Evaluating the Local Employment Dynamic Program as an Alternate Source of Place of Work Data For Use by Transportation Planners

Wende A. Mix, Ph.D.
Associate Professor
Department of Geography and Urban Planning
Buffalo State College
1300 Elmwood Avenue
Buffalo, NY 14222
PH (716) 878-6216
FAX (716) 878-4009

Word count: 8783
Submission Date:

ABSTRACT

In 2003, a pilot study was conducted for the Bureau of Transportation Statistics, USDOT on the feasibility of generating place of work data from the Longitudinal Employment – Household Dynamic (LEHD) program (now called Local Employment Dynamic (LED)) at the U.S. Census Bureau. This study resulted in several data files containing place of residence, place of work, wage, and industry classification data for the states of Illinois and Florida. All data was referenced to Census block level geography. This paper compares results from this study to data from the 2000 Decennial Census and the American Community Survey. The emphasis of this work is on comparing data products that are readily available to ordinary data users, and not on how these data products are created. Suggestions are made to improve the LED products to further enhance their utility to transportation planners.

INTRODUCTION

Urban transportation planners require information about where people live and work, as well as characteristics of those workers, to analyze system performance and evaluate alternative planning and policy strategies. Traditionally, this information has been obtained from local travel surveys and the long-form sample questionnaire of the decennial census. In the future, journey-to-work data will be available from the Census Bureau's new American Community Survey (ACS) which will replace the long-form questionnaire. Planners are always seeking more current information during the periods between surveys and censuses with which to monitor changes in the distribution of population and employment as well as the commuting patterns that result.

In an effort to address this need, in 2003, the U.S. Department of Transportation's (DOT) Bureau of Transportation Statistics (BTS) contracted with the Census Bureau to conduct a pilot study of the feasibility of producing journey-to-work data from the Census Bureau's Local Employment Dynamic (LED) program. The LED program (formerly the Longitudinal Employment – Household Dynamic (LEHD) program) is an effort to link

and integrate administrative databases from a variety of state and federal sources using sophisticated statistical models to produce local labor market information.[1] The objective of the BTS pilot study was to develop origin-destination work trip data for two states, Florida and Illinois.

This paper compares 2001 work trip data for Illinois that were derived from the LED program as part of the BTS study with comparable information gathered by the decennial census and the ACS. First, additional background information on this study and other similar studies and an overview of the three data sources are provided. These sections are followed by comparisons of these data sets with respect to coverage, spatial distributions, flow patterns, and content. Finally, suggestions are made for improving the final product from the LED program to facilitate use by transportation planners.

STUDY BACKGROUND

BTS had several requirements for the product developed in this pilot study. First, BTS insisted on block level summaries so planning agencies could summarize the results using their own transportation analysis zones. Wage and industry classifications also were required in the data products. Much of the BTS pilot study's effort went to working with state departments of employment security to clean and verify employer address information found in their administrative databases. Further, procedures for insuring confidentiality at the block level had to be developed and implemented. The result was two data sets containing several tables on home-to-work based flows.

Many of the issues associated with using administrative data for transportation planning are discussed by Souleyrette, Plazak, Strauss, and Andrle (1) in a paper demonstrating linking ES202 data to driver license information in Iowa to determine Origin/Destination work flows. These authors provide a brief summary of the issues surrounding ES202 employment data, such as uncovered work force, inaccurate and non-specific address information, reporting problems, and confidentiality. Examples are provided of how some state transportation agencies are using administrative employment data.

Several differences in Souleyrette et al's approach to estimating and validating trip tables and the BTS pilot described here are important to note. First, by linking ES202 to driver license data, a question arises on accounting for workers who do not drive. The LED approach links employer information to federal tax forms. In addition, Souleyrette et al. validate three different synthesized home based work (HBW) trip tables using a base 1990 model, comparing assigned flows to actual ground counts on links with greater than 2000 average daily traffic. During the past decade, home based work trips comprise a smaller proportion of the daily traffic. According to the 2001-2002 National Household Travel Survey (NHTS) 45 percent of daily trips are taken for shopping and errands, 27 percent of daily trips are social and recreational, such as visiting a friend, and only 15

[1] More information on the LED program and technical papers describing the methodology may be found at http://lehd.dsd.census.gov/led/00/index.html.

percent of daily trips are taken for commuting. This makes validation to ground counts complex.

This study compares data from the 2000 Census, the 2001 and 2002 ACS, and the LED 2001 Prototype Origin-Destination Matrix and Block Characteristics Files. The purpose of this comparison is to provide some insight into the validity of the LED data in hopes of addressing the question that BTS now faces with regard to pursuing this option for providing transportation planners with annual flow data. The next section provides relevant background for these data sources and discusses substantial differences that impact direct comparison.

OVERVIEW OF DATA SOURCES

Decennial Census

The 2000 census long-form questionnaire was sent to a 1-in-6 (about 17 percent) sample of all U.S. households. The journey-to-work data from the long form are weighted estimates of the entire population of workers 16 years of age and older. The reference period for journey-to-work data from the census refers to the location where workers usually worked during the week prior to Census Day, April 1st, of the census year.

Journey-to-work data have been collected in the long-form component of each decennial census since 1960. Beginning with the 1970 census, place-of-work responses were geographically coded to the block level within the urbanized portions of metropolitan areas. Also beginning with the 1970 census, the Census Bureau produced special tabulations tailored to the characteristics and geography needed for transportation planning. These special tabulations, called Census Transportation Planning Package (CTPP), have been produced after each subsequent census. The data they provide have become the benchmark for urban transportation planning in U.S. metropolitan areas.

The 2000 census was the last census that will include the use of a long-form sample questionnaire with which to collect detailed social, economic, and housing information like the journey-to-work. Instead, the Census Bureau is replacing the long-form questionnaire a new program, the American Community Survey (ACS). The ACS will provide data comparable to that obtained from the census long form, but at a frequency greater than once every ten years.

This paper compares LED data with 2000 census journey-to-work data at the county level and for tract-to-tract commuter flows. The county level data come from standardized tables on Census 2000 Summary File 3, available on the Census Bureau's web site at www.census.gov. Journey-to-work flow tables at the tract and block group levels were obtained from part 3 of CTPP 2000, available through the Bureau of Transportation Statistics' web site at www.bts.gov.

The American Community Survey

Unlike the "one-point in time, every-ten-years" census long-form, the ACS is a continuous survey in which a (new) sample of households in each U.S. county receives a questionnaire each month throughout the year. When it is fully implemented nationwide, beginning in January, 2005, the ACS will collect data from about 3 million households per year. That sample size will be sufficient to produce yearly estimates for counties, places, and metropolitan areas with a population of 65,000 or more. Multi-year averages will be required to produce data for smaller areas. For example, census tracts will require data averaged over five years to reach a cumulative sample size comparable to the census long-form. The ACS reference week for journey-to-work data will include every month and season of the year for annual averages and every month and season of the years that comprise the multi-year averages.

Although, as noted above, the ACS will not be fully implemented until 2005, comparable data from ACS "supplemental surveys" are available on the Census Bureau's web site for approximately 1,200 U.S. counties, cities, and metropolitan areas. Included among these areas are nine counties in Illinois: Cook, DuPage, Kane, Lake, McHenry and Will counties in the Chicago-Naperville-Joliet, IL-IN Metropolitan Statistical Area (MSA), Madison and St. Claire counties in the St. Louis, MO-IL MSA, and Winnebago County in the Rockford-Freeport-Rochelle, IL MSA. For this paper, LED data are compared with journey-to-work data for these counties.

Tract level journey-to-work data are not yet available from the ACS, so no comparison with the LED data can be made at this time. However, data users may be interested in perusing the results of a Census Bureau study comparing 1999-2001 three-year average estimates at the tract level for 36 ACS test counties to the 2000 census. The results for Lake County, IL, the only Illinois county among the test sites, can be found at http://www.census.gov/acs/www/AdvMeth/acs_census/17097.htm. Also, the results of a study funded by the Federal Highway Administration (Mix, 2003) that found that tract-to-tract commuter flow data from the 1999-2001 ACS test for Broward County, Florida and San Francisco, California were closely comparable to 2000 census results are available at www.trbcensus.com/notes.html.

LED 2001 Prototype Origin-Destination Matrix and Block Characteristics Files

The LED data sets are created using administrative data from multiple sources.[2]. Statistical models are used to assign workers to places of work. The primary file used to determine workplace locations is from participating[3] states' employment security programs. These files, referred to as "ES202 Files", contain the legal and "doing business as" name of each employer, their physical and mailing address, and the number of employees. For the BTS demonstration project, quarterly ES202 Files were combined to represent annual workplace data. As with many administrative databases, missing and

[2] For a detailed explanation of how the LED estimates are created, see Abowd, Lengermann, and Vilhuber (2002).
[3] Currently, participation in the LED program is voluntary, and not all states participate.

incomplete information required initial efforts to clean the data. Staff from the State of Illinois reviewed and modified location information for employers whose physical address was not geocodable to the block level. Generally, place of residence information is derived from the Census Bureau's Statistical Administrative Records System (StARS). For the prototype, however a geocoded extract from federal tax forms was used to identify residence block. LEHD used only those workers who were geocoded to a Census block, eliminating 10% of the workers from the pilot study (see Roemer, 2003).

BTS requested the origin-destination (OD) tables at the block level to facilitate creation of transportation analysis zones by data users. Confidentiality requirements resulted in publishing data if the block of residence had at least five workers living in it who traveled to at least three different blocks of work. According to the documentation provided with the LED datasets, there should be a 2% difference in workers in the OD dataset (as compared to the Home Block Characteristics File) due to confidentiality requirements.

Census produced three files from the LED Program for the BTS pilot study (see Roemer, 2003). The OD Matrix File contains references to the residence block, work block, and the number of workers traveling between these blocks. The Home Block Characteristics File contains a reference to the home block, the number of workers living in the block, the proportion of workers with low ($0 < annual earnings ≤ $12,000), medium ($12,000 < annual earnings ≤ $35,000), and high (annual earnings > $35,000) earnings, and the average wage of workers residing in the block. The Work Block Characteristics File contains a reference to the work block, the mean monthly pay per worker, the disclosure status of the mean monthly pay, and ten binary flags indicating that an establishment within the referenced Standard Industrial Classification (SIC) code operates on the block. For example, if the first SIC flag is 1 then an industry in the SIC Division "Agriculture, Forestry, and Fishing" operates on the block.

Since the LED data derive from state Unemployment Insurance System information, and represent private, non-farm wage and salary employment, exclusions in its coverage must be taken into account when comparing the data with other employment data.[4] The LED data exclude members of the Armed Forces, self-employed workers, domestic workers, unpaid family workers, most farmers and agricultural employees, and federal government workers.

RESULTS

LED data at the county level were compared with both 2001 and 2002 ACS and 2000 census data. LED data portraying worker flows at the census tract level were compared with 2000 census data only since data at that level are not available from the ACS. The 90 percent confidence intervals (CI) for the ACS data were obtained from the ACS data tables and calculated for all decennial data. Currently, it is not possible to estimate confidence intervals for the LED data.

[4] For a discussion of the types of employment that are not covered by state unemployment insurance laws, see Stevens (2002).

Number of Workers

Decennial data are available for all 102 counties in Illinois. Total workers in the State of Illinois (from Table P049: Universe: Employed Civilian Population 16 years or older) is 5.83 million. If, however, total workers includes anyone who worked during the reference week (Table P026: Universe: Workers 16 years and over) this number is 5.75 million and the estimate of those workers who live and work in Illinois (comparable to LED coverage) is 5.56 million. The estimate of civilian workers from Table P049 drops to 5.2 million when considering only those employed civilians covered in the LED dataset. The 5.56 million figure from Table P026 includes categories of workers that are not covered in the LED.

Two LED estimates for the State of Illinois may be derived from the Home-Based (HB) Characteristics file (workers) and the OD Matrix File (Travelers/Commuters). As noted previously, 10% of the workers are ineligible for inclusion in the HB file because they were not geocodable to the block level. In addition, LED staff estimated that another 2% of workers[5] are missing in the OD file because of confidentiality requirements. The unadjusted number of total workers from the LED HB file is 5.15 million which increases to 5.67 million with a 10% adjustment. The unadjusted number of total commuters from the LED OD file is 4.30 million which increases to 4.39 million with a 2% adjustment. LED total workers from the HB file is comparable to decennial census numbers. However, the commuter flows represented in the OD matrix are significantly lower than the decennial estimates[6].

The 2001 and 2002 ACS included six counties (Cook, DuPage, Kane, Lake, McHenry, and Will) in the Chicago-Naperville-Joliet, IL-IN-WI MSA, one county (Winnebago) in the Rockford, IL MSA, and two counties (Madison and St Clair) in the St. Louis, MO-IL Metropolitan Statistical Area. These nine counties contain 72% of all workers, according to the decennial data. However, the LED HB and OD tables indicates that these nine counties have 75% and 76%, respectively, of all workers in Illinois

Table 1 shows the estimated number of workers residing in each county in Illinois, based on the ACS and the LED[7]. The ACS estimates in the table are derived in order to approximate the LED universe. The employed civilian population 16 years old and over working in all industries including agriculture, forestry, fishing and hunting, and mining was obtained for each county (2001, 2002 ACS Table P068). In each of these industries,

---

[5] The actual difference in the totals from the HB and OD files is 16.5%.

[6] All LED estimates are below the lower bound on the 90% CI of adjusted Decennial estimates , for all 102 counties, when total workers from the LED is determined using the OD summary file. When total workers is calculated using the HB summary file, four counties have estimates that fall within the 90% CI for the adjusted Decennial estimate. They are Grundy, Kankakee, Macon, and Whiteside. Otherwise, 85 counties have LED estimates below the decennial lower bound and 13 counties have LED estimates that are above the decennial upper bound.

[7] LED numbers are unadjusted. The only way to distribute the 10% of workers associated with geocoding errors to each county is based on the population distribution in each county. Similarly with the 2% error associated with confidentiality suppression.

private for-profit wage and salary workers; employee of private company, private not-for-profit wage and salary workers, local government workers, and state government workers were included in the totals. Census summaries by place of residence still include workers who work outside the state but live in the area of interest.

TOTAL WORKERS

Modification of ACS P068: SEX BY INDUSTRY BY CLASS OF WORKER FOR THE EMPLOYED CIVILIAN POPULATION 16 YEARS AND OVER - Universe: EMPLOYED CIVILIAN POPULATION 16 YEARS AND OVER

AREAS	**ACS 2001 Estimate**	*Lower Bound*	*Upper Bound*	**ACS 2002 Estimate**	*Lower Bound*	*Upper Bound*	**LED Estimate OD Summary File**	**LED Estimate HB Summary File**
Cook County, IL	**2,206,950**	*2,169,319*	*2,244,581*	**2,134,998**	*2,093,728*	*2,176,268*	**1,923,997**	**2,242,317**
DuPage County, IL	**429,570**	*415,465*	*443,675*	**378,835**	*359,878*	*397,792*	**375,982**	**454,572**
Kane County, IL	**201,030**	*191,613*	*210,447*	**179,509**	*168,736*	*190,282*	**154,449**	**191,873**
Lake County, IL	**280,438**	*275,163*	*285,713*	**251,004**	*244,490*	*257,518*	**246,070**	**295,126**
McHenry County, IL	**126,363**	*119,172*	*133,554*	**119,706**	*112,312*	*127,100*	**108,154**	**132,039**
Will County, IL	**244,832**	*234,421*	*255,243*	**223,671**	*211,647*	*235,695*	**204,108**	**257,214**
Madison County, IL	**106,691**	*100,081*	*113,301*	**65,086**	*55,669*	*74,503*	**72,044**	**89,804**
St. Clair County, IL	**98,700**	*91,290*	*106,110*	**76,684**	*68,482*	*84,886*	**65,707**	**81,875**
Winnebago County, IL	**114,213**	*106,145*	*122,281*	**104,019**	*95,961*	*112,077*	**106,473**	**131,035**

Table 1: Total workers by place of residence.

The LED estimate from either the OD or HB summary file falls within the 90% CI of the 2001 or 2002 ACS estimate of total workers in all but one county. The only county where both LED estimates are different from the two ACS estimates is Will County. The LED OD estimate of workers is different from both ACS estimates in five counties (Cook Co., Kane Co., McHenry Co., St. Clair Co., and Will Co.) In each of these counties, the LED OD estimate is below the lower bounds of the ACS estimates. The LED HB estimate of workers is different from both ACS estimates in five counties (DuPage Co., Lake Co., Madison Co., Will Co., and Winnebago Co.). In all but Madison Co., the LED HB estimate is above the upper bound of the ACS estimates. In Madison Co. the LED HB estimate is below the lower bound for ACS 2001 and above the upper bound for ACS 2002.

Estimates of total workers, both from the ACS and the Decennial, include people who live in the county but work outside the state. LED estimates do not include this group. Table P043 from the ACS and table P026 from the decennial SF3 contain county level information concerning where people work who work in the state of residence. The

proportion of workers who work in the same county of residence is shown for the ACS data sets in Table 2. The 90% confidence intervals have been estimated for these proportions. None of the proportions are statistically different from each other for these datasets or from the decennial estimates (not shown). However the LED estimate falls below the lower bound of the CIs for all nine counties[8]. The lower proportion of internal county trips in the LED data is partially explained by the multi-unit/headquarter issue in the administrative data. Workers are assigned to headquarters of multi-units instead of to the actual workplace. These multi-unit/headquarter records are the records that need more attention during administrative data cleaning.

Proportion of those workers who work in the state of residence and county of residence Source ACS 2001 and 2002 Table P043							
County	**ACS 2001 Estimate**	*Lower Bound*	*Upper Bound*	**ACS 2002 Estimate**	*Lower Bound*	*Upper Bound*	**LED Estimate**
Cook County, IL	**88.3**	*87.4*	*89.1*	**89.0%**	*88.2%*	*89.8%*	**84.3%**
DuPage County, IL	**57.9**	*55.2*	*60.5*	**59.7%**	*57.0%*	*62.5%*	**50.1%**
Kane County, IL	**58.6**	*55.3*	*61.9*	**58.7%**	*54.7%*	*62.7%*	**46.0%**
Lake County, IL	**67.6**	*66.2*	*69.1*	**67.0%**	*65.5%*	*68.6%*	**57.4%**
McHenry County, IL	**47.6**	*44.0*	*51.3*	**45.5%**	*41.4%*	*49.6%*	**42.3%**
Will County, IL	**46.8**	*43.7*	*49.8*	**41.7%**	*37.7%*	*45.7%*	**34.6%**
Madison County, IL	**83.9**	*79.5*	*88.2*	**86.1%**	*82.4%*	*89.9%*	**75.1%**
St. Clair County, IL	**89.3**	*84.7*	*93.9*	**89.9%**	*84.5%*	*95.2%*	**74.8%**
Winnebago County, IL	**89.7**	*86.6*	*92.9*	**89.2%**	*86.7%*	*91.7%*	**78.5%**

Table 2: Proportion of internal county workers to internal state workers.

Origin – Destination Flows

It is not possible, at this time, to explore more detailed flow data using the ACS. Even though the estimates explored so far between the LED and the decennial data indicate that the two datasets are not comparable with respect to total workers, further analysis of distribution of flows and flow patterns, reveal more similarities in these datasets.

Flow Distributions by Area Type

Statewide flow data in the CTPP part 3 is available at the census tract level. In metropolitan areas, flow data will be available at the block group level. The LED flow data may be aggregated to both the block group and tract level.

Of interest in comparing these datasets are the general flow patterns by area type for a region. These nine flow patterns are characterized as intra-urban flows, intra-suburban flows, intra-rural flows and flows between urban, suburban and rural areas.

[8] In fact, for all 102 counties, the LED estimate falls below the lower bound of the decennial estimate.

Four additional flows are considered; those associated with urbanized areas outside of a metropolitan area.

To identify suburban areas within the state of Illinois, metropolitan area boundary files (MSA and PMSA), urbanized areas (UA), central city (CC), tract and block group boundary files were downloaded from the Census Bureau web site. First, block groups are classified as urban using the central city boundaries within MSA/PMSA boundaries in the state of Illinois. Any block group whose geometric center falls within a central city is classified as urban. Next, all block groups outside the central city but within an urbanized area boundary in an MSA/PMSA are classified as suburban. All other block groups are classified as rural unless they fell within an urbanized area that is not part of a metropolitan area.

The proportion of a tract's area that is classified urban, suburban, urbanized, or rural is used to characterize each tract. Consequently, a tract that contains 50% or more urban area, based on the area of the block groups within the tract, is classified as urban. If a tract's area has a majority of suburban land, the tract is classified as suburban. A similar procedure is used for rural area classification. Table 3 shows the proportion and area of block groups and tracts that are classified as urban, suburban, urbanized area, and rural. As expected the majority of the area within the state is classified as rural. However, urban and suburban groups contain more than a third of the total geographic features.

Area Type	Percent of Total Block Groups	Percent Area	Percent of Total Tracts	Percent Area
Urban	37%	1%	41%	1%
Suburban	35%	4%	31%	4%
Urbanized Area	7%	1%	3%	0%
Rural	21%	94%	25%	95%

Table 3: Summary of geographical assignment to area types.

Tables 4 and 5 show the distribution of flows to each area type from CTPP and LED tract level data. Since the LED contains only information on trips internal to the state of Illinois, only trips originating in AND destined to tracts in Illinois are considered in this analysis. The CTPP has nearly one million more internal state work flows than the LED data. However, this analysis does not exclude flows in the CTPP that are not significant at the 90% confidence level (very low flows between interchanges).

Total Flow	Urban	Suburban	Urbanized Area	Rural	Total Productions
Urban	1,229,541	406,166	1,193	97,456	1,734,356
Suburban	533,849	1,502,074	771	180,032	2,216,726
Urbanized Area	6,335	1,725	56,830	47,836	112,726
Rural	257,386	213,908	93,936	661,527	1,226,757
Total Attractions	2,027,111	2,123,873	152,730	986,851	5,290,565
NOTE: 3471 interchanges have destination tract ID 999999 (200280 total trips); this includes 2810 trips from Urban origins, 1876 trips from Suburban origins, 23,138 trips from Urbanized Areas, and 174,456 trips from Rural origins.					

Table 4: CTPP flow summary

Total Flow	Urban	Suburban	Urbanized Area	Rural	Total Productions
Urban	929,540	372,109	4,125	85,967	1,391,741
Suburban	518,504	1,214,899	3,285	151,235	1,887,923
Urbanized Area	10,521	5,939	45,665	43,322	105,447
Rural	206,908	194,792	74,398	440,557	916,655
Total Attractions	1,665,473	1,787,739	127,473	721,081	4,301,766

Table 5: LED flow summary

In both datasets the greatest number of trips is between suburban areas. Suburban areas both produce and attract more trips than urban areas, consistent with sprawl occurring in the region. Low flow levels between suburban (or urban) and urbanized areas are expected given that no suburban (or urban) area is adjacent to urbanized areas. Urbanized areas, by the definition used here, are surrounded by rural areas. For all area types, the number of trips is greatest along the diagonal (or from/to the same area type). For example, most rural trip productions are allocated to rural destinations in both datasets. Flows originating in rural areas are almost equally likely to go to an urban area or a suburban area. However, flows originating in urbanized areas are at least twice as likely to end in urban areas as opposed to suburban areas.

Comparing the proportion of flow between area types indicates that the two databases actually distribute trips quite similarly at this level of aggregation. Tables 6 and 7 show the flow distributions for each dataset. Both datasets allocate 28% of the

total flows between suburban areas. Overall, the distributions are virtually the same, especially for urbanized areas.

Total Flow	Urban	Suburban	Urbanized Area	Rural	Total Productions
Urban	23%	8%	0%	2%	33%
Suburban	10%	28%	0%	3%	42%
Urbanized Area	0%	0%	1%	1%	2%
Rural	5%	4%	2%	13%	23%
Total Attractions	38%	40%	3%	19%	100%

Table 6: Distribution of flows from the CTPP.

Total Flow	Urban	Suburban	Urbanized Area	Rural	Total Productions
Urban	22%	9%	0%	2%	32%
Suburban	12%	28%	0%	4%	44%
Urbanized Area	0%	0%	1%	1%	2%
Rural	5%	5%	2%	10%	21%
Total Attractions	39%	42%	3%	17%	100%

Table 7: Distribution of flows from the LED.

Intraurban trips are further analyzed as flows within the same central city and flows between different central cities. Tables 8 and 9 show the results from comparing intraurban trips. The census data indicates a slightly higher proportion of flows within the same central city.

Between Urban trips	Flows	Interchanges	% of Total Flow	Avg. Trips per Interchange
Same Central City	1,150,225	53,869	94%	21
Different Central Cities	79,316	4,892	6%	16

Table 8: CTPP intraurban flows.

Between Urban trips	Flows	Interchanges	% of Total Flow	Avg. Trips per Interchange
Same Central City	839,403	149,092	90%	6
Different Central Cities	90,137	31,489	10%	3

Table 9: LED intraurban flows.

An interchange is an origin-destination pair of tracts with positive flow. The census data has a much higher number of average trips per interchange than the LED data. In general, LED distributes trips much more widely throughout the region; a more dispersed pattern of origins. In the LED data, workers come from many more origins as evidenced by the low average trips per interchange in Table 11 compared to the census averages in Table 10.

Average trips per Interchange	Urban	Suburban	Urbanized Area	Rural	Total Productions
Urban	21	12	8	17	18
Suburban	14	19	9	19	17
Urbanized Area	12	9	109	32	41
Rural	20	15	38	37	26
Total Attractions	19	17	48	29	19

Table 10: CTPP trips per interchange.

Average trips per Interchange	Urban	Suburban	Urbanized Area	Rural	Total Productions
Urban	5	3	2	3	4
Suburban	4	6	1	4	5
Urbanized Area	3	2	38	8	9
Rural	5	4	11	11	7
Total Attractions	5	4	10	7	5

Table 11: LED trips per interchange.

Internal Tract Flows

Table 12 provides statistics on how each dataset allocates within tract trips by area type. These flows represent people who live and work in the same tract. The LED dataset has slightly more tracts with no internal trips than the CTPP dataset. The LED allocates only 6% of all trips as internal. Eleven percent (11%) of the CTPP total trips are internal. This difference may be attributable to the LED data capturing only those employers who pay unemployment insurance, missing self-employed workers. For urban tracts, only 3% of the total trips that begin in urban areas are internal in the LED dataset, compared to 7% in the CTPP dataset. As expected, tracts in urbanized areas and rural tracts have the highest proportion on total flow that is internal in both datasets. This represents people who live and work in small towns as well as those who live and work in larger rural tracts.

Area Type	Total Tracts	Total Trips	Percent of Total Trips	Average Trips per Tract with Positive Flow	Minimum Internal Trips	Maximum Internal Trips	Std Dev	Tracts with no Internal Flows
LED Data								
Urban	1087	38,192	3%	35	1	714	52	
Suburban	907	89,033	5%	98	1	657	81	
Urbanized Area	94	16,967	16%	181	2	780	167	
Rural	752	129,607	14%	172	1	1338	189	
Total	2840	273,799	6%	96				124
CTPP Data								
Urban	1104	118,075	7%	107	4	2115	142	
Suburban	910	185,480	8%	204	10	1880	152	
Urbanized Area	94	26,420	23%	281	20	1010	214	
Rural	754	237,659	19%	315	4	1890	245	
Total	2862	567,634	11%	198				102

Table 12: Internal tract trips.

In general, the average number of internal tract trips in the LED dataset is much smaller than the average number in the CTPP dataset, for comparable area types. Overall, the LED dataset averages 96 internal trips per tract and the CTPP averages 198 internal tract trips. The difference in the average trips is particularly noticeable for urban tracts. On average, an urban tract in the LED dataset has 35 trips representing people who live and work in these small areas while the CTPP has 107 average internal trips per urban tract.

The distribution over area type of the average number of internal trips is slightly different in the two datasets. In the LED dataset, urbanized areas have the greatest average number of internal trips (181) whereas in the CTPP dataset, rural areas have the greatest average number of internal trips (315).

The maximum number of internal trips allocated to each different area type is, generally, much higher in the CTPP dataset. For instance, the CTPP allocates a maximum of 2115 internal trips to an urban tract compared to a maximum of 714 in the LED data. The standard deviation of the CTPP data, representing the spread of internal trips for each area type, is larger in the CTPP dataset.

Distance Decay

In the past, transportation planners have shown that people tend to live close to where they work. In other words, the number of trips between places is inversely related to distance. So, for any destination tract, we expect more trips from closer tracts and fewer trips from tracts farther away. A general form of this relationship is:

$$T_w = T_0 e^{bd-cd^2} \quad (1)$$

Where:

T_w = Total work flows
T_0 = Internal trips
d = Distance
b,c parameters to be estimated

When c = 0, the number of trips decreases exponentially with increasing distance. When b = 0, flows decline exponentially with the square of the distance. Figure 1 describes these three relationships graphically.

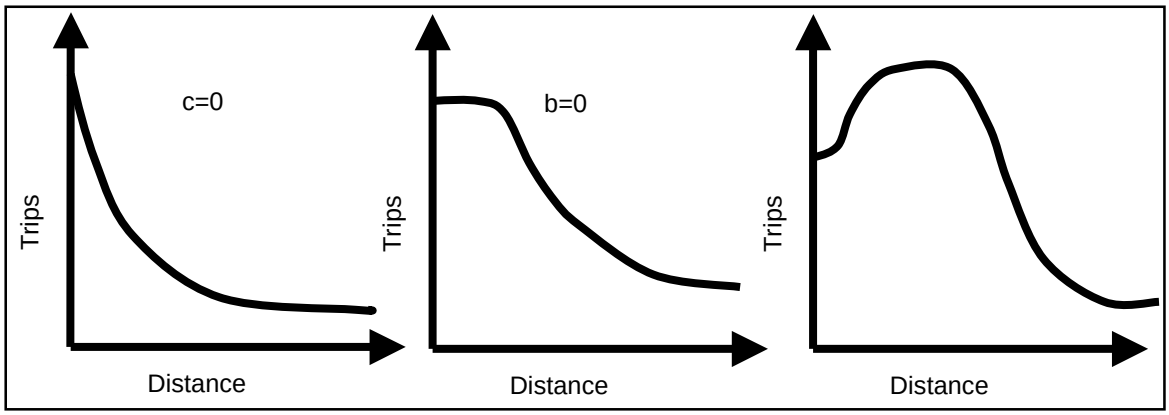

Figure 1. Distance decay relationships.

A scatter diagram of the data for suburban destinations from the CTPP and the LED indicates that the general form (b, c ≠ 0) may be warranted, see Figure 2. In fact, all area types, except urbanized areas, have fewer total trips between 0-10 miles from the work tract than between 10 – 20 miles from the destination.

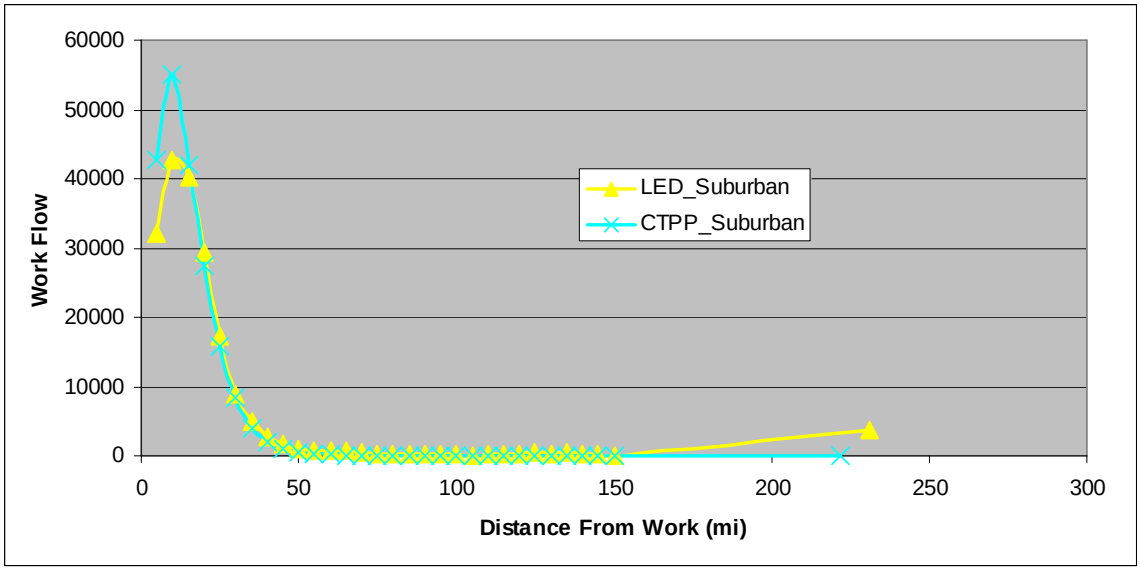

Figure 2: Relationship between flow and distance for suburban trips.

Equation 1 is transformed to the following relationship:

$$\text{Ln } T_w = \ln T_0 + bd - cd^2 \qquad (2)$$

For each of the four area types, the ten (10) destination tracts that attracted the most work trips in each dataset were analyzed (or 40 total destination tracts from each data set). Great circle distances were calculated between the geometric centers of each origin/destination pairs. Flow between tracts greater that 150 miles apart is excluded from the analysis.[9] Results of the regression analysis are shown in Table 13.

	LED		CTPP	
	Model	R^2	Model	R^2
Urban	$\ln(T_w) = 10.6 - 0.037d$	0.74	$\ln(T_w) = 11.3 - 0.072d$	0.84
	$\ln(T_w) = 12.3 - 0.11d + 0.0005d^2$	0.92	$\ln(T_w) = 13.2 - 0.15d + 0.0005d^2$	0.92
Suburban	$\ln(T_w) = 9.9 - 0.039d$	0.73	$\ln(T_w) = 10.5 - 0.070d$	0.86
	$\ln(T_w) = 11.7 - 0.12d + 0.0005d^2$	0.94	$\ln(T_w) = 12.4 - 0.15d + 0.0005d^2$	0.94
Urbanized Area	$\ln(T_w) = 7.9 - 0.028d$	0.73	$\ln(T_w) = 8.5 - 0.058d$	0.84
	$\ln(T_w) = 9.2 - 0.08d + 0.0003d^2$	0.90	$\ln(T_w) = 10.4 - 0.14d + 0.0005d^2$	0.95
Rural	$\ln(T_w) = 9.3 - 0.037d$	0.75	$\ln(T_w) = 9.8 - 0.077d$	0.83
	$\ln(T_w) = 11.1 - 0.11d + 0.0005d^2$	0.96	$\ln(T_w) = 12.1 - 0.16d + 0.0006d^2$	0.95

Table 13: Regression results

[9] In the LED dataset these interchange pairs represent those that need more attention during data cleaning. In the CTPP, few of these pairs have flows statistically different from zero at the 90% confidence level.

In general, the quadratic model performs better than the linear model relating trips to distance for all area types. For each area type, comparing the LED to the CTPP datasets, the linear model is a better fit to the CTPP data than the LED data. However, the quadratic model performs equally well for either dataset, where distance explains between 90 -96% of the variation in flows.

For the urban and suburban quadratic models, there is no significant difference (at the 95% confidence level) between the intercepts and the parameter estimates for either data set. None of the quadratic models have estimates for c that are significantly different between data sets. However, the urbanized area quadratic models for the LED and CTPP have both the intercept and the b estimates significantly different. For rural areas, the estimate for b is significantly larger than in the CTPP dataset.

All of the linear models have slopes that are significantly different; the CTPP slopes being consistently steeper than the corresponding LED slopes. So, according the CTPP data, distance from work (or proximity to work) is slightly more important than demonstrated by the LED data.

Tract flow patterns

A closer look at the flow data for a few tracts reveals that the some differences in work based flow patterns is in the details of these datasets. Figure 3 shows the importance of eliminating low flows in both of the datasets. The top set of maps depicts all work flows for two different tracts. The tract on the left is an urban origin tract and the flows are outbound from home to work. The tract on the right is an urbanized area destination tract and the flows are inbound from the tracts of residence to the tract of work. The LED dataset has flow patterns that are statewide for both of these tracts (reflecting records in the administrative database that need to be cleaned). For the urbanized area tract, CTPP flow pattern is dispersed more to the east (into the Chicago area) than seems logical. The bottom set of maps shows only flows greater than twenty (20) for the same two tracts. Under this condition, the flow patterns for both datasets quite similar.

Figures 4 and 5 demonstrate flow patterns from both datasets for an urban and a suburban tract. In these figures the CTPP and LED flows have been separated for easier visual comparison. Two flow levels are considered in these maps; flow greater than 20 and greater than 50. For the urban tract in Figure 4, the CTPP demonstrates a more dispersed flow pattern at both flow levels. In fact, for flows greater than 50 into this tract, the LED data indicates that these workers live predominantly northwest of the tract. The CTPP has workers coming from all directions (excluding Lake Michigan to the east). The flow patterns shown for the suburban tract in Figure 5 are quite similar for both datasets. The only glaring difference is a large flow of workers from a tract due west of this suburban tract (shown in pink) in the CTPP data. Also, the CTPP data contains slightly more flows in all directions from tracts further from the destination tract. CTPP flows originate more to the southeast than the LED flows, even at the higher flow level where Chicago is an origin.

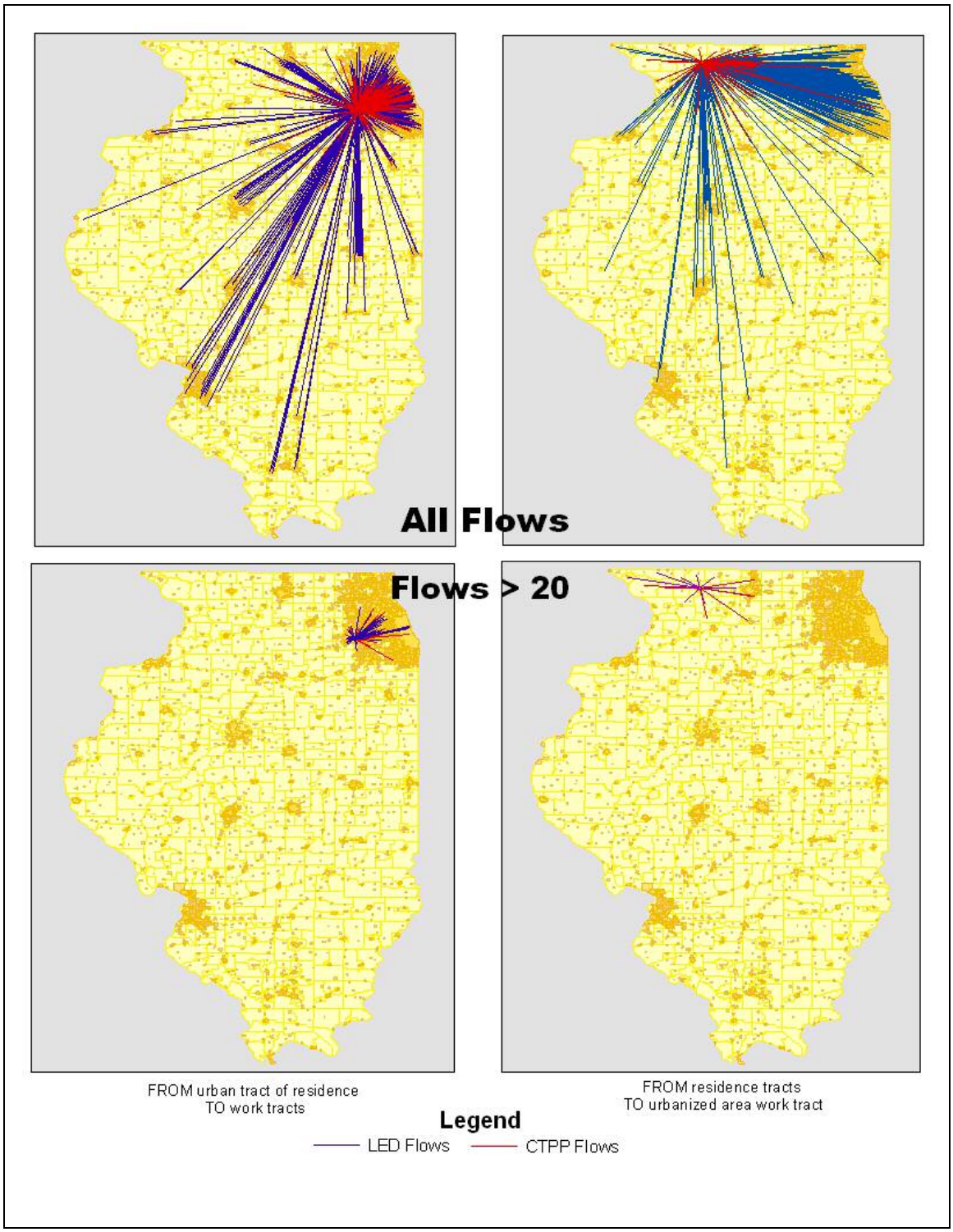

Figure 3: Comparison of tract level flows.

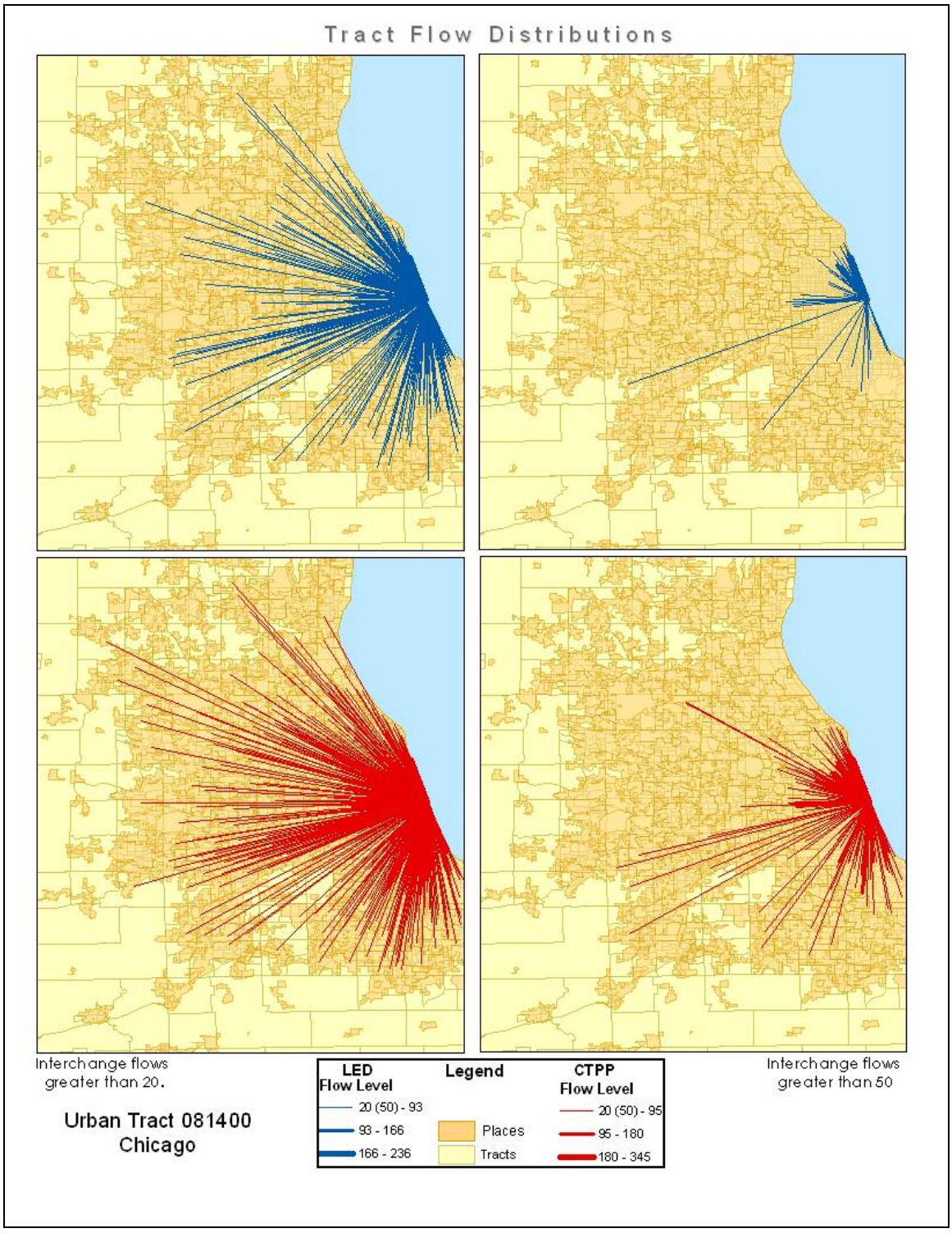

Figure 4: Home to work flows for an urban work tract.

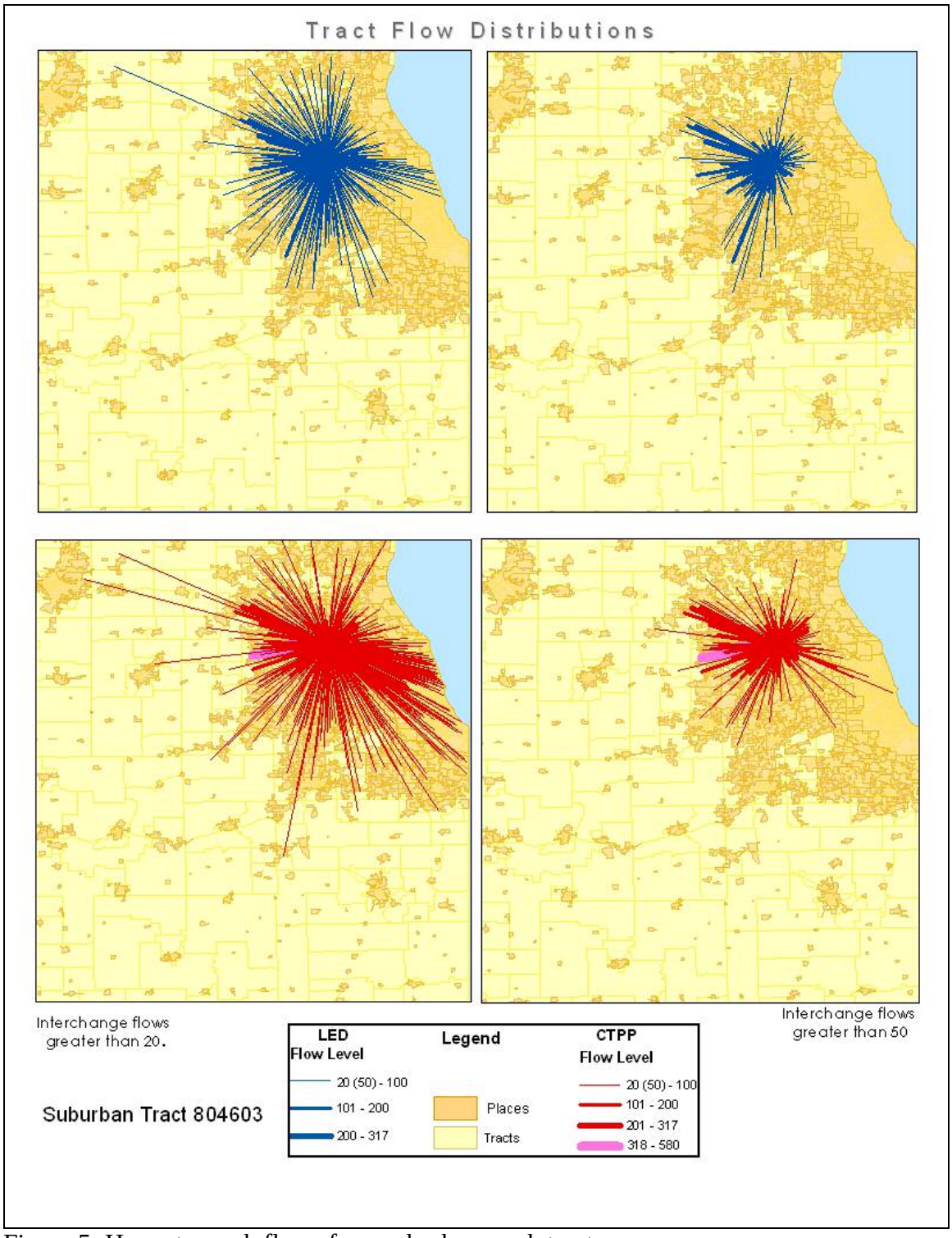

Figure 5: Home to work flows for a suburban work tract.

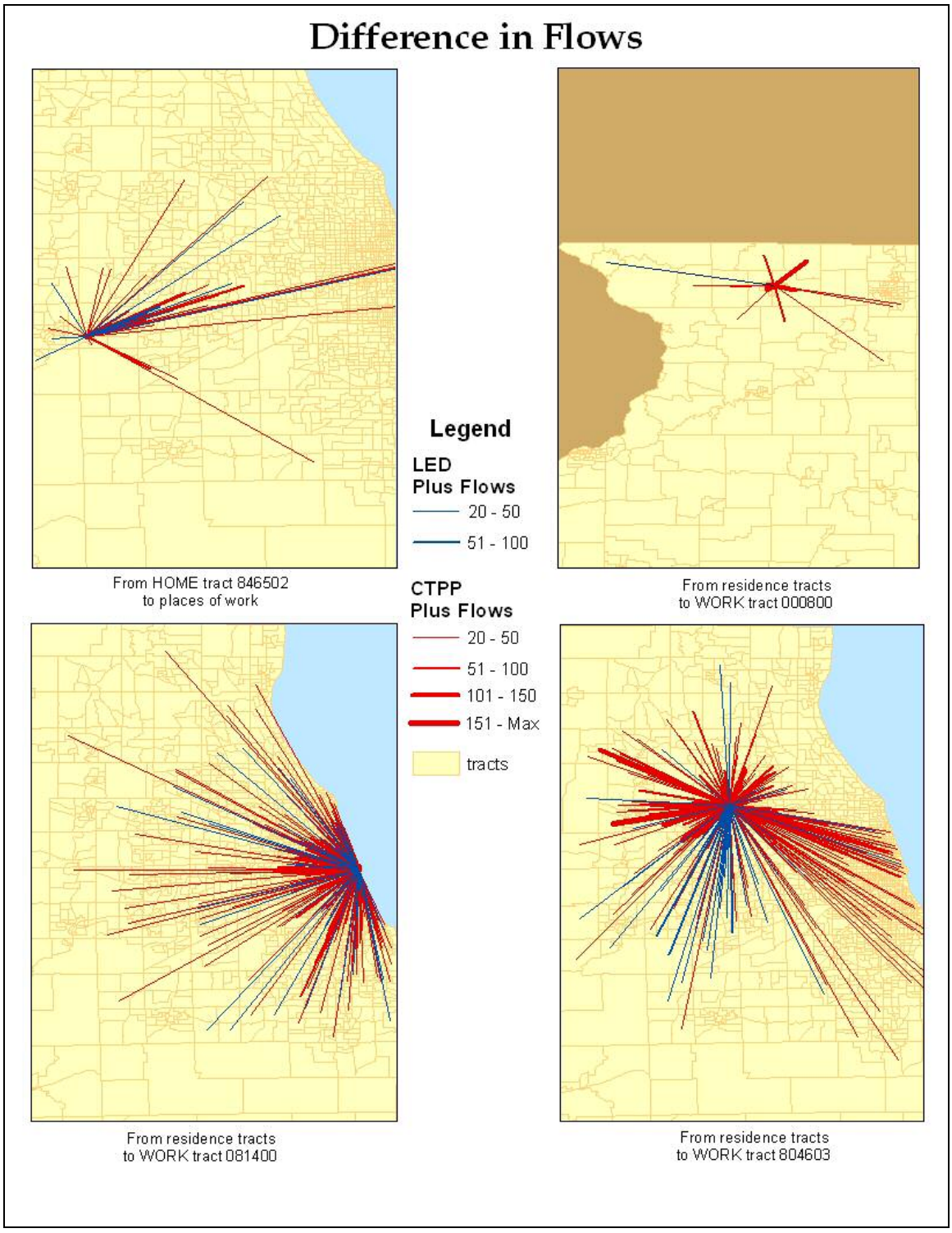

Figure 6: Absolute difference in flows between interchanges.

Figure 6 shows the actual differences in flow between the CTPP and LED datasets for each of the four tracts examined above. Red lines indicate that the CTPP flow exceeds the LED flow for an interchange by at least 20 trips. Blue lines indicate that the LED flow exceeds the CTPP flow for an interchange by at least 20 trips. These maps allow closer examination of directional bias of flows from an origin or to a destination. For instance, for the urbanized area tract 846502 – DuPage Co. (top right map) the CTPP draws more workers from the east-northeast direction than the LED. For the suburban tract 804603 Cook Co. (bottom right map) the larger LED flows from the south/southwest indicate an area for further investigation.

The CTPP flows, in general, will exceed the LED flows because, as noted previously, the CTPP has nearly a million more work trips. Where LED flows exceed CTPP flows indicates a distinct difference in the directional distribution of trips. For these four tracts, directional distributions of LED and CTPP flows are presented in Table 14. These distributions are calculated using only interchanges that are 150 miles or less from the tract of interest. In addition, directional distributions are provided after low flow interchanges (less than 10 travelers) are eliminated. Sectors where the absolute differences between the LED and CTPP proportion exceed 2.5% are shaded. As indicted in the maps in Figure 6 and this table, the directional distribution of flows from these datasets is quite comparable. Only tract 804603 has three sectors where the differences exceed 2.5%. In this tract, the LED has a slightly higher proportion of flows from the southwest to the southeast and the CTPP has a slightly higher proportion of flows from the southeast to the east.

Earnings/ Wage Data

Comparing wage data from the LED dataset to information in the ACS or the decennial census is not straightforward for a variety of reasons. First, the LED data contain the proportion of workers in three wage categories for each block and the total number of workers in each block. Each of the wage categories is assigned a percent carried to the first decimal place (e.g. .343 = 34.3%). These data are in the Home Based data file and, as mentioned, total workers in this file do not agree with total workers from the OD dataset. In addition, the categories used in the LED dataset to define low, medium, and high wages are inconsistent with summary tables provided by the Census Bureau. The closest table to compare from the Decennial is P84. (Sex by earnings in 1999 for the population 16 years and over with earnings [43] - Universe: Population 16 years and over with earnings) and table P111 from the ACS (Sex by earnings (in 2002 inflation-adjusted dollars) for the population 16 years and over with earnings in the past 12 months - Universe: Population 16 Years And Over With Earnings). Earnings are defined by the Census Bureau as follows:

Proportion of workers who live in a sector and commute to this tract. (Only flows less than 150 miles)				
Cook Co. Destination Tract 804603				
SECTOR	**LED**	**CTPP**	**LED (Flows ≥ 10)**	**CTPP (Flows ≥ 10)**
E-NE	7.3	7.8	7.9	7.9
NE-N	7.8	7.8	8.5	7.9
N-NW	5.0	5.4	5.6	5.5
NW-W	11.6	11.9	12.5	12.0
W-SW	17.3	19.3	18.9	19.7
SW-S	16.1	12.7	17.7	12.9
S-SE	14.1	11.2	13.9	11.0
SE-E	20.9	23.9	15.0	23.1
Cook Co. Destination Tract 081400				
E-NE	0.0	0.0	0.0	0.0
NE-N	0.0	0.0	0.0	0.0
N-NW	28.0	29.8	30.7	30.2
NW-W	29.5	29.0	29.3	28.8
W-SW	18.1	16.9	16.2	16.8
SW-S	16.1	15.6	15.3	15.4
S-SE	8.3	8.8	8.5	8.8
SE-E	0.0	0.0	0.0	0.0
Stephenson Co. Destination Tract 000800				
E-NE	9.2	10.2	8.2	10.2
NE-N	0.0	0.0	0.0	0.0
N-NW	3.2	4.2	3.8	4.3
NW-W	20.9	21.8	24.5	22.1
W-SW	28.0	29.7	33.1	30.0
SW-S	11.0	10.2	11.3	10.2
S-SE	11.4	11.9	12.6	11.9
SE-E	16.2	11.9	6.6	11.3
Proportion of workers who live in this tract and commute to these sectors. (Only flows less than 150 miles)				
DuPage Co. Origin Tract 846502				
SECTOR	**LED**	**CTPP**	**LED (Flows ≥ 10)**	**CTPP (Flows ≥ 10)**
E-NE	61.7	65.1	68.0	65.5
NE-N	12.9	11.9	11.5	11.9
N-NW	5.3	4.8	5.2	4.7
NW-W	7.0	6.4	6.5	6.5
W-SW	4.4	2.4	4.4	2.4
SW-S	1.1	0.5	0.0	0.5
S-SE	3.3	3.2	3.0	3.2
SE-E	4.3	5.6	1.5	5.3

Greater than 2.5% absolute difference

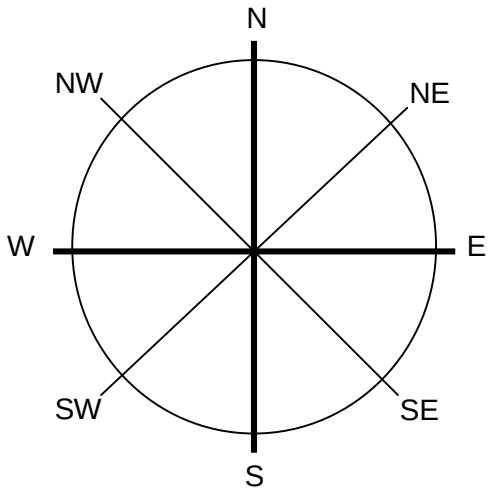

Table 14: Directional distribution of flows.

> Earnings are defined as the algebraic sum of wage or salary income and net income from self-employment. Earnings represent the amount of income received regularly before deductions for personal income taxes, Social Security, bond purchases, union dues, Medicare deductions, etc.

These tables contain the total number of workers 16 years in 20 earnings levels by two genders (or 40 classes). The LED data do not include income from self-employment or for those classes of workers (such as federal employees) discussed above. There is no way to accurately extract a proportion of workers from the Decennial or ACS tables that are not covered by the LED since there is no way to identify into which earnings range these workers fall. For example, if 15% of the total workers in a county are excluded from Table P051 or P058 because they represent categories of workers not covered in the LED, the only way to apply this proportion is uniformly across all 20 earnings ranges. Another difference in all three datasets is the reference year for the dollar value. The Decennial is 1999, the ACS is 2002, and the LED is the fourth quarter of 2001.

In addition, the LED data contain <u>average</u> (mean) wages for each block. Decennial census data (from table P85) contain <u>median</u> earnings at the tract level.

To the extent possible, data from the census tables were collapsed to provide estimates of the proportion of workers in similar low, medium, and high earnings categories by county. Differences in the category definitions used are:

Low: $\leq$ \$12,000 (LED), < \$12,500 (Decennial, ACS)
Medium: \$12,000 < x $\leq$ \$35,000 (LED), \$12,500 $\leq$ x < \$35,000 (Decennial, ACS)
High: > \$35,000 (LED), $\geq$ \$35,000 (Decennial, ACS)

There are two ways to calculate the proportion of workers in each wage category for each county from the LED dataset. Rounding errors result in slightly different values based on the method of calculation used. Method 1 is calculated as follows:

$$PW_{ic} = \sum_j \sum_k \sum_l W_l P_{lc} \; / \; \sum_j \sum_k \sum_l W_l$$

where

PW_{ic} = proportion of workers in county i, wage category c
W_l = total workers in block k
P_{lc} = percent workers in block l, wage category c
c = low, medium, or high wage category
k = block groups
j = tracts

Alternatively, method 2 using similar notation, is

$$PW_{ic} = \left[\sum_{j} \sum_{k} \sum_{l} P_{lc} \right] / (l+k+j)$$

Table 15 contains the results of the county calculations for the 9 counties included the ACS using the LED HB summary file and the two calculation methods described above.

	METHOD 1			METHOD 2		
County	Percent Low Earnings	Percent Medium Earnings	Percent High Earnings	Percent Low Earnings	Percent Medium Earnings	Percent High Earnings
Cook Co. IL	22.4%	36.0%	41.6%	22.2%	35.0%	42.9%
DuPage Co IL	16.6%	29.3%	54.1%	16.3%	28.3%	55.4%
Kane Co IL	20.1%	34.4%	45.5%	19.9%	34.5%	45.6%
Lake Co IL	17.9%	29.7%	52.4%	17.8%	30.0%	52.2%
McHenry Co IL	18.4%	31.0%	50.5%	18.9%	32.1%	49.1%
Madison Co IL	33.9%	37.5%	28.6%	34.4%	38.7%	26.9%
St Clair Co IL	38.2%	37.9%	24.0%	38.9%	38.7%	22.4%
Will Co IL	19.6%	31.6%	48.8%	20.2%	32.4%	47.3%
Winnebago Co IL	26.1%	39.7%	34.2%	26.8%	40.3%	32.9%

Table 15: LED Proportion of workers in wage categories.

Tables 16, 17 and 18 contain the proportion of workers in each wage category estimated using ACS table P111 from 2001 and 2002 and the Decennial table P84. The lower and upper bounds for the 90% confidence interval of these estimates are also provided in these tables. The proportion of workers in each wage category for the Decennial and ACS 2001 datasets is not significantly different in 20 (74%) out of 27 cases (3 wage categories by 9 counties). This improves to 22 (81%) out of 27 comparing the Decennial to the ACS 2002 data.

Despite the differences in these three datasets, the LED distribution of workers to wage categories compares favorably with the ACS estimates. Of the 27 possible comparisons in estimates, the LED and ACS 2001 estimates are not significantly different in 14 (52%) using Method 1 and 12 (44%) using Method 2. This improves to 19 (70%) using Method 1 and 18 (67%) using Method 2 when comparing the LED to ACS 2002. In contrast, the LED estimates are significantly different from the Decennial estimates in all but 4 (15%) cases using Method 1 and or 6 (22%) using Method 2. As with total worker estimates, the LED wage data compares more favorably with the ACS than the Decennial data.

ACS Table P111 2001	Population 16 years and over with earnings: Percent Low Earnings	*Lower Bound*	*Upper Bound*	Percent Medium Earnings	*Lower Bound*	*Upper Bound*	Percent High Earnings	*Lower Bound*	*Upper Bound*
Cook Co. IL	23.5	*22.6*	*24.4*	36.0	*34.8*	*37.1*	40.5	*39.5*	*41.6*
DuPage Co IL	21.6	*19.9*	*23.3*	28.3	*26.1*	*30.5*	50.1	*47.6*	*52.6*
Kane Co IL	26.6	*23.5*	*29.7*	35.7	*32.2*	*39.2*	37.7	*34.5*	*40.9*
Lake Co IL	20.8	*19.9*	*21.8*	30.4	*29.1*	*31.7*	48.8	*47.4*	*50.2*
McHenry Co IL	21.1	*18.0*	*24.1*	30.4	*26.6*	*34.3*	48.5	*43.8*	*53.2*
Madison Co IL	31.5	*27.8*	*35.2*	32.7	*29.0*	*36.5*	35.7	*31.8*	*39.7*
St Clair Co IL	26.8	*23.2*	*30.3*	42.4	*37.2*	*47.7*	30.8	*26.7*	*34.9*
Will Co IL	24.1	*21.3*	*26.8*	30.2	*27.3*	*33.0*	45.8	*42.5*	*49.1*
Winnebago Co IL	27.4	*23.7*	*31.2*	34.7	*30.5*	*38.9*	37.9	*33.8*	*41.9*

Table 16: ACS 2001 Proportion of Workers in Wage Categories

ACS Table P111 2002	Population 16 years and over with earnings: Percent Low Earnings	*Lower Bound*	*Upper Bound*	Percent Medium Earnings	*Lower Bound*	*Upper Bound*	Percent High Earnings	*Lower Bound*	*Upper Bound*
Cook Co. IL	23.6	*22.1*	*25.1*	34.0	*32.4*	*35.7*	42.4	*40.6*	*44.1*
DuPage Co IL	21.5	*18.8*	*24.1*	28.4	*25.4*	*31.3*	50.1	*46.8*	*53.5*
Kane Co IL	23.5	*19.1*	*27.9*	33.5	*28.2*	*38.9*	42.9	*37.9*	*48.0*
Lake Co IL	20.8	*19.1*	*22.4*	27.6	*25.8*	*29.4*	51.7	*49.6*	*53.8*
McHenry Co IL	22.1	*17.7*	*26.6*	31.7	*26.6*	*36.8*	46.1	*40.5*	*51.7*
Madison Co IL	29.8	*24.6*	*35.0*	35.4	*29.8*	*41.0*	34.7	*29.3*	*40.2*
St Clair Co IL	32.2	*25.8*	*38.6*	38.2	*31.3*	*45.0*	29.6	*23.6*	*35.6*
Will Co IL	23.9	*20.6*	*27.3*	27.4	*23.7*	*31.2*	48.6	*44.3*	*52.9*
Winnebago Co IL	23.0	*18.5*	*27.6*	42.5	*36.3*	*48.6*	34.5	*29.2*	*39.8*

Table 17: ACS 2002 Proportion of Workers in Wage Categories

Key to Tables 16 and 17.

- LED not in ACS CI regardless of calculation
- LED in ACS CI regardless of calculation
- LED in ACS CI Method 1 calculation only
- LED in ACS CI Method 2 calculation only

Decennial Table
P84 Population 16 years and over with earnings:

	Percent Low Earnings	*Lower Bound*	*Upper Bound*	Percent Medium Earnings	*Lower Bound*	*Upper Bound*	Percent High Earnings	*Lower Bound*	*Upper Bound*
Cook Co. IL	24.7	*24.5*	*24.9*	37.9	*37.6*	*38.1*	37.4	*37.2*	*37.7*
DuPage Co IL	20.9	*20.5*	*21.3*	30.0	*29.5*	*30.5*	49.2	*48.5*	*49.8*
Kane Co IL	23.2	*22.6*	*23.8*	36.6	*35.8*	*37.4*	40.2	*39.4*	*41.0*
Lake Co IL	22.6	*22.2*	*23.1*	32.3	*31.7*	*32.8*	45.1	*44.4*	*45.8*
McHenry Co IL	22.2	*21.5*	*22.9*	32.2	*31.3*	*33.0*	45.6	*44.6*	*46.7*
Madison Co IL	29.6	*28.7*	*30.5*	38.3	*37.3*	*39.3*	32.1	*31.2*	*33.0*
St Clair Co IL	29.1	*28.2*	*30.0*	41.6	*40.5*	*42.6*	29.3	*28.4*	*30.2*
Will Co IL	22.7	*22.2*	*23.3*	31.9	*31.3*	*32.6*	45.4	*44.6*	*46.1*
Winnebago Co IL	26.8	*26.0*	*27.6*	40.0	*39.1*	*41.0*	33.2	*32.3*	*34.0*

Table 18: Decennial Proportion of Workers in Wage Categories

Key to Table 18

LED not in DEC CI regardless of calculation
LED in DEC CI regardless of calculation
LED in DEC CI Method 1 calculation only
LED in DEC CI Method 2 calculation only

Industry Classification

The structure of the LED data files does not currently allow for meaningful comparison to the decennial and ACS data sets with regard to industry classification. The LED work place data set provides information on whether or not an industry falling into one of fourteen SIC classifications exists in a block. This information is useful to show, for example, where all industries of a certain classification are within a county or to identify tracts with a wide mix of industries.

Urban Part II (Table 2-3) of the CTPP provides information on the number of workers by place of work in the following industry categories:

Agriculture, Forestry and Fisheries
Mining
Construction
Manufacturing, non-durable goods
Manufacturing, durable goods
Transportation
Communication and other public utilities
Wholesale trade
Retail trade
Finance, Insurance, and Real Estate
Business and repair services

Personal Services
Entertainment and recreation services
Health services
Educational services
Other professional and related services
Public Administration
Armed Forces

Since LED does not provide the number of workers in each industry classification, these data were not analyzed.

CONCLUSIONS

Comparison of the LED administrative data products to the census and the ACS allows analysts to troubleshoot the datasets and improve the data products. The LED data appear to provide generally reliable information to transportation planners on home-to-work flows as well as earnings and pay. The most difficult aspect of comparing the data sets results from differences in coverage of employees and their earnings and different approaches to summarizing the data. Users should be cautioned when using the LED data about who is and is not included in the estimates.

An advantage of Census and survey based data is that confidence intervals for the estimates may be determined. An improvement to the LED data would be to add uncertainty measures associated with the estimates. In addition, spatial analysis of the LED products using techniques described above will aid in cleaning the ES202 employment location data used to generate the origin-destination tables.

At this point it is not possible to determine which of these data sets most closely models reality. It is reassuring to know that all of the datasets are telling similar stories. Given the benefits to using administrative data, including;

- lower costs for data collection and processing,
- more frequently available (quarterly or yearly based on LED program),
- avoidance of survey related issues such as non-response,
- and administrative data does not rely on people reporting of complex data such as income and earnings,

the LED data sets represent a potentially invaluable resource to transportation planners. An important drawback, however, is that the LED does not have the richness of information, such as mode choice, travel time, etc., provided by the ACS and Decennial long form. In summary, with some simple restructuring of the LED output, enhancements to the initial data cleaning process, and the addition of error estimates, the LED data could be a useful quarterly or annual source of work travel flow data for transportation planners.

ACKNOWLEDGEMENTS:

This research was indirectly supported by a grant from the Bureau of Transportation Statistics to the LED Program at the US Census Bureau. Special thanks to Dr. Julia Lane for providing an opportunity to work with the LED staff and data products for this pilot study. Also, special thanks to Dr. Philip Fulton reviewing this work and providing useful suggestions to improve this paper.

REFERENCES

Abowd, John M., John C. Haltiwanger, and Julia I. Lane. Integrated Longitudinal Employee-Employer Data for the United States. U.S. Bureau of the Census, Longitudinal Employer-Household Dynamics Technical Paper TP-2004-2, Suitland, Maryland, May 2004.

Abowd, John M., Paul A.Lengermann, and Lars Vilhuber, The Creation of the Employment Dynamics Estimates. U.S. Bureau of the Census, Longitudinal Employer-Household Dynamics Technical Paper TP-2002-13, Suitland, Maryland, July 2002.

Farmer, Tracy E. and Michael A. Searson. Use of Administrative Records in the Bureau of Labor Statistics' Covered Employment and Wages (ES-202) Program. Paper presented at the 1995 Bureau of the Census Annual Research Conference.

Mix, Wende A. A Comparison of Journey-to-Work Data in the 2000 Decennial Census and the American Community Survey. Westat Corporation, Rockville, MD, December 2003.

Souleyrette, Reginald R., David J. Plazak, Tim R. Strauss, and Stephen J. Andrle, Applications of State Employment Data to Transportation Planning. In *Transportation Research Record No 1768*, TRB, National Research Council, Washington D.C., 2001, pp. 26-35.

Stevens, David W. Employment That is Not Covered by State Unemployment Insurance Laws. U.S. Bureau of the Census, Longitudinal Employer-Household Dynamics Technical Paper TP-2002-16, Suitland, Maryland, January 2002.

Roemer, Marc. Origin-Destination Matrix and Block Characteristics Files: Prototype Developed for the Bureau of Transportation Statistics, U.S. Bureau of the Census, Longitudinal Employer-Household Dynamics Technical Paper TP-2003-07, Suitland, Maryland, September 2003.