EXAMINING THE RELATIONSHIP BETWEEN POPULATION AND PRECIPITATION: AN IMAGERY ANALYSIS APPROACH

Wade J. Martyn, Jr., J. Ronald Eyton

Department of Geography, Texas State University-San Marcos, 601 University Drive, San Marcos, Texas 78666
wm50187@txstate.edu

KEY WORDS: population, precipitation, MODIS, NDVI, nonparametric statistics

ABSTRACT:

This research explores the association between population, precipitation, and NDVI imagery on the county level for the state of Texas. Both parametric and nonparametric statistical procedures were employed in the analysis of three different sets of variables: NDVI and precipitation, precipitation and population, and NDVI (as a surrogate measure for precipitation) and population. The strongest associations exist between rural population and precipitation measures. Additionally, the use of nonparametric statistics helped to illuminate the association between total population and precipitation measures. In sum, we were able to demonstrate the utility of using NDVI as a surrogate measure when precipitation data are not readily available.

1. INTRODUCTION

In the course of history, water has been a critical determinant factor in the success of a human settlement. Human settlement patterns have naturally congregated in areas with access to water. Indeed, water is life: a society that is unable to meet their basic water demands is not able to survive for an extended period of time.

While it is quite easy to derive precipitation data from the plethora of NOAA weather stations located around the United States, vegetative patterns can be used in place of precipitation as a surrogate measure. The logic behind this is simple: vegetation grows where water is available. If human settlement patterns congregate in areas with access to water, vegetative and human settlement patterns should also converge. A quick examination of the population distribution of the state of Texas supports this observation. Of the ten largest cities in Texas, only El Paso is located in an area of sparse vegetation. (The other nine cities, all located in moderately vegetated areas, include Houston, San Antonio, Dallas, Austin, Fort Worth, Arlington, Corpus Christi, Plano, and Garland.) Examination of a Texas climate map available through the office of the State Climatologist of Texas (2004) reveals that only El Paso is located outside of the subtropical humid/subhumid zone (figure 1). According to the Köppen-Geiger system of climate classification, El Paso belongs to a BSh climate (dry, steppe, dry and hot); the other cities belong to a Cfa climate (warm temperature, sufficient precipitation in all months, warmest month mean over 71.6 degrees F) (Strahler & Strahler, 2003).

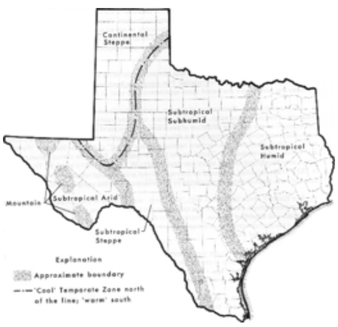

Figure 1. Climate region map of Texas

It is our assertion that it is possible to correlate the population distribution of the state of Texas with vegetative patterns inside state borders. In place of precipitation data, "greenness" values can be derived from satellite imagery using an imagery processing technique called the Normalized Difference Vegetation Index, or NDVI for short. An NDVI image is produced by applying a normalized ratio operation using red and infrared satellite image bands. According to Avery and Berlin (1992), NDVI images are effective in demarcating vegetated and non-vegetated areas. NDVI imagery-derived data values range from -1.0 to +1.0, with healthy vegetation represented by higher positive NDVI values. By using this data, "greenness" values can be used in place of precipitation data in our study of human population patterns.

Some may argue that it is not necessary to use a surrogate measure for precipitation, as NOAA makes weather station data readily available. This argument holds true when discussing areas such as the United States and Western Europe. However, in other areas of the world, there may be an advantage to using remote sensing data in place of sparsely available or otherwise unreliable in-situ data; China and North Korea immediately come to mind. Obviously, this hypothesized relationship between "greenness" values and population may not exist in more arid countries around the Earth, but it may hold true in more temperate climates. Texas, with both desert and forest within its borders, provides an ideal study area.

In this research, NDVI values are derived from imagery of the state of Texas provided by the Moderate Resolution Imaging Spectro-Radiometer (MODIS) sensor of the Earth Observing System (EOS) Terra/Aqua satellite platform. The MODIS sensor, with its 250m spatial resolution, offers a synoptic view of the state of Texas. This broad view facilitates the use of county-level population data in this research.

In Robinson and Bryson's (1957) classic article in the *Annals of the Association of American Geographers,* the authors successfully correlated rural population density and annual precipitation patterns in Nebraska. In this research, the authors were able to derive a correlation coefficient (Pearson's r) indicative of a strong relationship between population and precipitation: .80 (r^2 = .64). In other words, 64% of the total variation of rural population density in the state of Nebraska was explained by variation in annual precipitation.

More recently, Wang, Price, and Rich (2001) performed a study of spatial patterns of NDVI in Kansas. In this study, the authors were able to strongly correlate NDVI and precipitation for 60-95% of their study area. Furthermore, the authors found that "precipitation is a strong predictor of regional spatial patterns of NDVI" (2001). After examining these two pieces of research, we were left with the question: since NDVI is correlated with precipitation patterns, and rural population density is correlated with precipitation, could we show a more direct association between NDVI and population in Texas?

Most of the geographic literature exploring the intersection between remotely sensed data and human settlements has focused on estimating population values from aerial photography and Landsat imagery. The more well-known research in this vein has been undertaken by Hsu (1971), Lindgren (1971), and Lo and Welch (1977). With these exceptions, more complex examinations of the human-physical landscape using both satellite and U.S. census data have not been widely conducted. A recent notable exception to this trend was Lo and Faber's (1997) study integrating U.S. census data and Landsat Thematic Mapper (TM) data for quality of life assessment in Georgia's Athens-Clarke county. In this research, Lo and Faber distinguish between the two main dimensions of social space, the morphological and sociocultural environments:

The morphological environment is made up of a number of "subspaces" – biophysical and demographical – whereas the sociocultural environment, which is the material infrastructure of social life, is created in the human mind. Remotely sensed data can resolve a total environment into its components through an interpretation of the different facets of the same area. The result is a number of map layers, such as, vegetation, landform, houses, and so forth, each representing a subspace with its own characteristics. The common elements of these map layers will give us a better understanding of the different "social spaces"… In this paper, it is suggested that "greenness" is a desirable quality of the morphological environment. Greenness is the amount of green vegetation found in an urban environment (144).

Lo and Faber concluded that NDVI "greenness" values, derived at the census block group level for the Athens-Clarke county, were positively correlated with per capita income, median home values, and educational attainment, while negatively correlated with population density; green areas are desirable places to live. Examination of the human-environment landscape on a broader scale, such as the entire state of Texas, should reveal positive associations between "greenness" and population.

It is no great revelation to state that green areas are desirable places to live. Botkin and Beveridge (1997) provide a synthesis of the many lessons offered to urban planners through the instruction of history, with the ideas of Frederic Law Olmsted and Ebenezer Howard playing a central role. Both Olmsted and Howard were major proponents of incorporating "greenness" in urban landscape design, as can be witnessed by the establishment of Olmsted's Central Park and "garden cities" like Greenbelt, Maryland (which were influenced by Howard's ideas). These ideas have blossomed in the population landscapes of the 20th century, and it can certainly be argued that the recent mass move to suburbia is in no small way influenced by the "healthiness" seen in a green, country environment. Furthermore, even within densely settled urban areas, more expensive subdivisions are generally characterized by a strong "greenness" component.

2. DATA AND METHODOLOGY

2.1 Data

Three different data sets were used for this research. First, two different dates of MODIS imagery were acquired: October 14th, 2001, and October 23rd, 2003 (Figures 2 and 3). These dates offer two advantages: one, vegetative health is usually still quite strong in October in Texas; and two, the images are cloud-free, immediately eliminating a source of atmospheric interference. In these October images, arid west Texas is easily differentiated from an otherwise 'green' Texas.

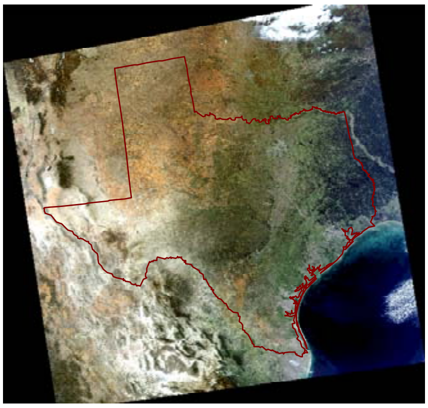

Figure 2. 2001 MODIS normal color composite of Texas

Figure 3. 2003 MODIS normal color composite of Texas

The MODIS sensor provides many different bands of imagery at varying resolution but only the red and infrared bands, with a spatial resolution of 250 meters, were needed for the production of our NDVI image. These images were downloaded through the Center for Space Research Texas Synergy website, which makes NASA's Earth Observing System (EOS) data readily available for researchers (http://synergyx.tacc.utexas.edu/).

Second, 2000 U.S. census data were obtained from the official census website (http://www.census.gov/). Total, urban, and rural population and housing values were downloaded and imported into Microsoft Excel. Using county spatial extents given in ArcGIS, county population and housing density values were calculated, and these density values were used in the statistical analysis.

Third, 30-year normal annual precipitation county totals were obtained from the *Texas Almanac*, which is derived from data provided by the Office of the State Climatologist of Texas. While Texas weather station precipitation totals are readily available from the National Climatic Data Center, the county-level totals provided by the *Texas Almanac* proved sufficient for our research design.

2.2 Image Processing

For both dates of MODIS imagery, two images were acquired and mosaiced together in ENVI 3.4 to provide complete coverage of the state. As three of the four source images were automatically projected by the Texas Synergy website into a UTM Zone 11N projection (NAD83 datum), the mosaiced images retained this projection. Extraneous areas outside of the state of Texas were then clipped out of the newly mosaiced images. Finally, the red and infrared bands for each date were output separately as raw binary data, resulting in a total of four imagery data files (Table 1).

Image	Year	Band	Wavelength (in nm)
1	2001	Red	620-670
2	2001	Infrared	841-876
3	2003	Red	620-670
4	2003	Infrared	841-876

Table 1. Imagery data files

With the exception of mosaicing and clipping the spatial extent of Texas in all the original MODIS images, image processing was performed in *Terra Firma* (Eyton, 2004). *Terra Firma* is an image processing and terrain modeling teaching software package used in undergraduate remote sensing courses at Texas State University-San Marcos.

NDVI images were produced using the infrared and red band images for the two separate dates (Figures 4 and 5). The lighter tones of the image represent more heavily vegetated areas, or areas with more 'greenness,' and darker tones correspond to less vegetated areas. A general distinction is seen between the arid (West) and subtropical portions of Texas, which corresponds to the distribution of "greenness" exhibited in the normal color composites of Figures 2 and 3. For the study area, raw NDVI values ranged from -0.65 to +0.95.

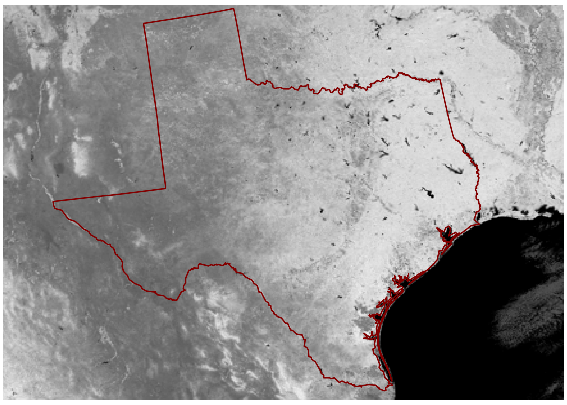

Figure 4. 2001 NDVI image

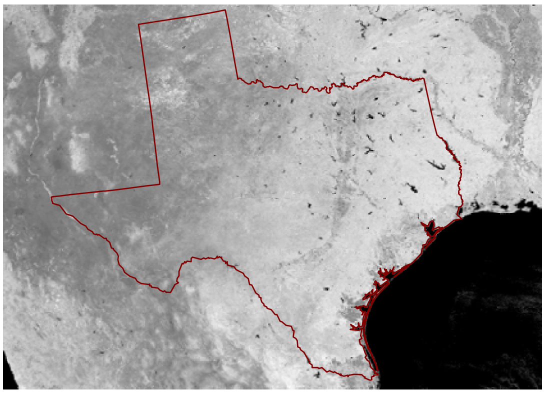

Figure 5. 2003 NDVI image

2.3 Mask Production

A county mask for the state of Texas was generated using ArcGIS and utilizing ESRI standard U.S. data (Figure 6). A shapefile containing Texas counties was projected into a UTM projection to match the image files, and the spatial extent of the county shapefile was clipped to the spatial extent of the images. This shapefile was converted to raster image format with a pixel spatial resolution of 250m, with each county given a value between 1 and 254 based on the county's alphabetical order. For example, pixels that correspond with the spatial extent of the first alphabetical county, Anderson county, were all given a value of 1. The pixels that correspond to the spatial extent of the last alphabetical county, Zavala county, were all given values of 254. Pixels associated with the remaining counties were given the appropriate numerical value. Any remaining pixels outside of the state of Texas in the raster image were given a value of 0. Figure 6 depicts the effect of giving counties a grayscale value equivalent to their alphabetical ranking, 0 (outside of Texas) toned black, 254 (Zavala county) toned near white, and the other counties grayscaled somewhere in-between the two. This raster mask was used to extract NDVI pixel values on a county basis from the MODIS images.

Figure 6. Texas county mask

2.4 Data Extraction

NDVI pixel data were tabulated using the county mask produced by ArcGIS. The mask was used to select NDVI pixel values for each particular county based on spatial extent, and then separate those values from the others; those pixels in the NDVI image that corresponded to the spatial extent of each county in the mask were treated differently. The pixels that represented Zavala county in the mask were paired up with the same pixels in the NDVI image, and the NDVI values for those pixels were set aside in a column of data that represented Zavala county. Once the NDVI pixel values were derived, county NDVI pixel means and standard deviations were calculated.

The NDVI county data values were then matched up with the appropriate precipitation and population/housing density values in Excel. The spreadsheet was then imported into SPSS for correlation and regression analysis.

3. RESULTS AND DISCUSSION

3.1 NDVI and Precipitation

Considering the relatively large sample size (n = 254) a Gaussian, or 'normal', distribution was assumed for the data, and standard parametric statistical procedures were followed. Pearson correlation coefficients (r) and coefficients of determination (r^2) were calculated for each date using NDVI and precipitation values for the state of Texas (Table 2). On the Texas county level, Pearson r values of .849 (r^2 = .721) and .691 (r^2 = .477) were calculated for 2001 and 2003. These r^2 values are indicative of a moderate to strong association between NDVI and precipitation, and helps justify the use of NDVI values as a surrogate for precipitation in the state of Texas.

	2001 NDVI Mean	2003 NDVI Mean
Precip. Totals	r = .849** (r^2 = .721**)	r = .691** (r^2 = .477**)

Table 2. Precipitation totals and NDVI means
*Note: ** denotes significance at the .01 level (2-tailed).*

3.2 Precipitation and Population

Additionally, because county precipitation totals are available, Pearson correlation coefficients (r) and coefficients of determination (r^2) were calculated for the precipitation totals and six measures of population and housing density: total population density, urban population density, rural population density, total housing density, urban housing density, and rural housing density (Table 3). This analysis shows some correlation, albeit weak, between population and precipitation values. The strongest correlation exists between rural density measures and population, with Pearson's r values of .559 (r^2 = .313) and .618 (r^2 = .381). Therefore, approximately 35% of the total variation that exists within rural population patterns is explained by the variance in precipitation. This is not true of total or urban density measures, however. This initial analysis highlights the association between precipitation and rural populations, as previously established by Robinson and Bryson (1957a), but the relationship does not still hold for urban populations or urban and rural population combined.

3.3 NDVI and Population

Correlation coefficients and coefficients of determination were calculated using the NDVI mean values as a surrogate for precipitation against the six measures of population and housing density (Tables 4 and 5). While we see the same general pattern as exhibited in Table 3, the association between NDVI means and population/housing density is not nearly as robust. In fact, only rural population/housing density measures exhibit an .01 level of statistical significance, and even that is weakly predictive in nature. These declining rates of correlation are most likely due to the non-perfect correlations between NDVI and precipitation values themselves (see Table 2). Population measures more readily correlate with precipitation totals (a direct influence) than with NDVI means (a surrogate measure), though not overwhelmingly so.

	Total Pop. Density	Urban Pop. Density	Rural Pop. Density
Precip. Totals	.198** (r^2 = .039**)	.158 (r^2 = .025)	.559** (r^2 = .313**)
(*Significance*)	.01	.05	.01
	Total Housing Density	Urban Housing Density	Rural Housing Density
Precip. Totals	.205** (r^2 = .042**)	.159 (r^2 = .025)	.618** (r^2 = .381**)
(*Significance*)	.01	.05	.01

Table 3. Precipitation totals and density measures

	Total Pop. Density	Urban Pop. Density	Rural Pop. Density
2001 NDVI Mean	.088 (r^2 = .008)	.053 (r^2 = .003)	.447** (r^2 = .200**)
(*Significance*)	.164	.397	.01
	Total Housing Density	Urban Housing Density	Rural Housing Density
2001 NDVI Mean	.091 (r^2 = .008)	.052 (r^2 = .003)	.490** (r^2 = .240**)
(*Significance*)	.149	.411	.01

Table 4. 2001 NDVI means and density measures

3.4 Nonparametric Statistical Analysis

An assumption was made at the beginning of this statistical analysis that the data fit a Gaussian, or 'normal' distribution. This assumption was made partially based on knowledge of the Central Limit Theorem, which states that as n increases, the more a sampling distribution will approach a normal distribution; the n used in this research is fairly large (254). In an effort to be completely thorough in this research, the data were tested for normality in SPSS (Table 6). Examination of both the Lilliefors and Shapiro-Wilk tests for normality shows that the data are not normally distributed, as was previously assumed.

	Total Pop Density	Urban Pop Density	Rural Pop Density
2003 NDVI Mean	.003 (r^2 = .000)	-.022 (r^2 = .000)	.307** (r^2 = .094**)
(*Significance*)	.957	.724	.01
	Total Housing Density	Urban Housing Density	Rural Housing Density
2003 NDVI Mean	.005 (r^2 = .000)	-.025 (r^2 = .001)	.344** (r^2 = .119**)
(*Significance*)	.943	.691	.01

Table 5. 2003 NDVI means and density measures

	Kolmogorov-Smirnov (Lilliefors)	Shapiro-Wilk
Precip.(Annual in Inches)	.089**	.967**
Pop. Density (Total)	.368**	.325**
Pop. Density (Urban)	.390**	.285**
Pop. Density (Rural)	.207**	.729**
Housing Density (Total)	.365**	.328**
Housing Density (Urban)	.389**	.285**
Housing Density (Rural)	.189**	.769**
NDVI 01 Mean	.104**	.949**
NDV1 03 Mean	.124**	.944**

Table 6. Tests for normality

According to Daniel (1990), nonparametric statistical procedures are useful when populations do not always meet the assumptions for parametric tests. In this case, since a Gaussian population distribution cannot be assumed, the use of nonparametric statistics is in order. There are two nonparametric tests commonly used to measure the strength of association between variables of interest: Spearman's rho and Kendall's tau. Both tests measure the degree of association between the rankings of observations rather than between the raw values of the observations.

Reexamining the association between NDVI means and precipitation totals with nonparametric tests in Table 7, the correlation remains moderate to strong. In fact, the statistics derived from the Spearman's rho test are very similar to the Pearson's r statistics in Table 2. Table 8, showing the nonparametric procedure-derived correlation between precipitation totals and population density measures, is even more revealing. In this instance, moderate to strong associations can be found between every density measure and precipitation except urban density measures, with the strongest association between rural density measures and precipitation. This same correlation relationship also holds true when comparing NDVI values and density measures using nonparametric statistical procedures (Tables 9 and 10).

		2001 NDVI Mean	2003 NDVI Mean
Precip. Totals	Kendall's tau	.694**	.521**
	Spearman's rho	.877**	.703**

Table 7. Precipitation totals and NDVI means

		Total Pop. Density	Urban Pop. Density	Rural Pop. Density
Precip. Totals	Kendall's tau	.410**	.252**	.537**
	Spearman's rho	.590**	.367**	.745**
		Total Housing Density	Urban Housing Density	Rural Housing Density
Precip. Totals	Kendall's tau	.440**	.254**	.573**
	Spearman's rho	.623**	.369**	.780**

Table 8. Precipitation totals and density measures

		Total Pop. Density	Urban Pop. Density	Rural Pop. Density
2001 NDVI Mean	Kendall's tau	.317**	.181**	.442**
	Spearman's rho	.478**	.277**	.637**
		Total Housing Density	Urban Housing Density	Rural Housing Density
2001 NDVI Mean	Kendall's tau	.334**	.181**	.466**
	Spearman's rho	.499**	.277**	.665**

Table 9. 2001 NDVI means and density measures

Three important points need to be made about the use of nonparametric statistics for this type of analysis. First, the strength of association is strongest between rural population measures and precipitation, regardless of the use of either parametric or nonparametric procedures. Secondly, nonparametric procedures most strongly reflect the association between total (urban and rural) population measures and precipitation, which may be of the most interest to researchers. Lastly, generally speaking, the use of the NDVI surrogate measure in place of precipitation will result in weaker associations with population measures, also regardless of the use of either parametric or nonparametric procedures.

		Total Pop Density	Urban Pop Density	Rural Pop Density
2003 NDVI Mean	Kendall's tau	.248**	.143**	.331**
	Spearman's rho	.372**	.220**	.478**
		Total Housing Density	Urban Housing Density	Rural Housing Density
2003 NDVI Mean	Kendall's tau	.256**	.139**	.355**
	Spearman's rho	.380**	.212**	.503**

Table 10. 2003 NDVI means and density measures

4. CONCLUSION

This research employed both parametric and nonparametric statistical procedures to examine the relationship between three different sets of variables: NDVI and precipitation, precipitation and population, and NDVI (a surrogate precipitation measure) and population. The common thread woven between all of the statistical procedures is the overarching strength of the relationship between rural population and precipitation. More than any other comparison, this relationship was always comparatively strongest. Most probably, impervious ground cover in urban areas complicated the analysis, despite the coarse resolution of the MODIS.

Also of significance is the possible utility of using nonparametric procedures for delineating the association between total population and precipitation measures. Lastly, this research showed the effectiveness of NDVI as a surrogate measure of precipitation. It is important to note, however, that population is more directly correlated to precipitation values than NDVI values.

This research has helped to further the case that water and population are intertwined in space. However, there are other avenues by which to explore the relationship between water and human settlements. The presence of groundwater or aquifers in a physical landscape could show a stronger correlation with population. Also of interest is how dammed water bodies may influence the association, particularly as humankind is increasingly able to modify the landscape to fit its own needs.

5. REFERENCES

Avery, T. & Berlin, G. 1992. *Fundamentals of Remote Sensing and Airphoto Interpretation*. Prentice Hall, Upper Saddle River, NJ.

Botkin, D. & Beveridge, C. 1997. Cities as environments. *Urban Ecosystems*, 1, pp. 3-19.

Daniel, W. 1990. *Applied Nonparametric Statistics*. Duxbury Thomas Learning, Australia.

Eyton, J. R. 2004. *Terra Firma*. Unpublished software package for digital image processing and terrain analysis. Texas State University-San Marcos, San Marcos, TX.

Larkin, T. & Bomar, G. 1983. *Climatic Atlas of Texas*. Texas Department of Water Resources, Austin, TX.

Lillesand, T., Kiefer, R., & Chipman, J. 2004. *Remote Sensing and Image Interpretation*. John Wiley & Sons, Inc., Hoboken, NJ.

Lo, C. & Faber, B. 1997. Integration of Landsat Thematic Mapper and census data for quality of life assessment. *Remote Sensing of Environment*, 62, pp. 143-157.

Office of the State Climatologist of Texas. http://www.met.tamu.edu/met/osc/osc.html (accessed 01 October 2004).

Ramos, M. (ed.) 2001. *Texas Almanac 2002-2003*. Dallas Morning News, Dallas.

Robinson, A. & Bryson, R. 1957. A method for describing quantitatively the correspondence of geographical distributions. *Annals of the Association of American Geographers*, 47 (4), pp. 379-391.

Strahler, A. & Strahler, A. 2003. *Introducing Physical Geography*. John Wiley & Sons, Inc., New York.

Texas Synergy Website. http://synergy1.csr.utexas.edu/ (accessed 01 October 2004).

United States Census Website. http://www.census.gov/ (accessed 01 October 2004).

Wang, J., Price, K., & Rich, P. 2001. Spatial patterns of NDVI in response to precipitation and temperature in the central Great Plains. *International Journal of Remote Sensing*, 22 (18), pp. 3827-3844

6. ACKNOWLEDGEMENTS

Mr. Martyn would like to thank Shirley Martyn, David Viertel, Debbie Bryan, and the reviewers of the fall 2004 class of GEO 7370 for their assistance in preparation of this article.