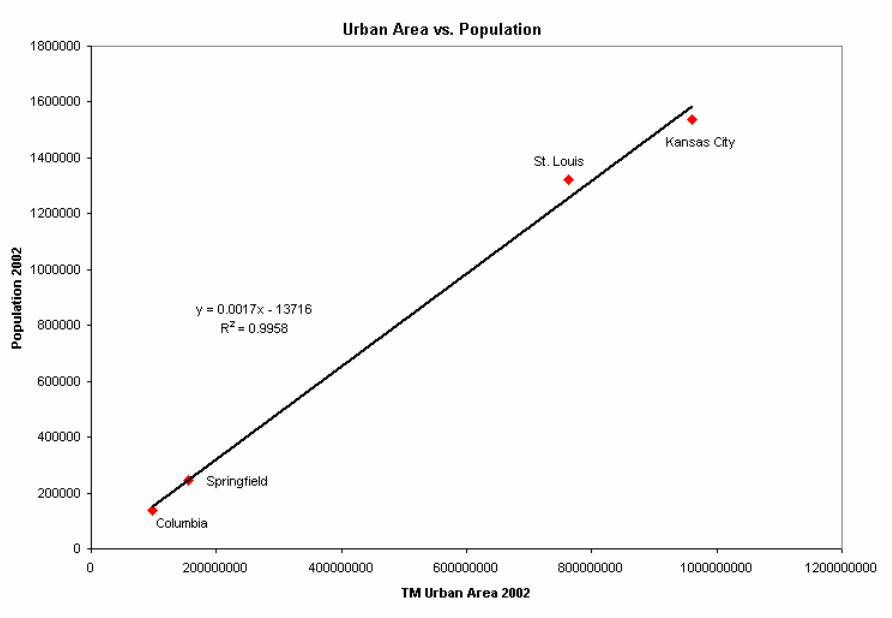
The products developed to date were created as a proof-of-concept for the Missouri test sties in response to the FSCPE research chair's request for a pilot study. The products relate to the two levels of estimation work conducted by the FSCPE ('state / county' and 'sub-county'). Specifically we are addressing a stated need for higher accuracy in estimating populations in fast growing areas. The accuracy of these estimates is evaluated by the FSCPE using two measures: Mean Absolute Percent Error (MAPE) and Mean Algebraic Percent Error (MALPE). We have adopted these as one component of our performance metrics. We have demonstrated above that the change extracted from the various sources of imagery evaluated (TM, ASTER, and MODIS) have significant (predictable) relationships (high r^2) with the observed aggregated population changes for specific MSAs in Missouri at various resolution groups. Simply stated, these geographic extent relationships to imagery source are: a) Metropolitan Statistical Areas – MODIS, TM, b) County – MODIS, TM, ASTER and c) Municipal (sub-county) – TM, ASTER.

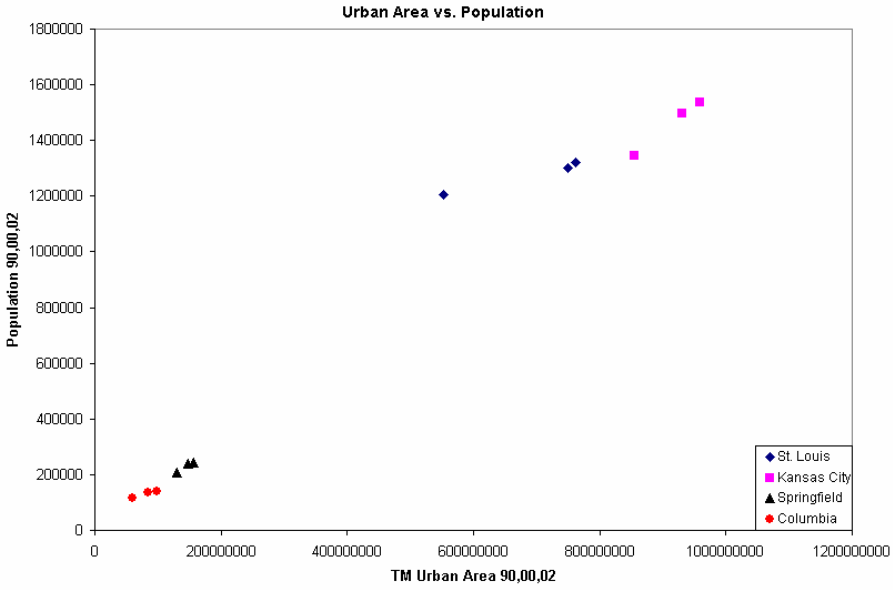
MSA level Population Estimation from TM:



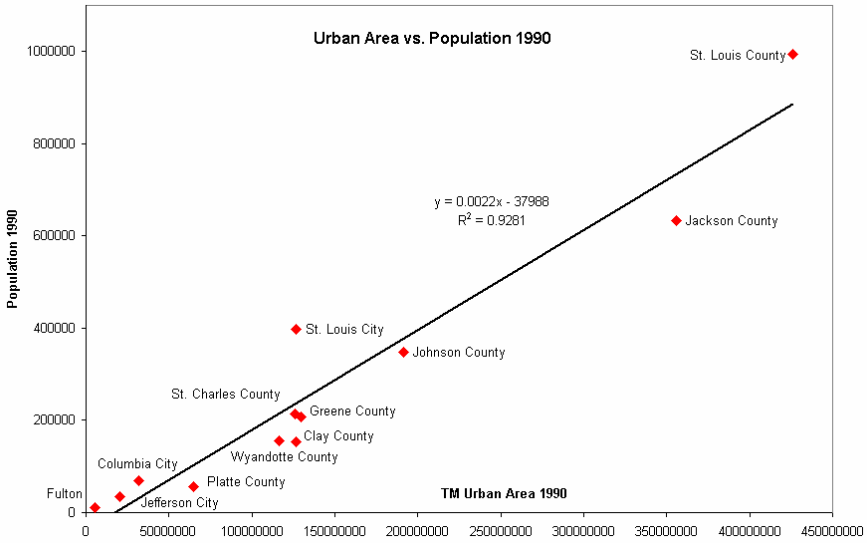


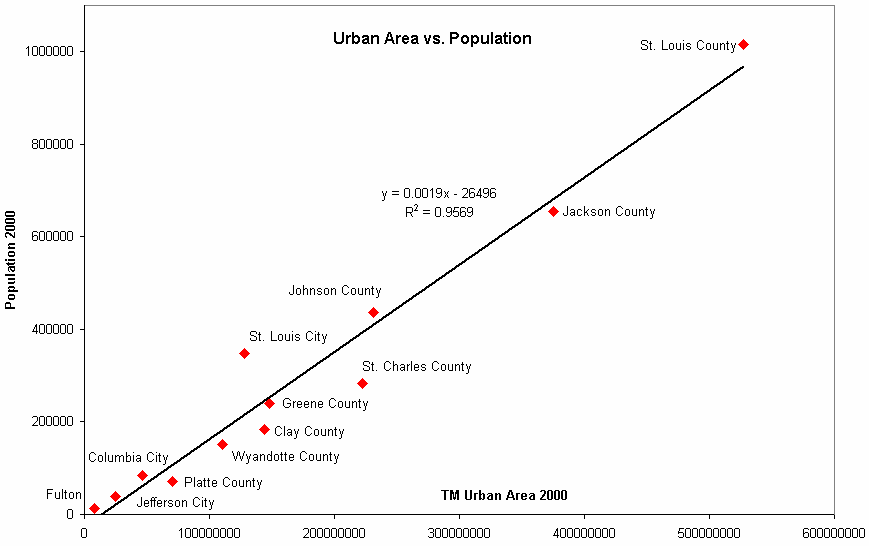
MSA level Population Estimation from TM (cont)



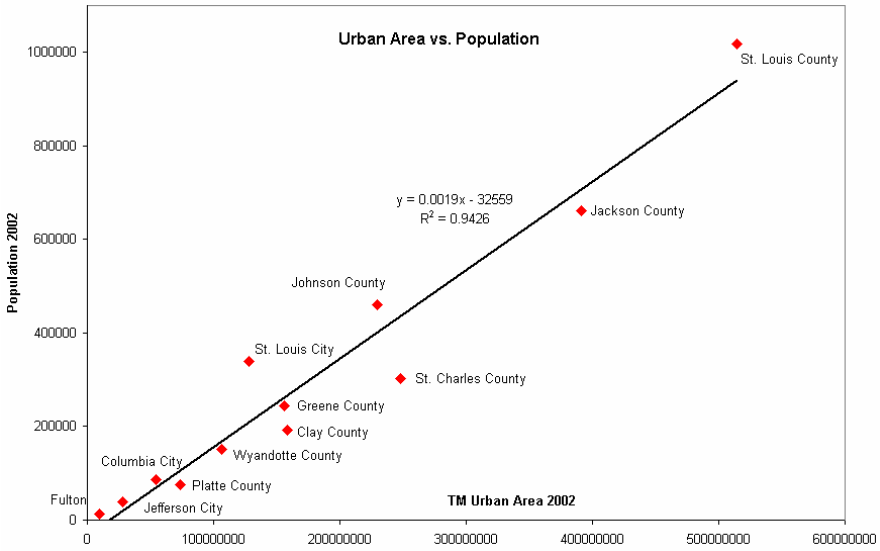


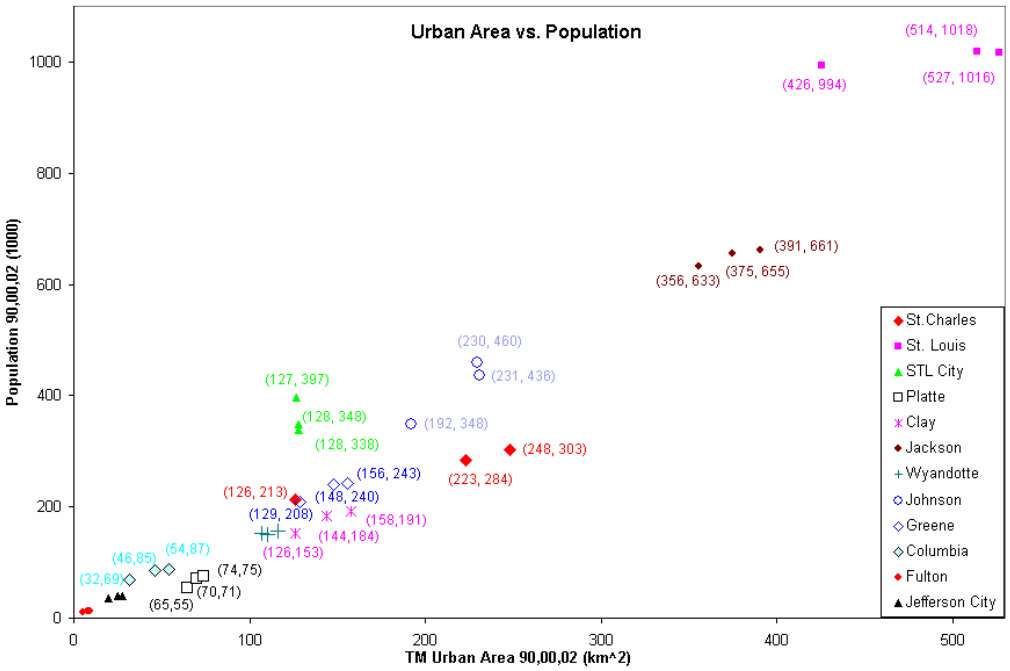
County level Population Estimation from TM:



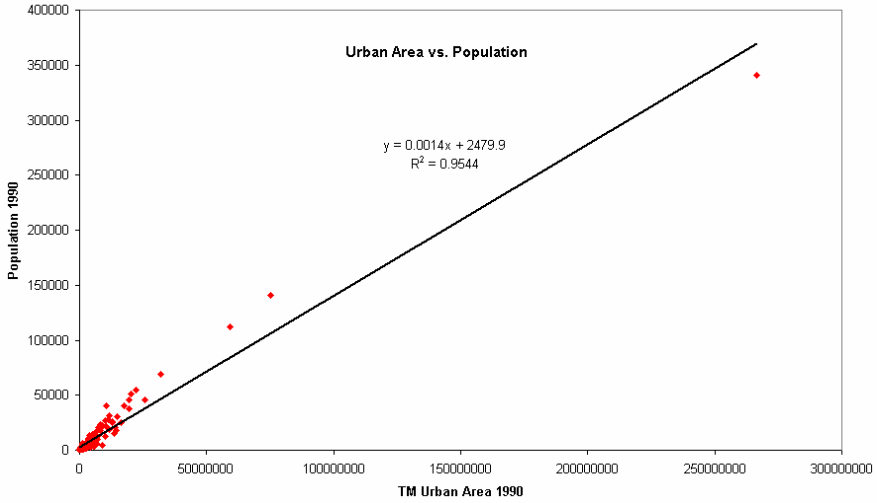


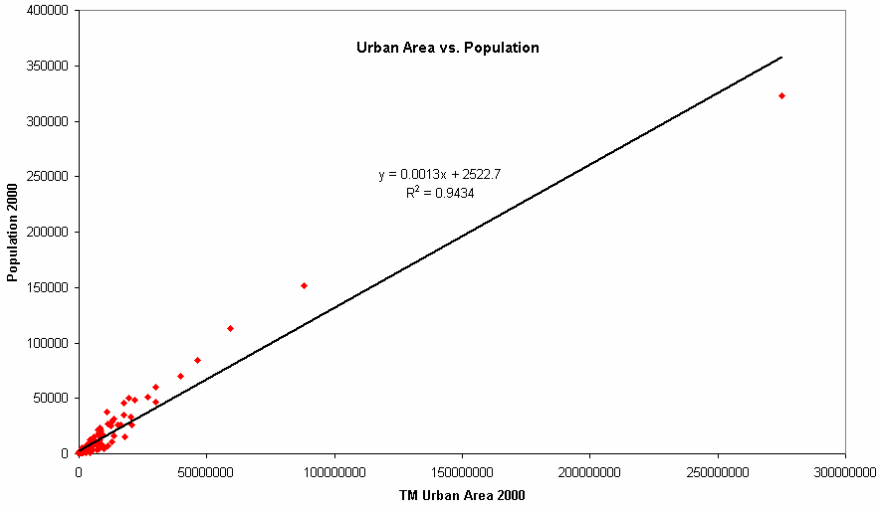
County level Population Estimation from TM: (cont)



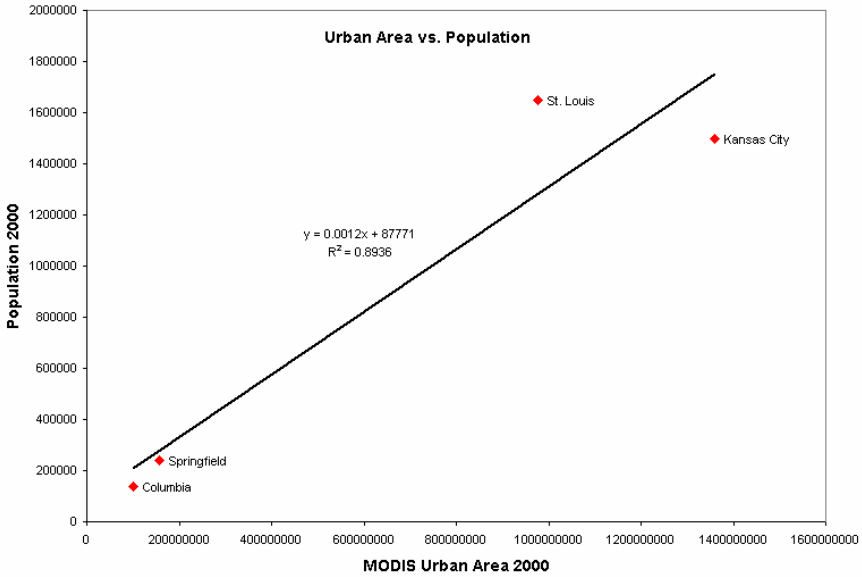


City level Population Estimation from TM:

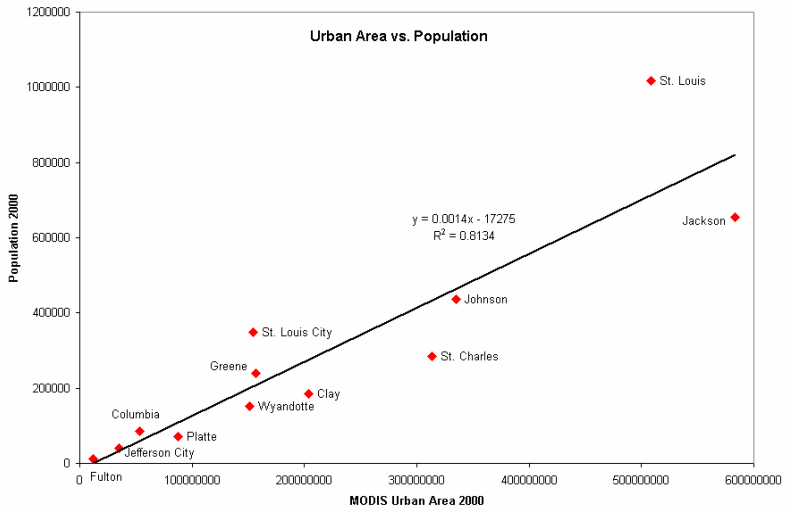




MSA level Population Estimation from MODIS:



County level Population Estimation from MODIS:



This research has successfully demonstrated the ability to improve population estimates in four Missouri high-growth urban areas by using remote sensing to set parameter(s) within the FSCPE decision system related to rate and extent of growth. This research utilized commercial high-resolution systems for validation purposes only.

User response to these products

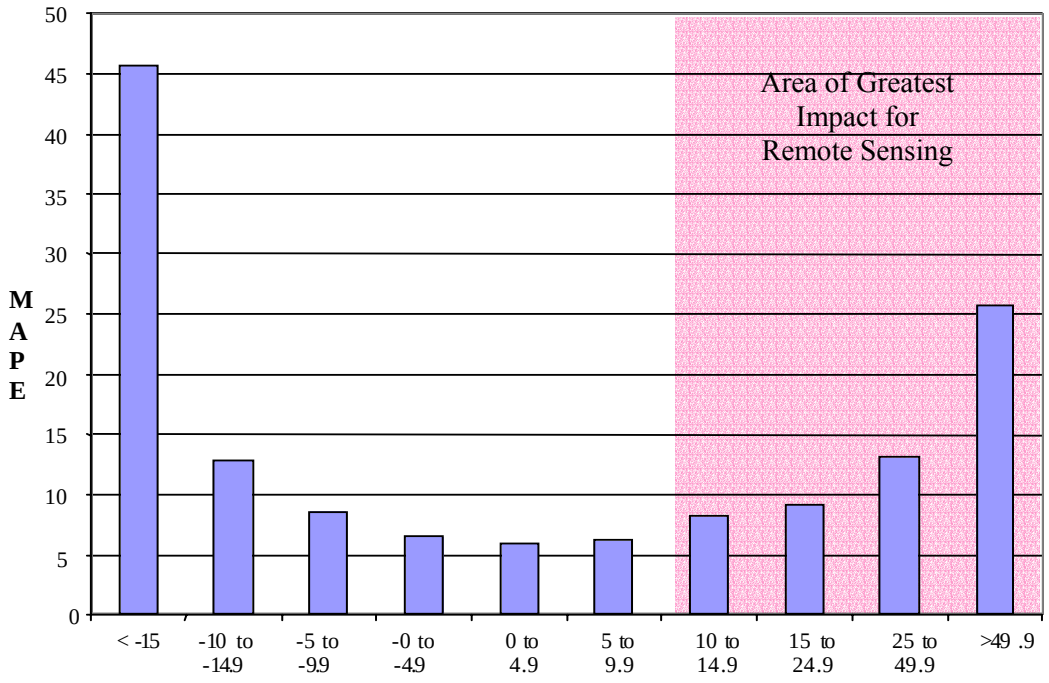
ICREST personnel conducted several meetings and discussions with the Missouri State Demographer to review the requirements of the U.S. Census Bureau's Federal-State Cooperative Program for Population Estimates regarding the image based population change and monitoring aspects as they relate to population estimates. From this and other conversations, the potential role for this application within their decision making process is very high in terms of identifying fast growing areas within and/or adjacent to urbanized areas. These areas are hard to model with their

current model inputs and provide the opportunity for remote sensing to have an impact on the policy of funding distribution equity that has direct benefit and value to society and the general citizenry.

User Requirements for FY2004 Activities:

1. *More research is needed into county characteristics and error*. This was the basic premise upon which we launched this research activity. The needed ability to monitor 'growth' or 'sprawl' for the purposes of generating a more accurate estimate is paramount to FSCPE's goal of improving the estimates generated for the nation.
2. *Decrease the error of the estimator* for those areas where growth is occurring faster than the trend data collected through other mechanisms would predict.

MAPE of current estimates versus the Census tabulations by place growth.



3. *Expand the population change modeling to national scope.* The overall MAPE for the nation was 3.3%. This however belies the fact that there exist regional differences in the accuracy as measured by the estimator. Both the South and West regions were higher (3.8 and 4.4% respectively) while the Midwest and Northeast were lower (2.5 and 2.3% respectively). The largest errors occurred in the following states: Hawaii (15.4%); Nevada (8.2%); Arizona (7.3%); Colorado (5.8%), and Florida (5.7%). The robustness of this remote sensing application needs to be tested at a variety of locations across the country. The designation of these areas will come from the FSCPE group as they represent several states and have the knowledge and data to provide input into the process.
4. *Development and implement benchmark , validation, and verification* procedures for the addition of remote sensing input to the current DSS application. It is our plan to follow the NASA DSS Guidelines for the National Application assimilation process.
5. *Assess the impact* of adding this remote sensing input to the current DSS application. Through a formal partnership with Census Bureau's FSCPE, coordinate the assessment of the efficacy of the upgraded State 2 population model to support the distribution of federal funds, policy decision-making, and human and fiscal resource management.