Agricultural Trip Generation - Linking Spatial Data and Travel Demand Modeling using SAS®

Alan Dybing, North Dakota State University – UGPTI, Fargo, ND

ABSTRACT

Data linkages from GIS and Cube Voyager require advanced knowledge of these software. The four-step travel demand modeling procedure (TDM) is commonly used to estimate and forecast traffic volumes for use in transportation planning. The trip generation step of the four-step model seeks to estimate the trip attractions and productions representing individual trips originating or terminating with a geographic boundary. Freight trip generation in rural areas within the great plains primarily result from agricultural production and marketing. This paper outlines a procedure for utilizing satellite imagery to estimate agricultural truck trip production at the township level linking ArcGIS and SAS®® to generate trip generation tables for use for TDM purposes. The National Agricultural Statistics Service (NASS) Cropland Data Layer (CDL) is a data sourced produced using a combination of satellite imagery, ground-truth surveys and data mining techniques resulting in a digital raster map providing land use type at a 30-meter (0.25 acre) resolution. At the township level, the raster data was converted to polygons and acreages by crop type were calculated using ArcGIS and output to a shapefile. The database file was imported for SAS® and using a combination of NASS county-level crop yield and fertilizer usage rate estimates, the total truck trips resulting from agricultural production activities were estimated by township. SAS® was used to convert these estimate to a specific file format utilized by Cube Voyager for use in development of a truck TDM.

INTRODUCTION

Travel demand models are used to provide estimates of current and future roadway traffic and are used for transportation planning. Conventional travel demand models consist of four steps: trip generation, trip distribution, mode choice, and traffic assignment. The trip generation step estimates the numbers of trips produced or attracted to a geographic area. The trip distribution step often utilizes a gravity model to pair potential origins and destinations for these trips. The mode choice estimates the number of trips by individual mode, and the traffic assignment step estimates the set of paths that are used to connect the origins and destinations. The result of these four steps are forecasted traffic volumes on individual segments of a network.

One use of travel demand models is to estimate and forecast the traffic volume over a set of highways for use in modeling roadway deterioration and timing of future improvements. A key factor in modeling pavement deterioration is the number of equivalent single axle loads (ESALs) over time. The ESAL is a standardized estimate of pavement damage resulting from a single 18,000-pound axle. To calculate ESALs for a particular highway segment, the number and configuration of truck trips is required.

The processes outlined in this paper were used as a part of a multi-part model for forecasting highway funding needs for rural highways in North Dakota. Three activities generate significant traffic on rural North Dakota highways: agricultural production, crude oil production and extraction, and manufacturing.

DATA SOURCES

To forecast trips generated by agricultural production in North Dakota, three data items are required: geographic boundary designations, quantifiable activity levels, and trip generation rates.

TRANSPORTATION ANALYSIS ZONES

Since the trip generation step is used to estimate trips generated within a given geographic boundary, it needs to be specified based upon the required level of aggregation dictated by the travel demand model. The size of these transportation analysis zones (TAZ) must balance the level of detail against software limitations and data size. As an example, the use of county boundaries for TAZ would be too aggregate to provide meaningful traffic forecasts for individual highways. Conversely, using sections (one square mile) as TAZ would provide great detail, but would be difficult to effectively use in ArcGIS and Cube software due to data size.

The base unit of production used in the agricultural model is the township, or county subdivision; this is typically a 6-mile by 6-mile area. Township shapefiles were obtained from the North Dakota GIS Hub, a source of public GIS data maintained by the State of North Dakota. However, organized townships do not exist in all North Dakota counties. Townships were selected for use as a geographic and not an organizational boundary. Where unorganized townships exist, a placeholder boundary was created to represent a geographic area similar in size to a township. A unique numeric identifier attribute is added for use in merging township shapefile data.

AGRICULTURAL PRODUCTION ACREAGE

The agricultural production activity levels in each township are estimated from satellite imagery. The National Agricultural Statistics Service (NASS) Crop Data Layer (CDL) is a satellite image of land use in North Dakota, with individual crop types represented by different colors as shown in Figure 1. Each pixel of the image represents a 30 meter by 30-meter area. Used in conjunction with ESRI ArcGIS, the CDL provides data regarding the total number of acres of each crop produced in each county subdivision.
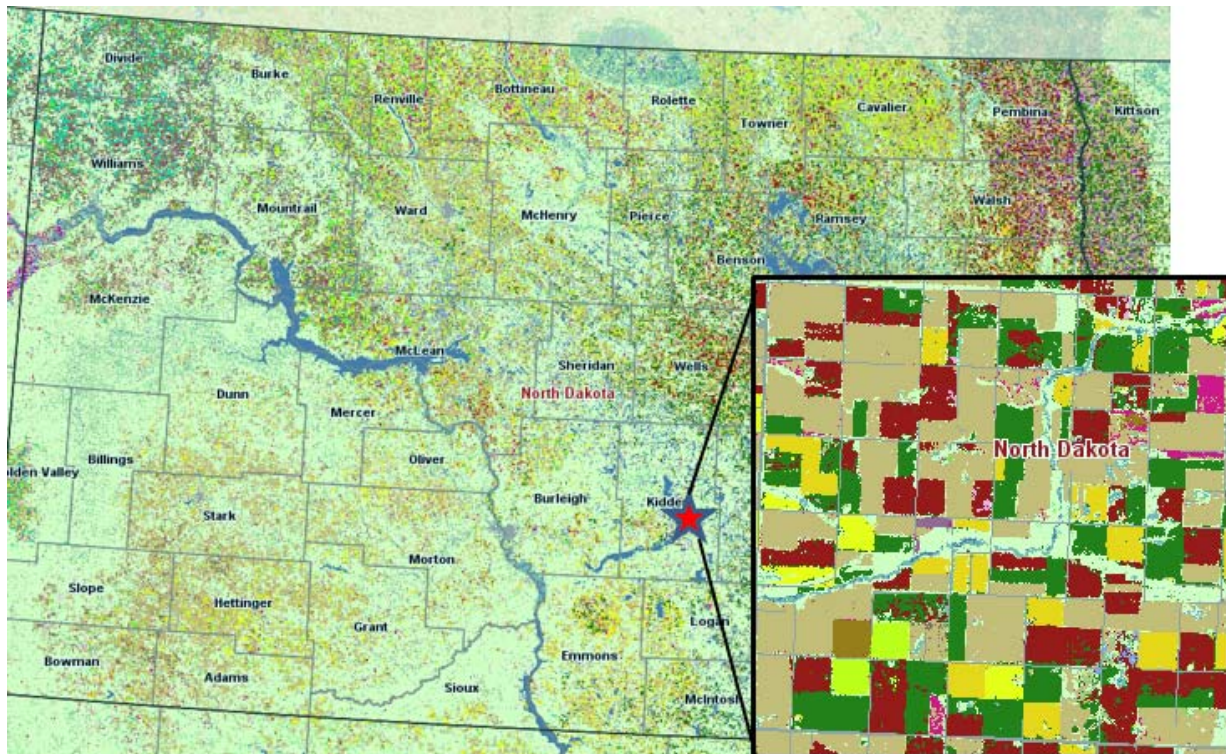

Figure 1. Sample Township Coverage of Crop Data Layer

CROP PRODUCTION DATA

Crop production data by county was obtained from the NASS website. This data provides the number of acres planted and harvested, as well as yields and total production by county, crop, and production practice. Data as to yield and production is not available at the township level, so it is assumed that townships within a county have similar agricultural output.

CDL DATA EXTRACTION PROCEDURE

The CDL in its original format is a raster-based GeoTIFF file. Without conversion, this image file does not readily contain quantifiable data required to estimate truck trips generated. Figure 2 outlines the process of the CDL conversion within ArcGIS.
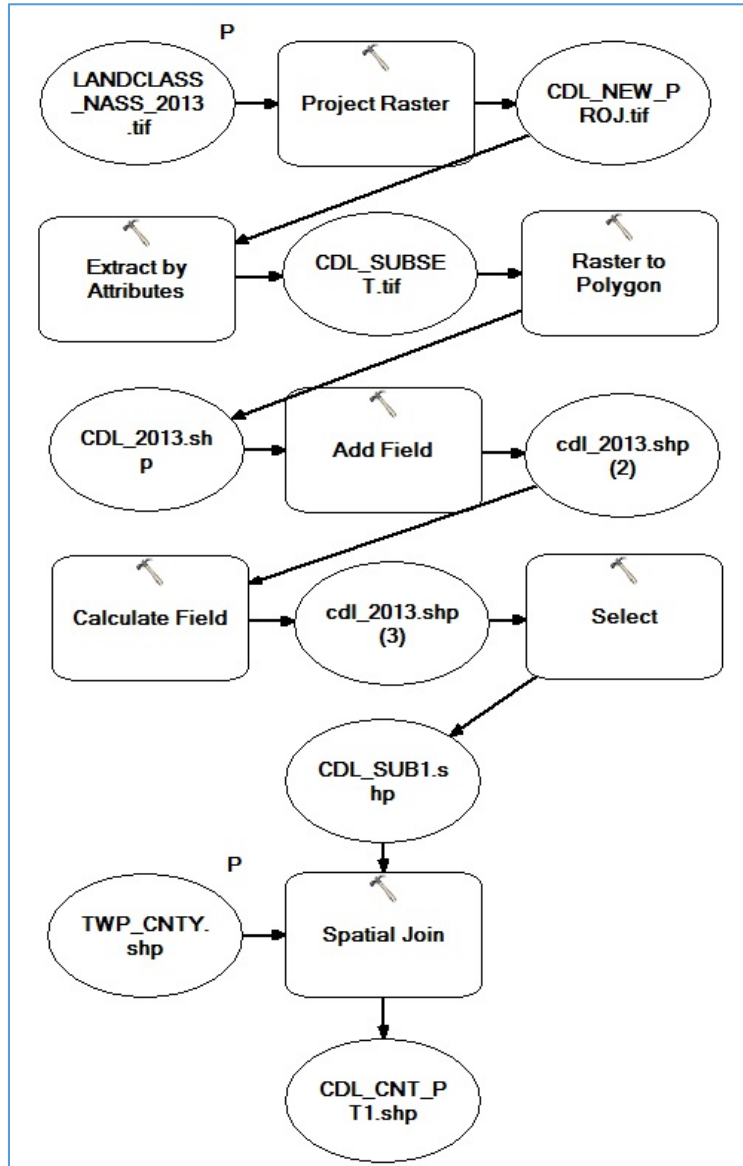

Figure 2. Model Builder Process Map

STEPS IN THE CDL DATA CONVERSION PROCESS

1. Project CDL – To ensure consistency between the CDL and the boundary files, the CDL image is projected to the same coordinate system as the boundary shapefile.

2. Select by Attributes – Since the objective of this procedure is to estimate truck trips generated by agricultural production, only a subset of the land uses is required. The CDL contains many possible land types including non-productive land such as wetlands, urban developments, prairie land, and water. Since not all of these land uses will be utilized in the trip generation step, a subset of the top ten crops produced in North Dakota by acreage are selected. This is done prior to conversion to polygon data to reduce the resulting file sizes.

3. Raster to Polygon – In order to perform spatial calculations, the raster image is converted to polgyons using the Raster to Polygon tool in Arc Toolbox. This produces a layer of polygons for each group of pixels representing crop acreage within the CDL.

4. Add Attribute – An acreage attribute is added for use in calculating acres of each commodity in step 5.

5. Calculate – Acreage is calculated by geometry in ArcGIS. This procedure calculates the acreage of each polygon in the crop data layer based upon the extent and projection specified in step 1 of this procedure. These acreage estimates will later be used to calculate total production of each commodity based upon the reported commodity yields from the year that the CDL was produced.

6. Spatial Join – Now that the estimates of the area of each polygon have been calculated, the CDL is joined with the township shapefile based upon spatial location. The resulting shapefile contains polygons within each township boundary with acreage estimates for each polygon along with commodity identification information.

7. Spatial Join – The final step in the process is to join the township file with estimated commodity acreages with the county shapefile. The county-level yield and production estimates produced by NASS require a county identifier with each township in order to calculate production at the township level.

The shapefile produced by the conversion process contains the individual acreage estimates by commodity for each polygon inside of the township boundaries. Figure 3 shows a subset of the attribute table produced by the CDL data extraction process. In this example, acreages calculated for each polygon in Arne Township, Benson County North Dakota are shown. These acreages will later be summed to the township level for estimating truck trips produced within the township.

CDL_TWP

FID	Shape	Join_Count	TARGET_FID	ID	GRIDCODE	CLASS_NAME	Acres	count	name	unique
21176	Polygon ZM	1	2117601	2117602	11	Dry Beans	0.152187	601	Arne	TOWN613
21176	Polygon ZM	1	2117602	2117603	2	Soybeans	0.142417	601	Arne	TOWN613
21176	Polygon ZM	1	2117603	2117604	3	Barley	0.142472	601	Arne	TOWN613
21176	Polygon ZM	1	2117604	2117605	3	Barley	0.752204	601	Arne	TOWN613
21176	Polygon ZM	1	2117605	2117606	6	Winter Wheat	0.320181	601	Arne	TOWN613
21176	Polygon ZM	1	2117606	2117607	8	Canola	0.222395	601	Arne	TOWN613
21176	Polygon ZM	1	2117607	2117608	5	Spring Wheat	0.152187	601	Arne	TOWN613
21176	Polygon ZM	1	2117608	2117609	5	Spring Wheat	0.142417	601	Arne	TOWN613
21176	Polygon ZM	1	2117609	2117610	3	Barley	0.44479	601	Arne	TOWN613
21176	Polygon ZM	1	2117610	2117611	2	Soybeans	0.222395	601	Arne	TOWN613
21176	Polygon ZM	1	2117611	2117612	2	Soybeans	0.142472	601	Arne	TOWN613
21176	Polygon ZM	1	2117612	2117613	2	Soybeans	0.152187	601	Arne	TOWN613
21176	Polygon ZM	1	2117613	2117614	11	Dry Beans	0.142417	601	Arne	TOWN613
21176	Polygon ZM	1	2117614	2117615	2	Soybeans	5.667977	601	Arne	TOWN613
21176	Polygon ZM	1	2117615	2117616	1	Corn	0.152187	601	Arne	TOWN613
21176	Polygon ZM	1	2117616	2117617	2	Soybeans	1.556764	601	Arne	TOWN613
21176	Polygon ZM	1	2117617	2117618	2	Soybeans	0.142417	601	Arne	TOWN613
21176	Polygon ZM	1	2117618	2117619	1	Corn	1.541441	601	Arne	TOWN613
21180	Polygon ZM	1	2118033	2118034	2	Soybeans	0.142472	601	Arne	TOWN613
21180	Polygon ZM	1	2118034	2118035	2	Soybeans	0.222395	601	Arne	TOWN613
21180	Polygon ZM	1	2118035	2118036	5	Spring Wheat	21.13602	601	Arne	TOWN613
21180	Polygon ZM	1	2118036	2118037	5	Spring Wheat	0.222395	601	Arne	TOWN613
21180	Polygon ZM	1	2118037	2118038	5	Spring Wheat	0.222395	601	Arne	TOWN613
21180	Polygon ZM	1	2118038	2118039	7	Oats	0.142472	601	Arne	TOWN613
21180	Polygon ZM	1	2118039	2118040	2	Soybeans	0.44479	601	Arne	TOWN613
21180	Polygon ZM	1	2118040	2118041	5	Spring Wheat	0.222395	601	Arne	TOWN613
21180	Polygon ZM	1	2118041	2118042	1	Corn	0.44479	601	Arne	TOWN613
21180	Polygon ZM	1	2118042	2118043	11	Dry Beans	0.142472	601	Arne	TOWN613
21180	Polygon ZM	1	2118043	2118044	11	Dry Beans	0.300191	601	Arne	TOWN613
21180	Polygon ZM	1	2118044	2118045	11	Dry Beans	0.152187	601	Arne	TOWN613
21180	Polygon ZM	1	2118045	2118046	11	Dry Beans	0.667185	601	Arne	TOWN613
21180	Polygon ZM	1	2118046	2118047	3	Barley	0.152187	601	Arne	TOWN613
21180	Polygon ZM	1	2118047	2118048	5	Spring Wheat	0.667185	601	Arne	TOWN613
21180	Polygon ZM	1	2118048	2118049	5	Spring Wheat	0.222395	601	Arne	TOWN613
21180	Polygon ZM	1	2118049	2118050	11	Dry Beans	0.152187	601	Arne	TOWN613
21180	Polygon ZM	1	2118050	2118051	1	Corn	0.142417	601	Arne	TOWN613
21180	Polygon ZM	1	2118051	2118052	5	Spring Wheat	0.667185	601	Arne	TOWN613
21180	Polygon ZM	1	2118052	2118053	1	Corn	0.222395	601	Arne	TOWN613

Figure 3. Sample Data from the CDL Data Extraction Process

GIS Shapefiles Consist of five individual files as shown in Figure 4. The .dbf file contains the attribute data, a sample of which is shown in Figure 3. The .prj file stores the coordinate system of the shapefile. This is used by ArcGIS to correctly display the data. The .shp file contains the feature geometry, and the .shx file contains the index of the feature geometry. The .shp.xml file contains geospatial metadata in XML format.

Figure 4. Files Include in CDL-Township Shapefile

ESTIMATION OF TOWNSHIP LEVEL TRUCK TRIP GENERATION

As mentioned above, the DBF file contains the attribute data including the selected commodity acreage estimates within the CDL. This file is imported into SAS® for further data processing. Similar data steps and calculations can be done within the ArcGIS infrastructure. However, working with database components of GIS shapefiles in SAS® is an efficient method of handling large data outside the limitations of ArcGIS.

The database file is effectively the attribute table stored in a separate file from the geographic components of the shapefile. Attributes can be added, deleted, modified or otherwise changed in SAS® to generate a new or updated database file which can be read directly into ArcGIS for display purposes. If the original database file is replaced following manipulation of the attributes steps should be taken to ensure that the observations are in the same order as the original file to avoid attribute display errors in ArcGIS. The addition of a count variable prior to sorting of the data and sorting the data set by the count variable prior to export will ensure that the observations are returned to the original order.

DATA VERIFICATION

Since the acreage estimates calculated from the CDL data extraction procedure are just that – estimates, they are compared to the reported NASS county level acreage data. As shown in figure 3, the shapefile attribute table contains the acreages for individual polygons, not county-level estimates. The PROC IMPORT procedure is used to import the DBF file into SAS®. The data is sorted by county and the PROC SUMMARY procedure of the acreage variable by county is used to calculate the estimated acres by county produced by the CDL data extraction procedure. Example code for this step is:

```
proc sort data=CDL_TWP;
   by County Commodity;
run;

proc summary data=CDL_TWP;
   var Acres;
   by County Commodity;
   output out=Acre_Check sum=;
run;
```

This data is compared against the NASS county-level acreage data. Two measures of acreage estimates are given in the NASS data: Acres Planted and Acres Harvested. The acres planted data is consistently higher than the acres harvested data. The reason for this is due to weather impacts and other production related impacts which may result in acres not being harvested. The production data is based upon the acres that were actually harvested during the calendar year. The CDL image is taken during the production season, that is – prior to harvest. To assess the accuracy of the CDL data extraction procedure the sum of the acres by commodity from the CDL is compared to the NASS reported planted acres. The ratio of the acres harvested to planted is calculated for each county to index the CDL acres when calculating the truck trips produced for each township.

PRODUCTION ESTIMATES

The required data items for the trip generation files are: TAZ identifier and trucks generated by each individual commodity. The initial GIS output shapefiles contain the TAZ identifier and the acreage by commodity. County level yields from the NASS website are used to convert acreages to bushels or pounds of commodities. The bushel estimates are converted to pounds using the conversion factors shown in Table 1.

Commodity	Conversion Factor
Hard Red Spring/Winter Wheat	60 pounds/bushel
Durum Wheat	60 pounds/bushel
Barley	48 pounds/bushel
Sunflowers	32 pounds/bushel
Corn	56 pounds/bushel
Flax	56 pounds/bushel
Oats	32 pounds/bushel
Soybeans	60 pounds/bushel
Canola	60 pounds/bushel
Dry Edible Beans	60 pounds/bushel

Table 1. Commodity Conversion Factors

The NASS production and yield data contain the average production per acre and the total county production for the calendar year by county within the state of North Dakota. Since sub-county yield and production estimates are not available, this data is used to calculate township production, and it is assumed that county production estimates are reflective of the township-level production within the county.

The PROC IMPORT procedure is used to import the NASS county-level production files. Both the county-level production data and the CDL/TWP acreage datasets are sorted by county and commodity. These data sets are merged by county name and commodity name. Commodity production is calculated by multiplying the number of acres by the county average yield:

```
proc sort data=NASS_PROD;
   by County Commodity;
run;

data CDL_TWP_VOL;
   merge CDL_TWP NASS_PROD;
   by County Commodity;
   Bushels=Acres*Yield;
run;
```

ESTIMATES OF TRUCK TRIPS GENERATED

Now that the township level estimates of production by commodity are calculated, they are converted to equivalent trucks. Each commodity has a unique density. As shown in Table 1, wheat has a density of 60 lbs./bushel and sunflowers a density of 32 lbs./bushel. The density of the crop being hauled impacts the number of bushels per truck due to truck size and weight restrictions.

The legal payload of a typical 5-axle semi with a grain trailer is 50,000 lbs. with a gross vehicle weight of 80,000 lbs. The 80,000 lb. load limit is typical for local roads, although certain areas allow for legal loads of 105,500 lbs., particularly state highways. Often, county and township roads are limited to 80,000 lbs. The cubic capacity of a typical grain trailer is 1,400 cubic feet. A bushel is 1.22 cubic feet, so a typical grain trailer has a maximum capacity of roughly 1,150 bushels.

Table 2 shows the weight and payload of selected commodities. As stated above, truck loads are either constrained by weight (50,000 lbs.) or cubic capacity (1,150 bushels). Given these two constraints, the maximum density per bushel that will not be constrained by weight is 43.47 lbs./bushel (50,000 lbs./1,150 bushels). As shown in table 2, using the cubic capacity of 1,150 bushels, all commodities except

sunflowers and oats exceed the legal 50,000 lb. payload. These two commodities will then be restricted by cubic capacity and have a truck capacity of 1,150 bushels. The commodities where the weight at 1,150 bushels exceeds 50,000 lbs. are weight constrained and the payload is calculated by dividing 50,000.

Commodity	Weight at 1,150 Bu. (pounds)	Constrained Payload (bushels)
Hard Red Spring/Winter Wheat	68,852	833
Durum Wheat	68,852	833
Barley	55,082	1,041
Sunflowers	36,721	1,150
Corn	64,262	892
Flax	64,262	892
Oats	36,721	1,150
Soybeans	68,852	833
Canola	68,852	833
Dry Edible Beans	68,852	833

Table 2. Truck Payload Estimates by Commodity

By multiplying the constrained payloads by the bushels of each commodity produced within a township the number of trucks estimated to be generated by agricultural production are calculated. Summation of the number of trucks by township result in the total trucks generated by production in each township. These estimates are used directly as an input to the travel demand model. Example code for these steps are:

```
proc sort data=CDL_TWP_VOL;
    by Commodity;
run;

proc sort data=LBS_BU;
    by Commodity;
run

data CDL_TWP_TK;
    merge CDL_TWP_VOL LBS_BU;
    by Commodity;
    if LB_PER_BU<43.47 then do;
        Trucks=Bushels/1150;
      end;
    else Trucks=Bushels*LB_PER_BU/50000;
run;

proc sort data=CDL_TWP_TK;
    by Township;
run;

proc summary data=CDL_TWP_TK;
    var Trucks;
    by Township;
    output out=TWP_TRUCK sum=;
run;
```

TRIP PRODUCTION AND ATTRACTION

The trip generation estimates in this paper were developed for use in Cube Voyager travel demand software. Typically, the trip generation step occurs within the travel demand software based upon generation rates by population and demographics. However, in this case, since additional data was available as to the type and scope of individual crops produced, the procedure resulted in improvements over the traditional trip generation step calculations.

The trips generated at the township level are trip productions, that is trips generated that originate at the township level. Potential destinations are also required, using alternate data sources. For grains, these include elevators, feedlots, and processors such as flour mills and ethanol facilities. These locations represent the trip attractions. Since these activities occur at distinct points, modeling of these activities are based upon annual throughput or processing capacity.

SUMMARY

This paper outlines a procedure to develop township-level truck trip generations for use in travel demand modeling software using publically available data. The CDL is used to estimate commodity specific acreage estimates at the township level. NASS yield data is used to calculate township-level production given acreage and yield estimates. Based upon the commodity density and payload capacity of grain trucks the production is converted to equivalent trucks. The resulting truck trips generated by agricultural production in ready for use in travel demand software.

CONTACT INFORMATION

Alan Dybing
Upper Great Plains Transportation Institute – North Dakota State University
701.231.5988
alan.dybing@ndsu.edu
www.ugpti.org

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are registered trademarks or trademarks of their respective companies.