KTC-03-11/MSC-97-1F

Assessment of the 1997 Commodity Flow Survey for State-Level Freight Transportation Planning

KTC Report

Darren M. Scott
School of Geography & Geology
McMaster University
Hamilton, ON L8S 4K1
Canada

December 2002

Table of Contents

List of Tables

List of Figures

Preface

This project was undertaken by Dr. Darren M. Scott as a subcontract from a project funded by the Bureau of Transportation Statistics Research Grants Program and the Kentucky Transportation Cabinet (KYTC). The PI and co-PI on the latter project were Drs. Lisa Aultman-Hall (University of Connecticut) and Ted Grossardt (University of Kentucky), respectively. Dr. Scott's participation was encouraged and supported by the KYTC.

At the beginning of the subcontract in the fall of 2001, Dr. Scott was employed by the Department of Geography and Geosciences at the University of Louisville. In the spring of 2002, Dr. Scott accepted a position in the School of Geography and Geology at McMaster University, Canada, where this work was completed.

1. Introduction

Urban travel demand modeling in the United States has a rich history in both practice and academic research. The same cannot be said, however, of freight demand modeling despite the intermodal planning requirements of ISTEA and TEA21. In most instances, agencies responsible for state transportation planning (e.g. Kentucky Transportation Cabinet) have applied, in a limited way, the methodology developed for urban travel demand modeling (i.e., Urban Transportation Modeling System or UTMS1) to the freight arena (see, for example, Black 1997). Obviously, this methodological similarity implies that the fundamental data requirements for both types of demand modeling are also similar. For example, both models require estimates of the amount of traffic (i.e., number of trips for urban modeling and amount of freight for freight modeling) produced within each zone comprising the study area. Yet, despite such similarities, the two modeling systems differ significantly in terms of the *availability* of data for modeling purposes.

For UTMS, such data are collected as part of the decennial U.S. Census and made available through the Census Transportation Planning Package (CTPP). The cost to metropolitan planning organizations is therefore negligible unless such information is supplemented by other surveys (e.g. household travel surveys). By comparison, two sources of data are readily available for freight demand modeling. First, the Commodity Flow Survey (CFS) is undertaken as part of the Economic Census by the U.S. Census Bureau in cooperation with the Bureau of Transportation Statistics (BTS) of the U.S. Department of Transportation. Second, Reebie Associates produces a proprietary commercial database, TRANSEARCH®, of nation-wide, county-to-county freight movements. This database was developed originally as part of a private/public partnership with the Federal Highway Administration (FWHA).

The primary differences between the two sources of data are cost and degree of spatial resolution. Simply put, data from CFS are free and aggregated to a maximum of 106 zones spanning the United States. These zones consist of some, but not all, metropolitan statistical areas, the remainders of states and some states. TRANSEARCH data, on the other hand, are available at commercial prices and at various levels of spatial aggregation, including the county level. The tradeoff between the two sources of data for modeling purposes is obviously one of cost versus spatial resolution. For the development of statewide freight transportation models, TRANSEARCH data appear to have the advantage over CFS data simply due to the need for modeling at a zonal level that is far less aggregate than that supported by CFS. The use of CFS data for freight demand modeling at the statewide level can be facilitated only if a means exists to disaggregate the data to a finer level of spatial aggregation.

The objective of this study is to explore the possibility of disaggregating CFS data to a zonal level that is compatible with statewide freight modeling. The obvious place to start in such an endeavor is with the first stage of model system development—that is, the production and attraction of freight at the zonal level (this stage is equivalent to "trip generation" in UTMS). If it can be demonstrated that the data can be disaggregated successfully at this stage, then they can be used in further development of a modeling system. The reason for this is that the output from the first stage serves as the input to the second stage of model development. The Commonwealth of Kentucky serves as the study area for this investigation and its counties form the zonal system used in the analysis.

Two tasks are undertaken as part of this study. First, regression models are estimated using CFS data. The models are then applied to Kentucky counties to predict freight productions and attractions. Second, in an effort to ascertain how well the modeled results conform to reality, they are compared to TRANSEARCH data, which serve as surrogates for actual freight productions and attractions.

The remainder of this report consists of four parts. In the next section, the CFS and TRANSEARCH data are described briefly with emphasis being placed on differences between the two databases. These differences are nontrivial as they can impact the interpretation of the study results, which are presented in Section 4. The third section discusses briefly the development of regression models for use in freight production and attraction at the county level. Finally, conclusions and recommendations are presented in Section 5.

2. Data Description

2.1. 1997 Commodity Flow Survey

The Commodity Flow Survey (CFS) is undertaken in the United States to collect information on the movement of goods within the nation. Two such surveys were completed in the 1990s: one in 1993 and the other in 1997. Although the information requested on the survey instruments was virtually identical in both years (see Table 1), the schemes employed to classify the goods differed. In 1993, goods were classified using the Standard Transportation Commodity Classification (STCC) coding system, which was developed by the American Association of Railroads. In 1997, the Standard Classification of Transported Goods (SCTG) coding system was used. This system was created jointly by the U.S. Department of Transportation (i.e., Volpe National Transportation Systems Center) and Statistics Canada (i.e., Standards and Transportation Divisions) based on the Harmonized System (HS) of product classification, which is used worldwide. Although every effort was made by the development team to make the SCTG as compatible as possible with the STCC, comparisons of data collected using the two schemes is difficult at best without a concordance of commodity categories. This can be seen in Table 2, which details the two-digit codes and categories used in the 1993 and 1997 CFS, respectively.

As indicated in Table 1, CFS also collects information on the modes used to transport the commodities listed in Table 2. The modes included in the 1997 survey are: rail; for-hire truck; private truck; air; shallow-draft vessel; deep-draft vessel; pipeline; parcel, U.S. Postal Service or courier; other; and unknown.

2.2. TRANSEARCH® Database 2000

As mentioned previously, Reebie Associates produces a proprietary database, TRANSEARCH®, for understanding freight movement in the United States. Unlike CFS, this database is constructed from over 100 proprietary, commercial and public sources of freight data (see Aultman-Hall *et al.* [1999] for a representative list of these sources). The data are available for several zoning systems including counties, zip codes, metropolitan areas and states. Also, seven modes are included in the database: for-hire truckload, for-hire less than truckload, private truck, rail carload, rail intermodal, air and water. Commodities are reported using either the STCC or

Table 1

Comparison of data collected for each shipment using the 1993 and 1997 CFS instruments

(Source: Commodity Flow Survey 1993, 1997)

1993	1997
Total value	Total value
Total weight	Total weight
Major commodity (STCC)	Major commodity (SCTG)
All modes of transport	All modes of transport
Multiple origins	Single origin
Destination	Destination
Containerized (yes/no)	Containerized (yes/no)
Hazardous material (yes/no)	Hazardous material (UN/NA[1] codes)
Export (yes/no)	Export (yes/no)
If export, mode of export, foreign country, and city of destination	If export, mode of export, foreign country, and city of destination

[1] United Nations (UN) or North American (NA) hazardous material codes.

Table 2

Comparison of two-digit level STCC and SCTG commodity categories used in the 1993 and 1997 CFS, respectively (Source: Commodity Flow Survey 1993, 1997)

STCC		SCTG	
Code	Category	Code	Category
01	Farm products	01	Live animals and live fish
08	Forest products	02	Cereal grains
09	Fresh fish	03	Other agricultural products
10	Metallic ores	04	Animal feed and products of animal origin, n.e.c.
11	Coal	05	Meat, fish, seafood, and their preparations
13	Crude petroleum, natural gas or gasoline	06	Milled grain products and preparations, and bakery products
14	Nonmetallic ores, minerals, excluding fuels	07	Other prepared foodstuffs and fats and oils
19	Ordnance or accessories	08	Alcoholic beverages
20	Food and kindred products	09	Tobacco products
21	Tobacco produces, excluding insecticides	10	Monumental or building stone
22	Textile mill products	11	Natural sands
23	Apparel or other finished textile products or knit apparel	12	Gravel and crushed stone

24	Lumber or wood products, excluding furniture	13	Nonmetallic minerals, n.e.c.
25	Furniture or fixtures	14	Metallic ores and concentrates
26	Pulp, paper or allied products	15	Coal
27	Printed matter	17	Gasoline and aviation turbine fuel
28	Chemicals or allied products	18	Fuel oils
29	Petroleum or coal products	19	Coal and petroleum products, n.e.c.
30	Rubber or miscellaneous plastics products	20	Basic chemicals
31	Leather or leather products	21	Pharmaceutical products
32	Clay, concrete, glass or stone products	22	Fertilizers
33	Primary metal products	23	Chemical products and preparations, n.e.c.
34	Fabricated metal products	24	Plastics and rubber
35	Machinery, excluding electrical	25	Logs and other wood in the rough
36	Electrical machinery, equipment or supplies	26	Wood products

Table 2 (cont.)

Comparison of two-digit level STCC and SCTG commodity categories used in the 1993 and 1997 CFS, respectively (Source: Commodity Flow Survey 1993, 1997)

STCC		SCTG	
Code	Category	Code	Category
37	Transportation equipment	27	Pulp, newsprint, paper, and paperboard
38	Instruments, photographic goods, optical goods, watches or clocks	28	Paper or paperboard articles
39	Miscellaneous products of manufacturing	29	Printed products
40	Waste or scrap materials not identified	30	Textiles, leather, and articles of textiles or leather
41	Miscellaneous freight shipments	31	Nonmetallic mineral products
42	Containers, carriers or devices, shipping, returned empty	32	Base metal in primary or semi-finished forms and in finished basic shapes
48	Waste hazardous materials or waster hazardous substances	33	Articles of base metal
--	Commodity unknown	34	Machinery
		35	Electronic and other electrical equipment and components, and office equipment

36	Motorized and other vehicles (including parts)
37	Transportation equipment, n.e.c.
38	Precision instruments and apparatus
39	Furniture, mattresses and mattress supports, lamps, lighting fittings, and illuminated signs
40	Miscellaneous manufactured products
41	Waste and scrap
43	Mixed freight
99	Commodity unknown

Standard Industrial Classification (SIC) coding systems. For this reason alone, a commodity-level comparison with the 1997 CFS is virtually impossible. Thus, the remainder of this study focuses on *total* freight generated and attracted at the county-level in Kentucky.

3. Model Development

As argued previously, the use of CFS data for freight demand modeling at the statewide level can be facilitated only if a means exists to disaggregate the data to a finer level of spatial aggregation. Under the CFS zoning system, two zones comprise Kentucky: the Louisville MSA (i.e., Bullitt, Jefferson and Oldham counties) and the remainder of the state. Obviously, two zones are inadequate for the development of freight production and freight attraction models. However, a maximum of 106 zones can be (and were) constructed for use with the CFS data. This number *is* adequate for the construction of such models. Again, another complication arises when using traditional regression analysis—that is, if a Y-intercept is included in the model estimation, the parameters will represent the characteristics of the existing national zoning system. To overcome this problem, regression models were estimated by excluding the Y-intercept from the analysis. This approach forces the least-squares regression line through the origin, which means that if no workers or no people are available to produce or consume freight, then none will be produced or attracted for a given zone. The appeal of this approach is that the models developed can be used at any level of spatial aggregation. The county level was chosen for this project as this is the level for which County Business Pattern data were available for 1997. These data, produced by the U.S. Census Bureau each year, were used to develop independent variables for the regression models.

In total, 14 regression models were estimated. These models are found in Table 3. Four dependent variables were used in the analysis: freight production measured in thousands of tons (i.e., Models 1 – 3), freight production measured in millions of dollars (i.e., Models 4 – 6), freight attraction measured in thousands of tons (i.e., Models 7 – 10) and freight attraction measured in millions of dollars (i.e., Models 11 – 14). The independent variables were of two types—namely, the number of mid-March employees in various economic sectors (i.e., all sectors, mining, manufacturing and retail trade) and the resident population. As shown in Table 3, at the national level, the models varied in terms of their fit (i.e., r^2 values). Generally,

Table 3

National-level regression models for freight production and attraction

Model Number	Regression Equation	r^2
01	PTON = 0.0858 × EMPALL	0.648
02	PTON = 0.4580 × EMP52	0.725
03	PTON = 5.7240 × EMP10 + 0.3600 × EMP52	0.831
04	PVAL = 0.0650 × EMPALL	0.922
05	PVAL = 0.3600 × EMP20	0.950
06	PVAL = 1.2340 × EMP10 + 0.3410 × EMP20	0.964
07	ATON = 0.0894 × EMPALL	0.742
08	ATON = 0.0358 × POP	0.784
09	ATON = 0.4720 × EMP52	0.816
10	ATON = 4.2950 × EMP10 + 0.3990 × EMP52	0.879
11	AVAL = 0.0658 × EMPALL	0.949
12	AVAL = 0.0256 × POP	0.945
13	AVAL = 0.3330 × EMP52	0.957
14	AVAL = 0.5780 × EMP10 + 0.3230 × EMP52	0.959

Dependent Variables:

PTON	Freight production (000 tons)
PVAL	Freight production (millions $)
ATON	Freight attraction (000 tons)
AVAL	Freight attraction (millions $)

Independent Variables:

EMPALL	Number of mid-March employees in all economic sectors
EMP10	Number of mid-March employees in mining (i.e., SIC mining)
EMP20	Number of mid-March employees in manufacturing (i.e., SIC manufacturing)
EMP52	Number of mid-March employees in retail trade (i.e., SIC retail trade)
POP	Population

however, the models calibrated using "value" for the dependent variable outperformed those calibrated using "ton." This was the case for both production and attraction.

4. Freight Production and Attraction in Kentucky

4.1. Results for CFS Zones

The models described in the preceding section were used to predict freight production and attraction for Kentucky counties. For each model, the results were then aggregated to the original CFS zoning system consisting of the Louisville MSA and the remainder of the state to allow for comparisons with the observed CFS data, which are presented in Table 4.

Table 5 presents the results for freight production and Table 6, freight attraction. In Table 5, it is clear that Models 1 – 3 overpredict the tonnage of freight produced in the Louisville MSA and underpredict the tonnage produced in the rest of the state by as much as 74 percent as is the case for Model 1. Overall, Model 3, which is calibrated based on the number of workers employed in mining and retail trade, performs best for tonnage of freight produced (i.e., 34 percent underprediction for the state). With respect to the value of freight produced, Model 6, which is calibrated using the same independent variables as Model 3, performs best (i.e., less than 2 percent overprediction for the state). Generally, when compared to the models for tonnage of freight produced, those developed for value of freight produced perform better at the state level in terms of their estimates. This is not surprising given the fact that the *r*-square values for these models are much higher than those for tonnage (see Table 3). The final observations to be made regarding Table 5 concern the overprediction of value of freight produced in the Louisville MSA and, for Models 5 and 6, the underprediction for the remainder of the state. These observations are opposite to those described for tonnage of freight produced. Together, all of the observations made with regards to Table 5 point to the fact that the models do not capture the complexity of Kentucky's economy. It is clear from the table, that the economy of the Louisville MSA is quite distinct from the rest of the state.

Interpretation of the results presented in Table 6 for freight attraction is similar to that discussed above for freight production. Again, models calibrated based on the number of workers employed in mining and retail trade performed best overall for both tonnage (i.e., Model 10) and

Table 4

1997 CFS data for Kentucky

Geography	Production		Attraction	
	Weight[1]	Value[2]	Weight	Value
Louisville MSA	31,778	44,136	46,711	35,367
Rest of Kentucky	331,615	84,880	236,417	86,580
Kentucky	363,393	129,016	283,128	121,947

Notes:

1 Weight measured in thousands of tons.

2 Value measured in millions of dollars.

Table 5

Comparison of 1997 CFS data to model results for Kentucky CFS zones: Freight production[1]

Weight (000 tons)				Value (millions $)			
Model	Estimate	Difference[2]	Difference (%)[3]	Model	Estimate	Difference	Difference (%)
Louisville MSA				Louisville MSA			
01	35,538	-3,760	-11.83	04	26,922	17,214	39.00
02	39,101	-7,323	-23.04	05	24,209	19,927	45.15
03	33,064	-1,286	-4.05	06	23,434	20,702	46.91
Rest of Kentucky				Rest of Kentucky			
01	86,427	245,188	73.94	04	65,475	19,405	22.86
02	111,403	220,212	66.41	05	86,448	-1,568	-1.85
03	207,798	123,817	37.34	06	107,805	-22,925	-27.01
Kentucky				Kentucky			
01	121,964	241,429	66.44	04	92,397	36,619	28.38

02	150,504	212,889	58.58	05	110,657	18,359	14.23
03	240,862	122,531	33.72	06	131,239	-2,223	-1.72

Notes:

1 County-level estimates were aggregated to the original CFS zoning system for Kentucky, which consists of the Louisville MSA and the rest of the state.

2 CFS data – model estimate.

3 (CFS data – model estimate) / CFS data × 100.

Table 6

Comparison of 1997 CFS data to model results for Kentucky CFS zones: Freight attraction[1]

Weight (000 tons)				Value (millions $)			
Model	Estimate	Difference[2]	Difference (%)[3]	Model	Estimate	Difference	Difference (%)
Louisville MSA				Louisville MSA			
07	37,029	9,682	20.73	11	27,254	8,113	22.94
08	28,149	18,562	39.74	12	20,129	15,238	43.09
09	40,296	6,415	13.73	13	28,429	6,938	19.62
10	35,812	10,899	23.33	14	27,811	7,556	21.37
Rest of Kentucky				Rest of Kentucky			
07	90,053	146,364	61.91	11	66,280	20,300	23.45
08	113,360	123,057	52.05	12	81,062	5,518	6.37
09	114,808	121,609	51.44	13	80,998	5,582	6.45
10	187,268	49,149	20.79	14	90,707	-4,127	-4.77

Kentucky				Kentucky			
07	127,082	156,046	55.12	11	93,534	28,413	23.30
08	141,508	141,620	50.02	12	101,190	20,757	17.02
09	155,104	128,024	45.22	13	109,427	12,520	10.27
10	223,080	60,048	21.21	14	118,517	3,430	2.81

Notes:

1 County-level estimates were aggregated to the original CFS zoning system for Kentucky, which consists of the Louisville MSA and the rest of the state.

2 CFS data – model estimate.

3 (CFS data – model estimate) / CFS data × 100.

value (i.e., Model 14) of freight attracted. Also, the models developed for the value of freight attracted (i.e., Models 10 – 14) performed better at the state level than those developed for the tonnage of freight attracted (i.e., Models 7 – 10). Once more, this reflects the fact that the r-square values for the former models are higher than those for the latter models (see Table 3). The one departure from the results presented in Table 5 is the finding that, with only one exception, the models in Table 6 tend to underpredict the value of freight attracted to CFS zones (i.e., Louisville, MSA and the remainder of the state) and the state itself. The most likely explanation for this is that the models are not capturing the full complexity of consumers for the freight. For example, some of the freight may be bound for warehouses in the state and, thus, would not be consumed immediately by either industry or people. Moreover, such freight may not even be consumed within the state. Given the importance of logistics to Louisville's economy and to the state in general, it is highly likely that this may indeed be the case.

4.2. County Results

The spatial distributions of freight productions and attractions for Kentucky counties, as predicted by the regression models described in Section 3, were mapped using a Geographic Information System (GIS)—namely, ArcView GIS. Four figures representing each of the four dependent variables or "model groups" are shown on the following two pages. Specifically, Figures 1 and 2 present the results of Models 1 (i.e., tonnage of production) and 4 (i.e., value of production), respectively. Figures 3 and 4, on the other hand, present the results for Models 7 (i.e., tonnage of attraction) and 11 (i.e., value of attraction), respectively. The remaining figures are found in the Appendix.

As discussed in the preceding section, the models forming each model group performed differently when compared to the observed CFS data. This finding suggests that the spatial distributions of the model results may also vary from one another within each model group. To test for this possibility, paired t-tests were run comparing the models defining each model group. The results of this analysis are presented in Table 7. As expected, the models calibrated for tonnage of freight production (i.e., Models 1 – 3) produced different results at the county level. This can be seen by comparing Figures 1, 9 and 10. The same finding is also the case for the

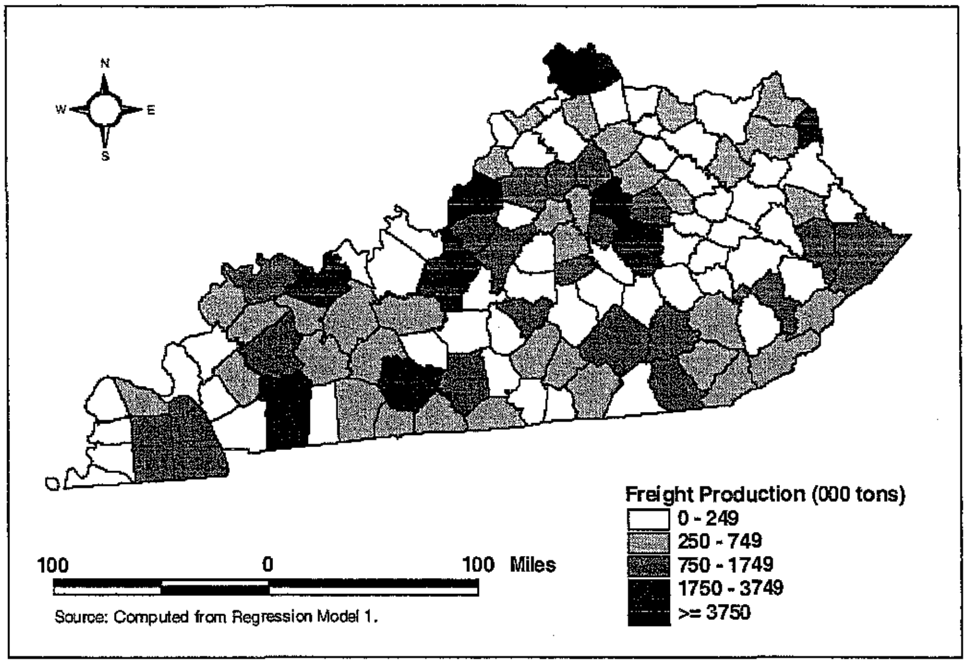

Figure 1

Results of Model 1

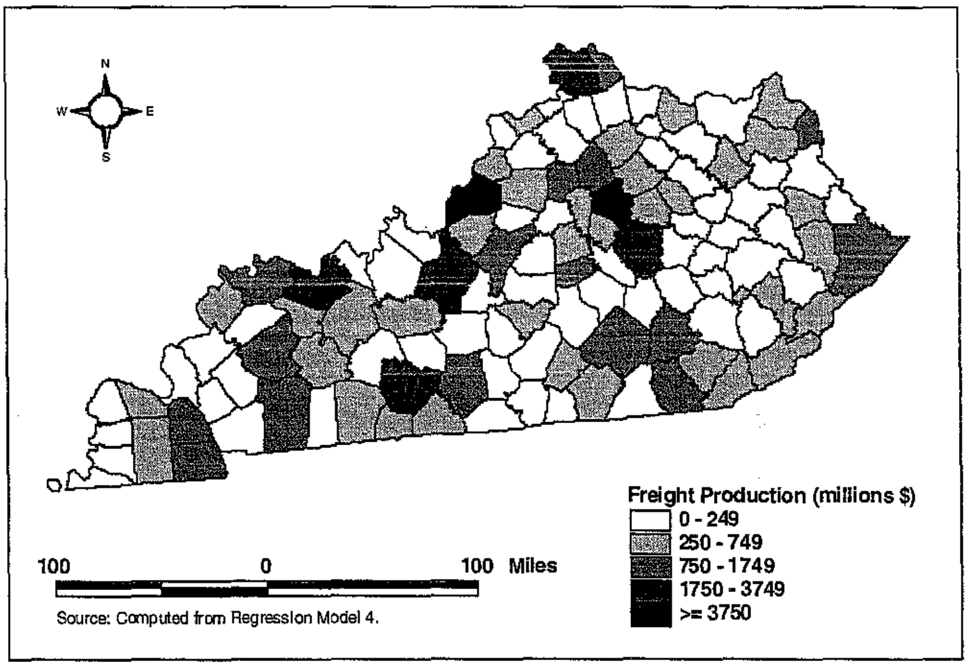

Figure 2

Results of Model 4

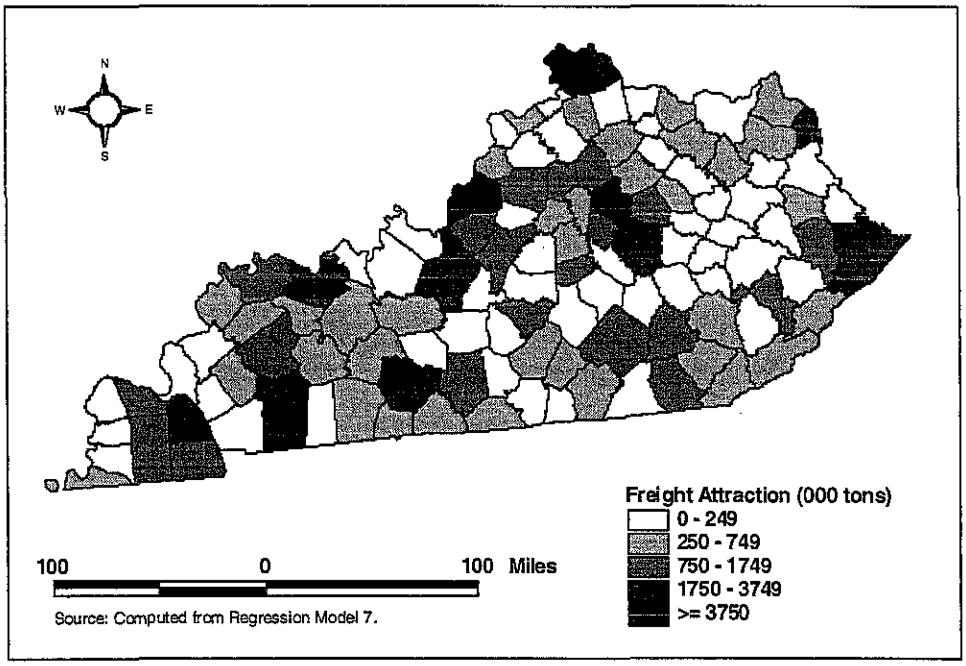

Figure 3

Results of Model 7

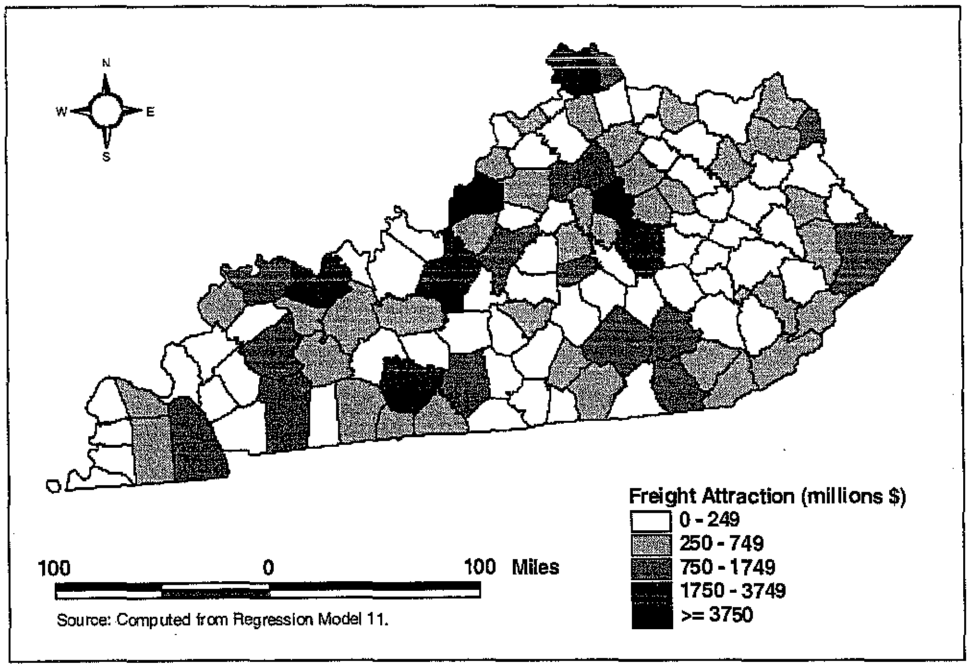

Figure 4

Results of Model 11

Table 7

Comparison of county-level model results using paired *t*-tests

	P-values
Model	0102030405060708091011121314
01	–0.0000.001———
02	0.000–0.007———
03	0.0010.007———
04	——0.0040.000———
05	——0.004–0.005———
06	——0.0000.005———
07	———0.2460.0000.000——
08	———0.246–0.4030.003——
09	———0.0000.403–0.007——

10	——0.0000.0030.007——
11	———0.4270.0000.000
12	———0.427–0.4630.117
13	———0.0000.463–0.007
14	———0.0000.1170.007–

models calibrated for value of production (see the results presented in Table 7 for Models 4, 5 and 6, and see Figures 2, 11 and 12).

Unlike freight production, the results are mixed for freight attraction. For tonnage of freight, two model pairs exhibit similar spatial distributions—namely, Models 7 and 8, and Models 8 and 9 (see Table 7). The spatial distribution of tonnage of freight attracted to Kentucky counties as produced by Model 10 is dissimilar to that of all other models forming the group (i.e., Models 7, 8 and 9). These similarities and differences are evident when comparing Figures 3 (Model 7), 13 (Model 8), 14 (Model 9) and 15 (Model 10).

The results presented in Table 7 for value of freight attracted to Kentucky counties are, for the most part, similar to those presented above for tonnage of freight. Specifically, Model 12 produces a spatial distribution that is similar statistically to that of both Models 11 and 13. Also, Model 12's distribution is similar to that of Model 14, which was not the case for tonnage of freight attracted to Kentucky counties. Once again, these similarities and differences among models can be seen by comparing Figures 4 (Model 11), 16 (Model 12), 17 (Model 13) and 18 (Model 14).

Although the comparison of models within model groups is useful in revealing similarities and differences in the spatial distributions of freight productions and attractions, the comparison does not suggest how well the modeled distributions conform to those observed in reality. In regression analysis, this can, however, be ascertained by the *r*-square value and by residual analysis. To undertake a residual analysis, observed data are required for the observations under scrutiny, which, in this investigation, are freight productions and attractions for Kentucky counties. The CFS data preclude such an analysis. For this reason, TRANSEARCH data are used. However, as discussed in Section 2, there are nontrivial differences between CFS and TRANSEARCH data. To further complicate the residual analysis, the TRANSEARCH data used are for 2000, not 1997—the year for which the CFS data were collected. Also, the analysis is restricted to tonnage of freight produced and attracted at the county level because this is how freight flows are measured in the TRANSEARCH database. As discussed in Sections 3 and 4.1, the models calibrated for value of production and attraction appear to outperform their tonnage counterparts. With all of this said, the residual analysis presented in this report can, at best, provide an approximation of how well the modeled distributions conform to those in reality, but not a definitive account.

Table 8, which compares the 1997 CFS data and the 2000 TRANSEARCH data for Kentucky, further suggests that caution must be exercised in the residual analysis. For the state as a whole, CFS freight production exceeds TRANSEARCH by 18 percent. This amount is even greater for freight attraction (i.e., 37 percent). It is unclear whether these discrepancies are attributed to structural differences in the data collection methodologies including commodity coverage, to temporal variations in freight production and attraction or to some combination of both.

Freight productions and attractions for Kentucky counties, as found in the TRANSEARCH database, are shown in Figures 5 and 6, respectively. The data underlying these figures were used to generate residuals for Models 1 – 3 (i.e., tonnage of production) and Models 7 – 10 (i.e., tonnage of attraction). The residuals were then mapped using ArcView GIS. Figures 7 and 8 present the results for Models 1 and 7, respectively. The residuals for the other models are found in the Appendix (i.e., Figures 19 – 23).

Figures 7, 8 and 19 – 23 provide a visual and, therefore, qualitative assessment of how well the modeled distributions conform to those obtained from the TRANSEARCH database (i.e., Figures 5 and 6). To obtain a more quantitative assessment, paired *t*-tests were run comparing the model results with the TRANSEARCH distributions. The results of this analysis are found in Table 9. For freight production, the spatial distributions for all models are significantly different from that obtained from TRANSEARCH whereas, for freight attraction, three of the four distributions are similar.

5. Conclusions and Recommendations

The objective of this study was to explore the possibility of disaggregating CFS data to a zonal level that is compatible with statewide freight modeling. To this end, 14 regression models were calibrated using 106 CFS zones. These models were then used to estimate freight productions and freight attractions by both tonnage and value for Kentucky counties. These estimates were

Table 8

Comparison of 1997 CFS and 2000 TRANSEARCH data for Kentucky

Geography	CFS Data		TRANSEARCH Data		Difference[3]			
	Prod[1]	Attr[2]	Prod	Attr	Prod	Prod (%)[4]	Attr	Attr (%)
Louisville MSA	31,778	46,711	33,075	58,231	1,297	3.92	11,520	19.78
Rest of Kentucky	331,615	236,417	274,443	147,914	-57,172	-20.83	-88,503	-59.83
Kentucky	363,393	283,128	307,518	206,144	-55,875	-18.17	-76,984	-37.34

Notes:

1 Freight production measured in thousands of tons.

2 Freight attraction measured in thousands of tons.

3 TRANSEARCH data – CFS data.

4 (TRANSEARCH data – CFS data) / TRANSEARCH data × 100.

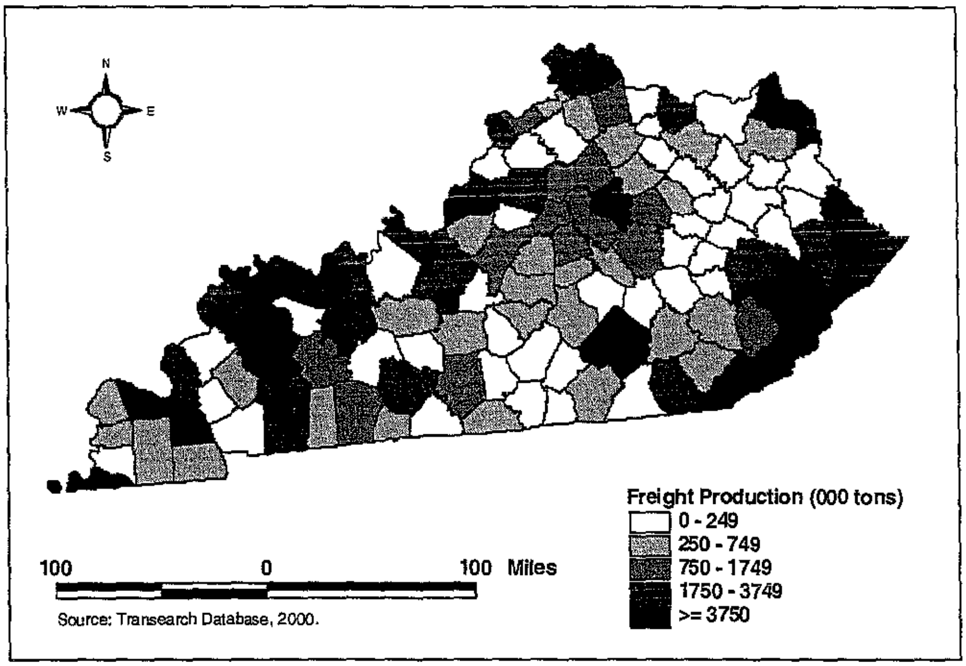

Figure 5

TRANSEARCH data freight production for Kentucky counties

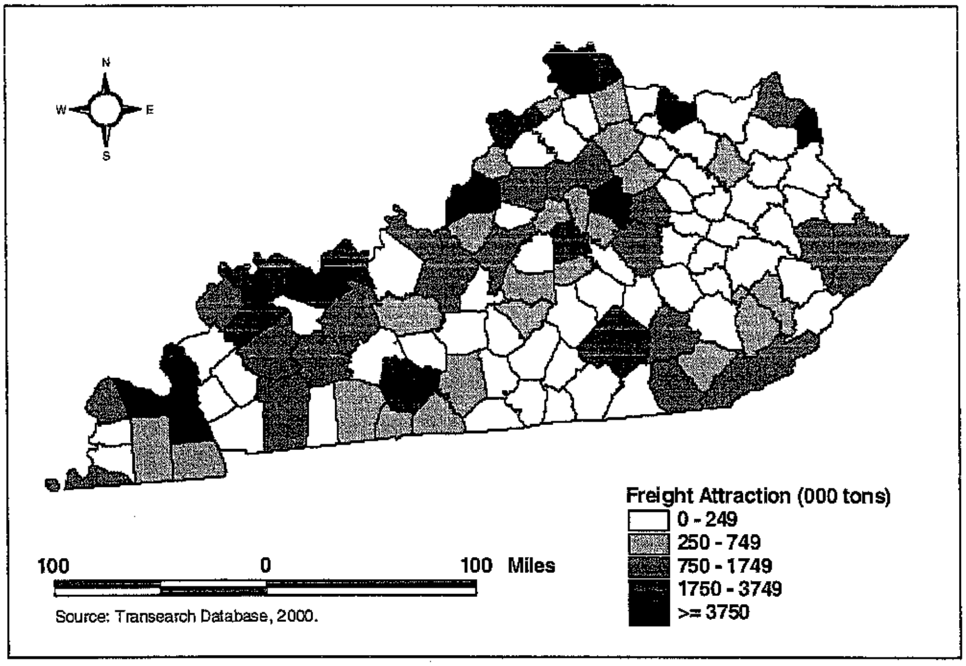

Figure 6

TRANSEARCH data freight attraction for Kentucky counties

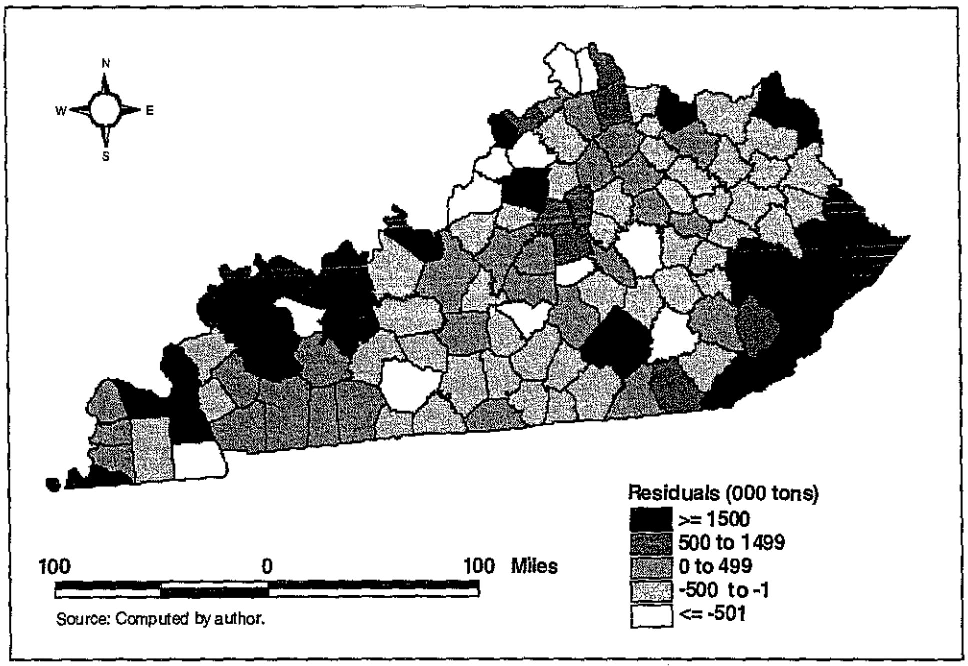

Figure 7

Comparison of TRANSEARCH data to Model 1 results (residuals are obtained by subtracting Model 1 results from TRANSEARCH data)

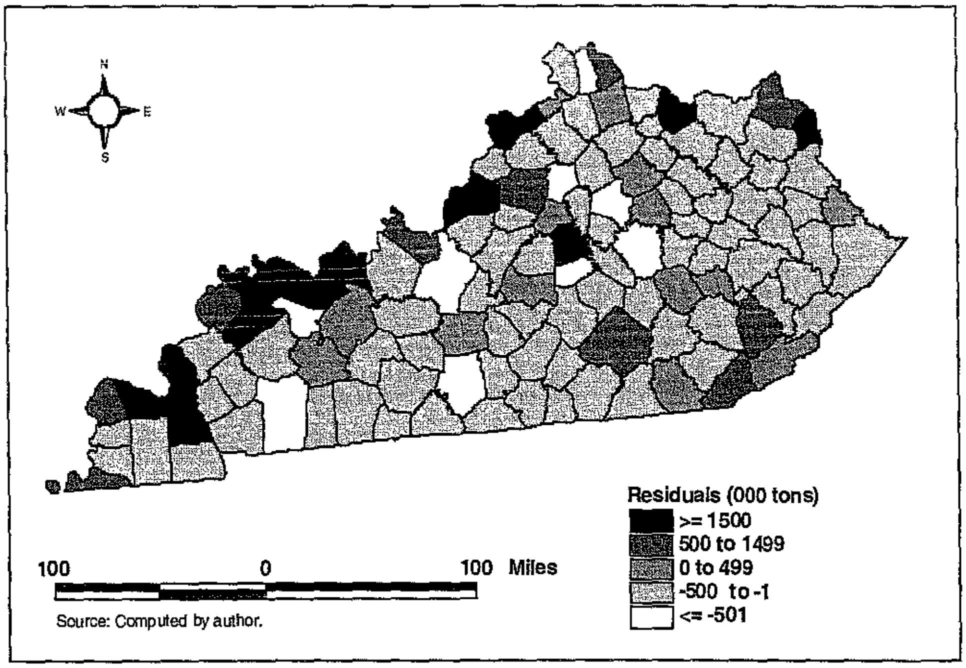

Figure 8

Comparison of TRANSEARCH data to Model 7 results (residuals are obtained by subtracting Model 7 results from TRANSEARCH data)

Table 9

Comparison of county-level model results to TRANSEARCH data using paired *t*-tests

Model	P-values	
	TRANSEARCH Production	TRANSEARCH Attraction
01	0.000	–
02	0.002	–
03	0.041	–
07	–	0.020
08	–	0.121
09	–	0.122
10	–	0.696

then aggregated to the two CFS zones comprising Kentucky and compared to the CFS data. Following this, the predicted values were mapped and statistical comparisons were made among the spatial distributions for models within each of four model groups. Finally, to ascertain model fit at the county level, a residual analysis utilizing TRANSEARCH data was undertaken for freight tonnage (i.e., Models 1 – 3 and Models 7 – 10). Again, statistical comparisons were made. Generally, the results of this study suggest that CFS data can be disaggregated to the county level for freight modeling. However, the relative success of any such endeavor depends on both the dependent and independent variables used in calibrating the regression model. For example, of the 14 models calibrated as part of this study, the best performing model of each model group was the one containing independent variables for specific economic sectors. Also, better results were found when production and attraction were measured in terms of "value" rather than "tonnage."

While this study demonstrates that it is possible to disaggregate CFS data to the county level for use in state-level transportation planning models, further refinement of the methodology is necessary. The following recommendations are made in this regard. First, regression models should be developed and evaluated for commodity groups. It is hypothesized that such models can better capture the underlying structure of Kentucky's economy than those calibrated for "total" productions and attractions. While this is possible for freight production, it is not possible for freight attraction using the 1997 CFS data. The reason for this is that a destination table by commodity group is not provided. Thus, a second recommendation from this study is that such data be provided in the future.

Acknowledgements

The author wishes to thank Feng Guo for running the regression models presented in Table 3 of this report. The author gratefully acknowledges the financial support of the Kentucky Transportation Cabinet in undertaking this project. Finally, the author wishes to thank Dr. Lisa Aultman-Hall for her constructive comments on a previous version of this report.

References

Aultman-Hall, L., Johnson, B. and Aldridge, B. (1999) *Freight Commodity Flow in Kentucky*, KTC Report 99-65. Lexington, KY: Kentucky Transportation Center, University of Kentucky.

Black, W.R. (1997) *Transport Flow in the State of Indiana: Commodity Database Development and Traffic Assignment: Phase 2*. Bloomington, IN: Transportation Research Center, Indiana University.

Appendix

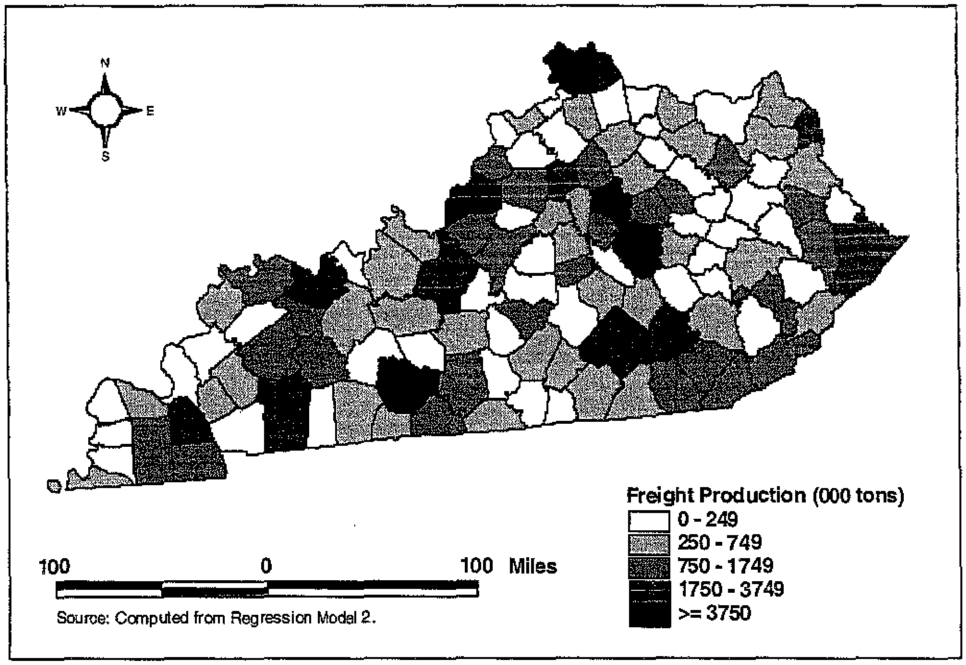

Figure 9

Results of Model 2

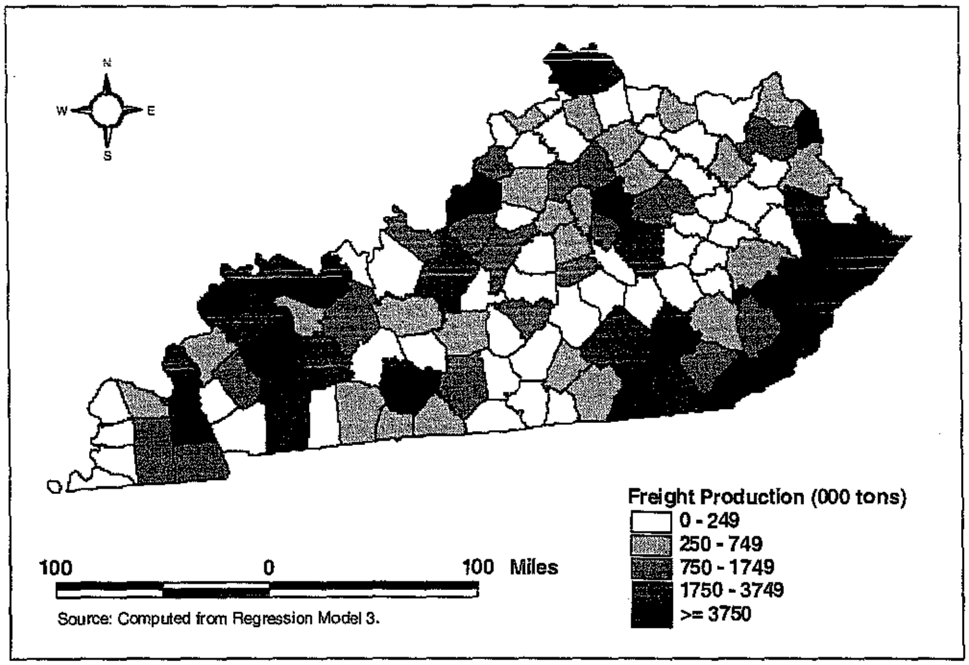

Figure 10

Results of Model 3

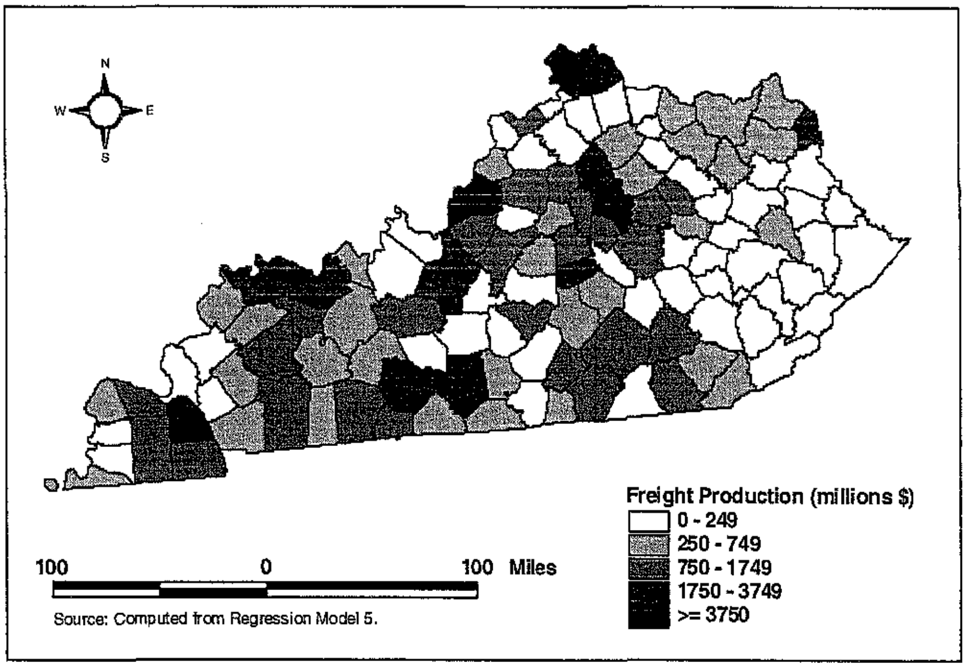

Figure 11

Results of Model 5

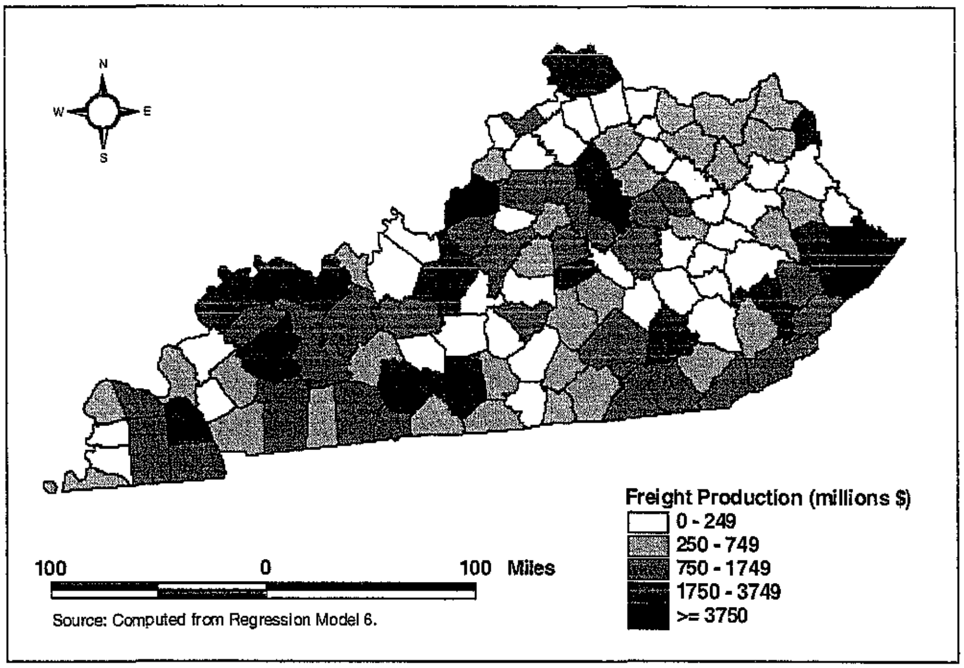

Figure 12

Results of Model 6

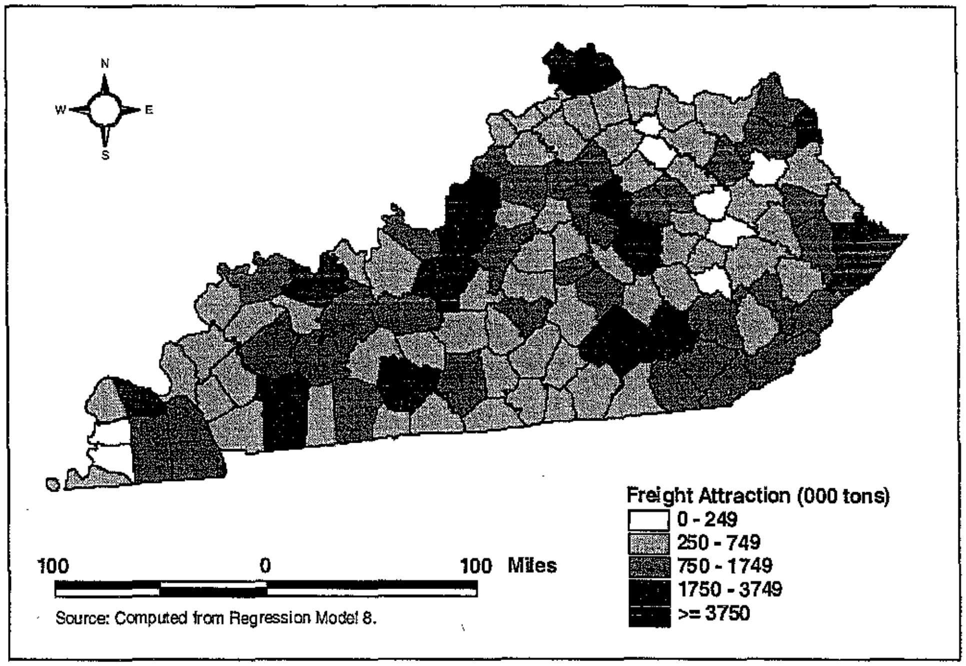

Figure 13

Results of Model 8

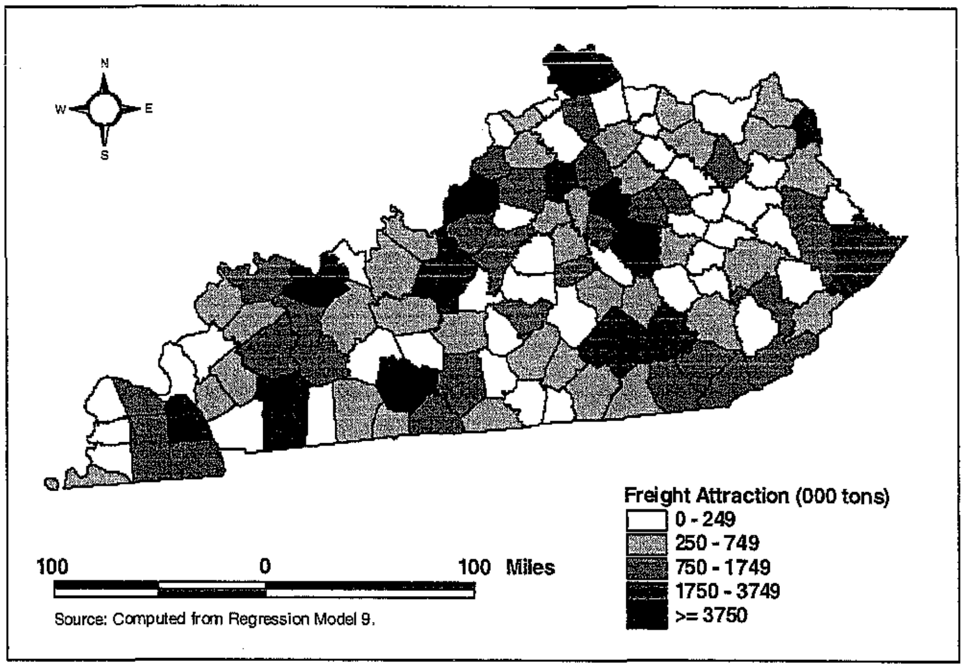

Figure 14

Results of Model 9

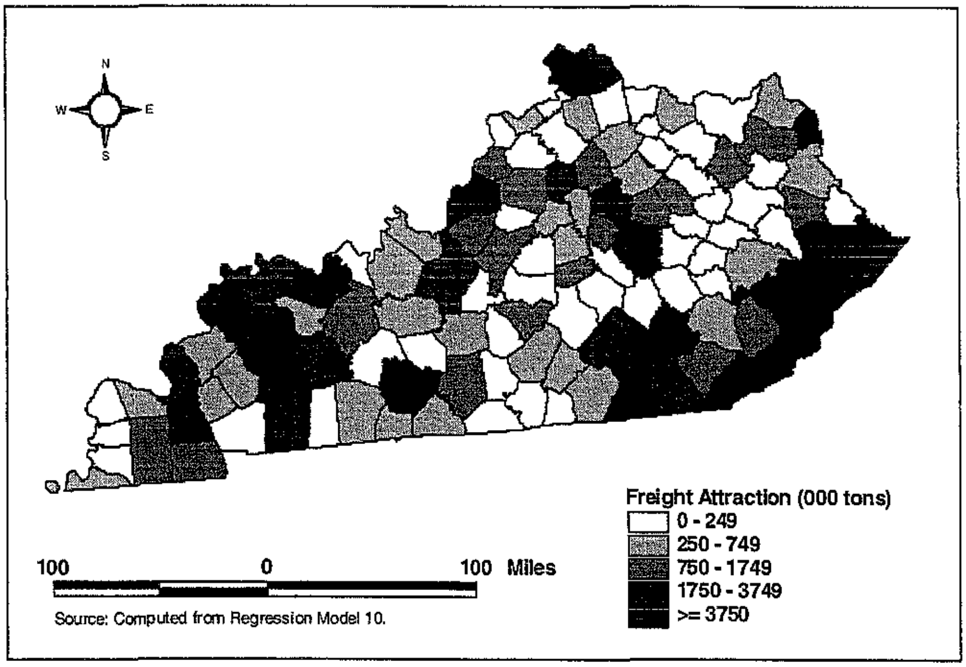

Figure 15

Results of Model 10

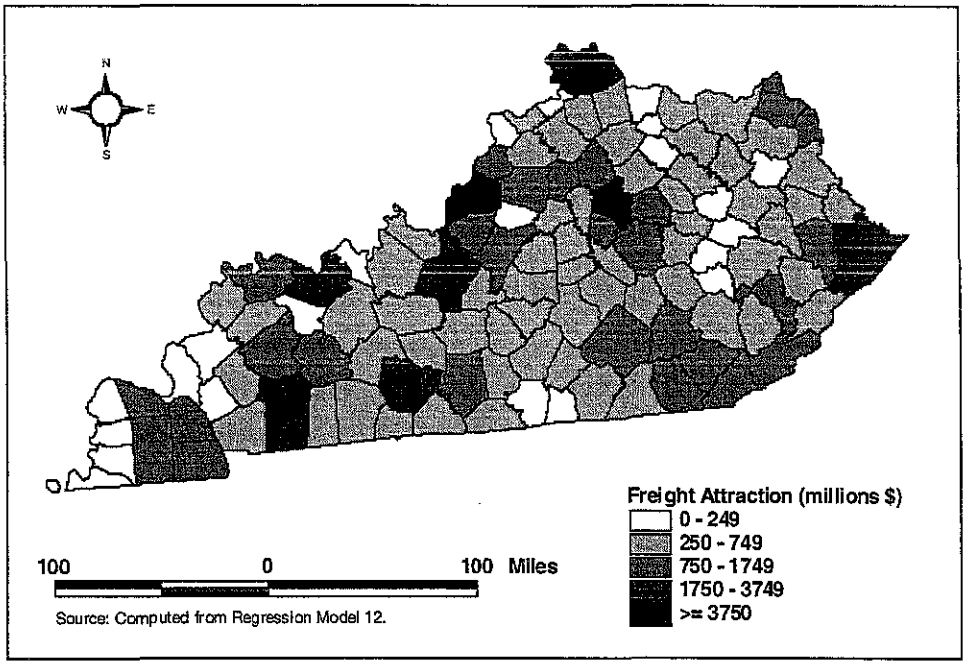

Figure 16

Results of Model 12

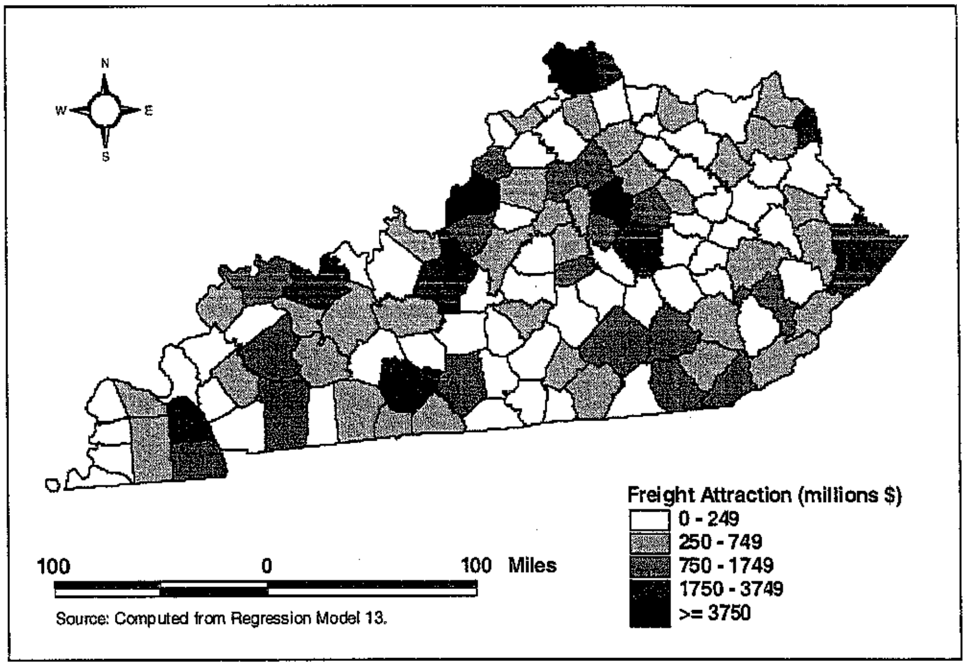

Figure 17

Results of Model 13

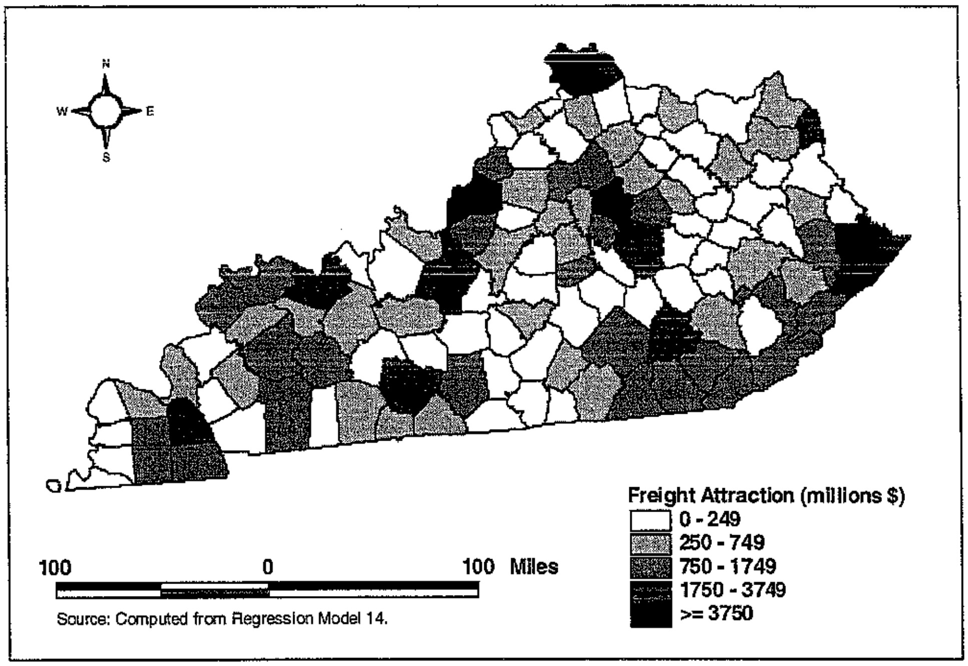

Figure 18

Results of Model 14

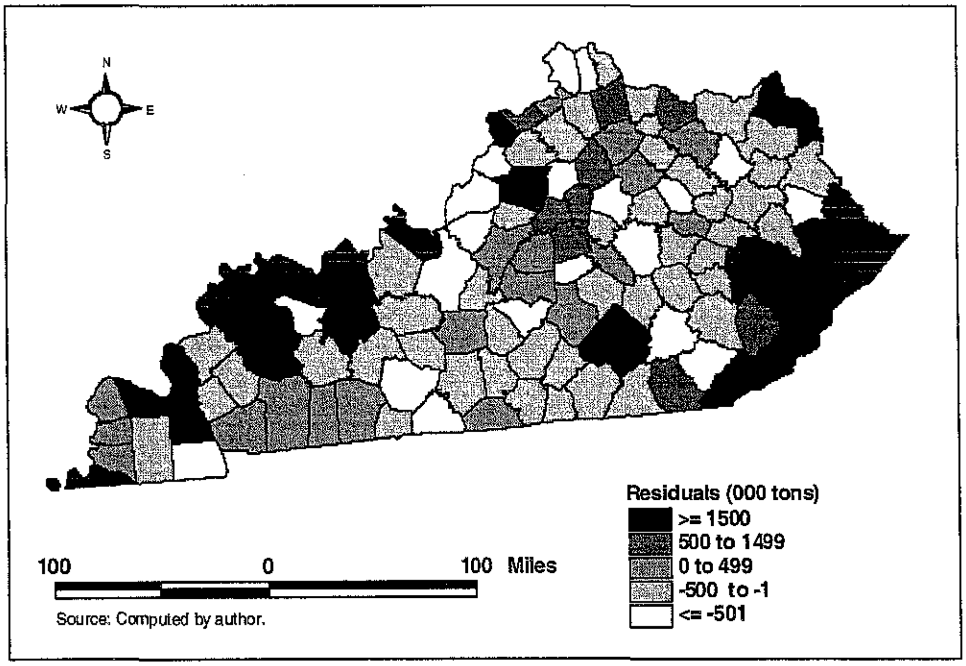

Figure 19

Comparison of TRANSEARCH data to Model 2 results (residuals are obtained by subtracting Model 2 results from TRANSEARCH data)

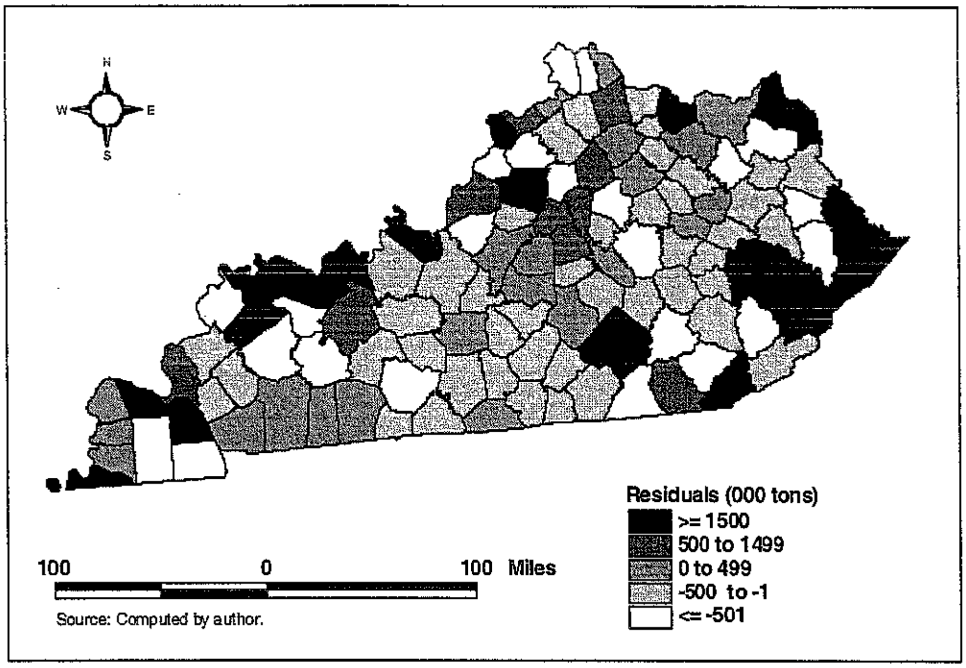

Figure 20

Comparison of TRANSEARCH data to Model 3 results (residuals are obtained by subtracting Model 3 results from TRANSEARCH data)

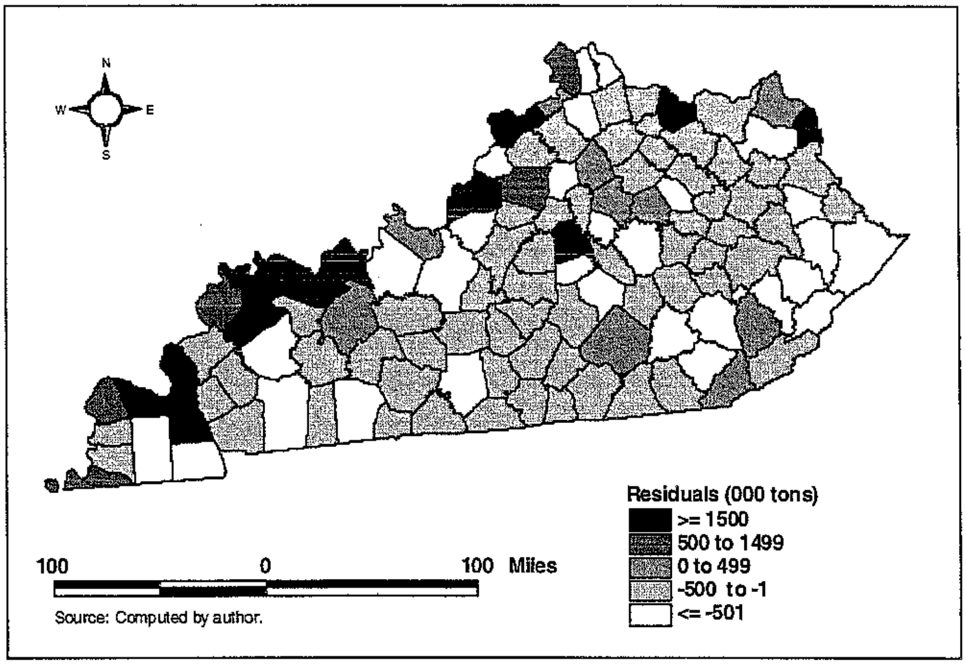

Figure 21

Comparison of TRANSEARCH data to Model 8 results (residuals are obtained by subtracting Model 8 results from TRANSEARCH data)

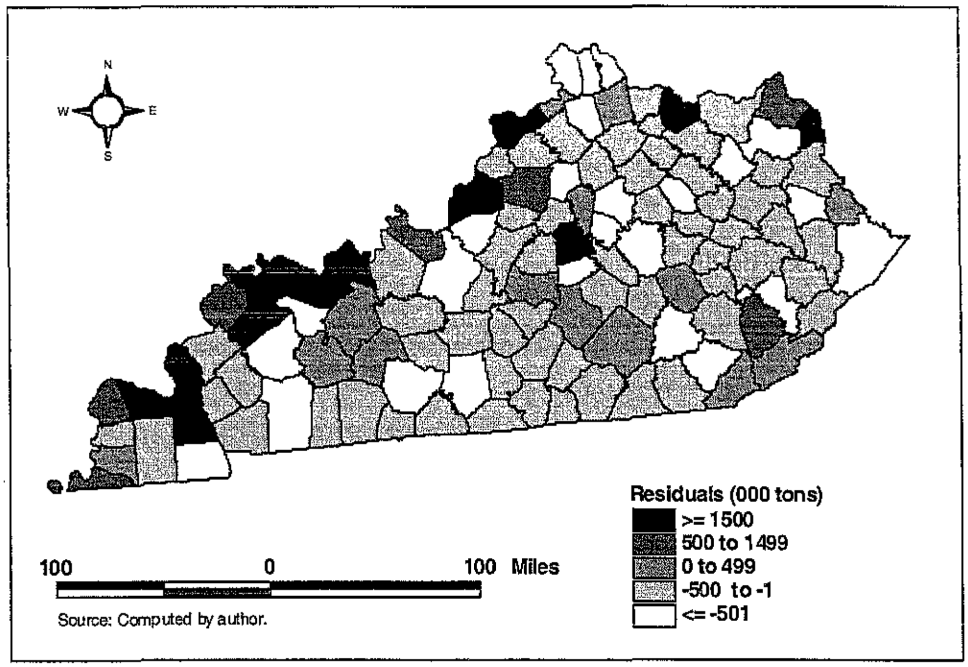

Figure 22

Comparison of TRANSEARCH data to Model 9 results (residuals are obtained by subtracting Model 9 results from TRANSEARCH data)

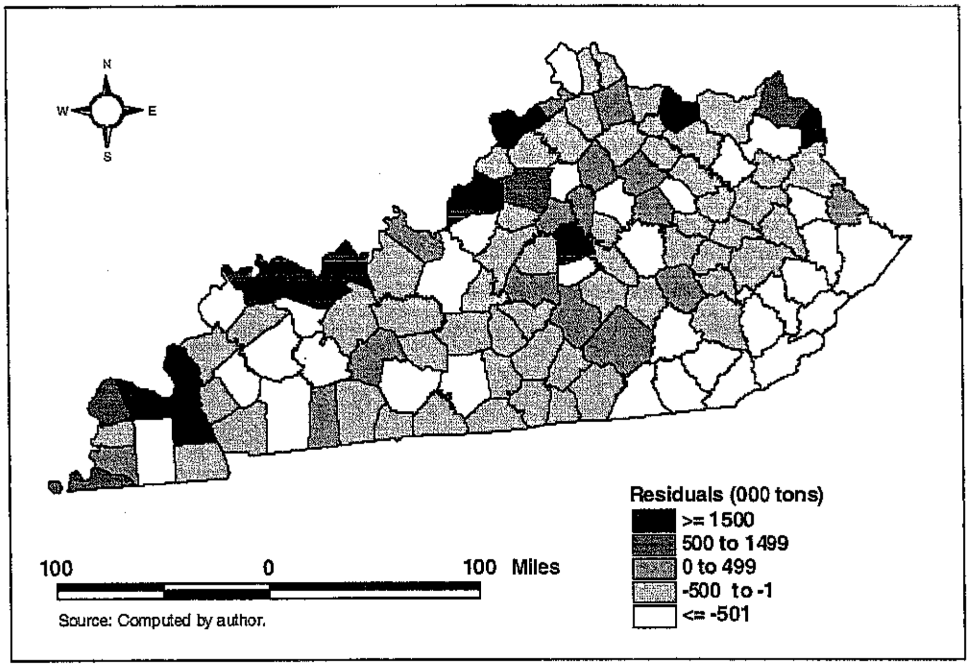

Figure 23

Comparison of TRANSEARCH data to Model 10 results (residuals are obtained by subtracting Model 10 results from TRANSEARCH data)

1 UTMS or the four-stage model consists of four models that are applied sequentially; notably, trip generation, trip distribution, modal split and traffic assignment.