Q. Let's briefly discuss model inputs. How important are they to the modeling process?

A. Without question, input choices are a critical step in any modeling process and can profoundly affect the costs that are developed. The inputs used in the modeling process influence the models' outputs, and some recent studies have indicated that differences in input values are sometimes the dominant reason for differences in outputs. For example, in a comparison of the Hatfield and BCPM models, Christensen Associates found that relatively few input items accounted for the majority of the difference in model outputs. [*Economic Evaluation of Proxy Cost Models for Determining Universal Service Support*, January 9, 1997]. Christensen concluded that ..."the surest path to a model that will be satisfactory to the Joint Board and the FCC is through a process that will focus on establishing a few key specifications that drive the proxy cost model results." [Id., p. 4]

Similar conclusions were reached in a Utah study that examined results from three models using similar input assumptions: Hatfield 2.2, the Cost Proxy Model (developed by Pacific Bell), and BCM2. Hatfield 2.2 was a predecessor of HAI, and the other two models were predecessors of BCPM:

> In sum, the three models yield estimates of the average monthly cost of an unbundled Utah loop that all fall within a very narrow range ($3). Hence, it appears that the models may be reconstructing the local network in a cost-comparable manner even though they employ different methodologies. Furthermore, it may suggest that what distinguishes one model from the others, in practice, are the values of the user-defined assumptions employed rather than inherent differences in the hardwired network architecture.

[Kevin T. Duffy-Deno, et al., *A Comparative Analysis of Loop Cost Proxy Models*, Preliminary Draft #2, Utah Division of Public Utilities, October 18, 1996, p. 5]. The Utah study also developed elasticities for changes in placement costs and fill factors. It found significant changes in loop costs when assumptions for placement costs were changed for BCM2 and CPM. It also found significant changes in costs when fill factors were changed for CPM and Hatfield. [Id., Table 2, p. 4].

While these studies confirm the importance of inputs, the algorithms within the models are also important. Even if the inputs are synchronized to achieve relatively similar statewide average results, different models still tend to produce different estimates for individual wire centers. In fact, one model may identify a particular wire center as having unusually high costs, while another model will indicate the wire center has relatively low, or normal, costs. Discrepancies of this type typically relate to differences in the geographic, or customer location, aspects of the model–either differences in the geographic input data, or differences in the customer location and cable routing algorithms, or both.

Q. Are some inputs and algorithms more important than others?

A. Yes. There is an informal consensus among model builders that certain aspects of the modeling process are the major "cost drivers". Right at the top of any list of factors explaining why certain areas have unusually high costs will be customer location or dispersion, and customer demand (line counts). If customers are widely dispersed over large areas, loop lengths will be long and line counts will be low, relative to the size of the area served, and thus costs will tend to be relatively high. As I indicated earlier, if the model does not accurately locate customers, it will be unable to accurately estimate the amount of cable required to serve those customers, the location of high cost areas, or the degree to which these areas have higher than normal costs.

If a model generates loop lengths that differ greatly from the actual loop lengths present in the existing network, this is an indication of potential weaknesses in the model; if the discrepancy is substantial, the most likely explanation is modeling error–a sign that the model isn't adequately locating customers, and/or isn't realistically and accurately deploying cable to serve those customers. The root cause of the problem may be weaknesses in the model algorithms, or errors in the input data that is used to run the model.

Customer demand is also an important cost driver, since customer density determines the extent to which economies of scale can be exploited; in remote areas with very few customers, the fixed costs of the network are spread over relatively few lines, driving up the cost per line. Therefore it is preferable for cost studies to use line counts which closely approximate, or match, the actual number of lines in each area. A comparison of actual versus modeled lines can provide an indication of modeling accuracy. An exact match can be achieved if the model is adjusted, or "conformed," to match the actual number of lines in each wire center. Even when this adjustment is performed, a comparison of the model's line counts *before* they have been conformed to the actual line count data can provide a useful indication of modeling error.

The relative number of wire centers where the raw line counts are substantially over or under the actual line counts, (e.g., error rates in excess of 25% or 50%) provides a strong indicator of the modelers' level of success in identifying and locating customers. And since no model is perfect, these data can also indicate where significant modeling error may be occurring with regard to other issues, besides simply the number of customer lines. Of course, once the model line counts have been conformed to match actual network data, they no longer provide an indication of the accuracy of the model's ability to locate customers. Hence, such comparisons need to be performed using the underlying model data, prior to "matching" a specific set of actual line counts.

The interaction of customer location and demand level largely determines cost per loop, the single most important network cost element. Most high-cost areas are ones with a handful of customers sprinkled far from the wire center. It is obviously much more costly to serve 50 customers spread over miles of roads within a 50 square mile area than it is to serve 50 customers concentrated in a single city block.

Q. Why are loop lengths important?

A. The BCPM, HAI and FCC models all use the actual sites of wire center central offices as nodal points in the network. In other words, existing wire center locations are taken as givens. Then the model "builds" a feeder/distribution system out to the identified end-users. Ideally, the model's cable routes should mirror the available rights of way to a large extent; those rights of way are generally along the road system that connects the community's homes and businesses.

To the extent the average loop lengths or total route mileage of the modeled network substantially exceeds or falls short of the corresponding data for the actual network, the model's cost results are suspect, since in practice there are relatively few efficient choices available for routing cable from the wire center to customers. The actual network may reflect some inefficiencies which could be overcome in a forward looking model, and thus the modeled route miles could be somewhat lower than the actual route miles. However, any such inefficiencies are unlikely to be extreme, and thus if substantial discrepancies are observed in the total number of route miles within a wire center, this is likely to be an indication of modeling error, perhaps reflecting inaccurate input data. Similarly, where the modeled route miles greatly exceed the route miles of the incumbent LEC's actual network, this is most likely an indication of modeling error, which should be identified and corrected, to the extent feasible.

As with discrepancies in line counts, substantial discrepancies in loop lengths or route mileage comparisons are useful indicators of weaknesses, errors, or problems in the model. This type of comparative data can help the Commission and the parties evaluate the model, and identify areas where better input data, or other refinements would be beneficial. For instance, if the model indicates that a wire center has high costs, but the model assumes much longer loop lengths than required by the actual network, the measured "high costs" might not exist in reality, but merely result from modeling error.

Q. Are there other types of inputs that are also important?

A. In addition to the "geographic" inputs discussed above, certain financial inputs tend to affect costs more than others. Some of the more important ones include:

Installed cable costs.
Fill factors
Structure sharing
Asset lives and depreciation rates
Cost of capital
Copper/fiber deployment
Drop lengths
Conduit costs
Fiber electronics costs
NID costs
Switch purchase prices and associated vendor discounts
Inflation factors and productivity adjustments
Maintenance and other expense factors

Q. Does the Commission need to determine the values of specific inputs in this phase of the proceeding?

A. No, this can occur at a later date. In fact, the Commission has already indicated that it intends to defer resolution of inputs until a subsequent phase or proceeding. [Order No. 5, Feb.3, 1999, p.8]. It is my understanding that the Commission intends to select an appropriate costing methodology or approach based upon the record in this phase; this will provide a foundation for a focused discussion of inputs in the next phase or proceeding.

At the time of prefiling this testimony, the FCC had not yet selected a specific set of inputs for use with its model; only illustrative input sets had been distributed by the FCC staff. Based upon these illustrative examples, there is reason to anticipate that the Commission may want to modify at least some of the FCC inputs for use in Kansas. For example, the cable cost assumptions reflected in the FCC staff's illustrative inputs appear to be rather high. Conversely, their conduit cost assumptions appear to be rather low.

Q. As you mentioned, the FCC has not chosen a geocode data source. Would you please discuss some of the options available to the Commission for selecting a data set for use with the FCC model?

A. Yes. An obvious choice is the PNR data set, which we have used in evaluating the FCC model. However, this data is slowly becoming outdated; it was developed during 1998 using 1996 and 1997 sources. The Commission might want to select a more current source. Furthermore, the PNR data is derived from mailing lists and other secondary sources–not actual list of telephone subscribers. Other sources are available, including white page address listings and LEC billing, customer service and engineering records.

In general, precise address listings for telephone customers should be used, to the extent these can be obtained and geocoded. In locations where available data sources fail to precisely locate a large fraction of the customers, estimating techniques, or alternative data sources (e.g. census data) must necessarily be relied upon.

The white pages are an excellent data source. This is inexpensive, public data that is readily available in a computerized, geo-coded format. By definition, however, this data excludes unlisted and non-published numbers, and thus the location of these customers must be estimated. This is not a particularly difficult problem, but it does require additional effort and it introduces potential for inaccuracy.

Mailing lists (used by PNR) have a similar, but opposite problem: while they include households with unlisted numbers, they also include households that do not have a phone. The latter locations need to be estimated and removed from the data, in order to match reality.

Internal LEC data bases offer an excellent alternative or supplemental data source. These generally include unlisted and non-published numbers, but they typically will require application of additional effort (e.g. geocoding) to use them with the FCC model. Another potentially useful data source is the census bureau. It provides detailed data concerning the number of households in each census block, which can be particularly helpful in areas where address listings cannot be adequately geocoded (e.g. in areas where rural routes or post office boxes dominate the address data). However, the 1990 census data are now nearly a decade old. The available updates are merely estimates--and not necessarily very accurate. Although white pages data exclude unpublished numbers, they are generally more accurate than census data in capturing the impact of differential growth rates and movement of customers within wire centers.