Disaggregated Spatial Modeling of Irrigated Land and Water Use

Pedro V. Garay[1],
Jeffrey M. Peterson[2],
Craig Smith[3],
Bill Golden[4]

May 4, 2010

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association Annual Meeting, July 23 -25, 2010.

[1] Graduate Research & Teaching Assistant, Department of Agricultural Economics, Kansas State University
[2] Associate Professor, Department of Agricultural Economics, Kansas State University
[3] Graduate Research Assistant, Department of Agricultural Economics, Kansas State University
[4] Research Assistant Professor, Department of Agricultural Economics, Kansas State University

Introduction

Water scarcity is already a critical issue across much of the world, and it will become an even greater concern in the semi-arid region of the eight U.S. states overlying the Ogallala aquifer. The Ogallala aquifer is one of the largest water resources in the world, underlying 111.6 million acres including portions of Colorado, Kansas, Nebraska, New Mexico, Oklahoma, South Dakota, Texas and Wyoming (U.S. Geological Survey, 2006). This area represents one of the most important agricultural regions in the nation.

The Ogallala aquifer contributes to the development of agricultural industries including irrigated crops, cattle feeding, and meat processing. After reliable technology became available to withdraw groundwater for irrigation in the mid Twentieth Century, the acreage of irrigated crops increased dramatically. As the Ogallala aquifer is the principal source of water for irrigation in the High Plains region, a natural consequence of a high demand of this resource for irrigation is a decline of water level availability. The area-weighted average water-level declined 14.0 feet from 1950 to 2007 (McGuire, 2009). Historically, the portion of the aquifer in Kansas has experienced one of the highest levels of depletion. For instance, for the period between 1950 and 2007 the water level change in the portion of the aquifer in Kansas declined 22.7 feet.

While there is an extensive literature developing economic simulation models of irrigated land and water use (see Brouwer and Hofkes (2008) and Harou et al. (2009) for recent reviews), little attention has been paid to the appropriate level of aggregation for these models. In most cases, researchers model the representative land unit in a relatively large region. This approach ignores any spatial variability within the region, which masks distributional issues that may be important for policy and, if many if the model relationships are highly nonlinear, may increase

the model's prediction error. In particular, even if one is only interested in aggregate measures at the regional scale, a more accurate prediction may be obtained by constructing many versions of the model – with each version calibrated to a distinct spatial unit—and then aggregating the results.

The main objective of this paper is to determinate the level of data aggregation that can give us the most accurate prediction of observed aggregates. To do so, we analyze the results of a Positive Mathematical Programming (PMP) model calibrated to data at varying levels of aggregation. In the first step of this approach, crop choices and water use at three level of data aggregation (county, crop report district and state aggregation) are simulated by a Positive Mathematical Programming (PMP) model (Howitt, 1995; Clark, 2008). A separate version of this model calibrates the observed crop allocation and water use in the base year (average from 2000 to 2008) to the three levels of data aggregation for comparison. This allows us to compare the approaches of (a) running a model with county-level data and aggregating the results to the state level, (b) running a model with data aggregated to multi-county crop reporting districts and then aggregating the results on further step to the state level and (c) running the models on state-level aggregate data.

Comparing the results and analyzing their prediction errors for the observed years following the calibration year (2000-08), allows us to test for aggregation bias. The measures that this paper uses to evaluate which of the simulation models fit better with the actual situation are the root-mean-square simulation error and the mean simulation error. The preliminary results suggest that the more disaggregated the data used in the calibration and simulation processes, the model fits better with the reality.

The rest of this paper is organized as follows. First, the models and methods are discussed. Second, the database that is used to calibrate and simulate the models is presented. Next, the results of the simulation at different levels of data aggregation are presented. The final section of the paper summarizes and concludes.

Methods and Procedures

The specific approach considered for this paper is The Positive Mathematical Programming (PMP). This method was initially used on policy model at different levels of production such as sectoral, regional and farm level. A peculiar aspect of this method is that, in spite several papers had used this method; it was not published until 1995 by Richard E. Howitt. In this article Howitt explains the characteristics and advantages of the PMP approach. An advantage of the PMP model is that, once calibrated, its solutions reproduce the observed patterns within a small tolerance, without adding 'flexibility' constraints in the model to artificially limit crop acreages (Howitt, 1995). This characteristic has made PMP a widely-used approach for policy modeling.

Our study region is a 31-county area overlying the Ogallala aquifer in western Kansas. We refer to this as the "region" level of aggregation in what follows. This region encompasses approximately the western third of Kansas. Following the boundaries of the USDA-NASS crop reporting districts, this region can be subdivided into three multicounty areas (including the northwest, west central, and southwestern Kansas crop reporting districts). These boundaries yield the intermediate "district" level of aggregation.

This paper applies the PMP model developed by Clark (2008) for irrigated agriculture in

western Kansas. Our analysis proceeds in three major steps. First, the PMP model is calibrated to observed crop allocation and water use in the base period (a nine-year average from 2000 to 2008) at each of the three levels of data aggregation. Second, the calibrated models are simulated over the data period, allowing prices to fluctuate yearly. A difference between this paper and Clark 2008 is that this paper uses the base period average price for calibration purpose, while for the simulation process; the yearly prices from the Kansas State University Extension budgets are used. Therefore, the prices for the simulation process are not constant. Third and finally, the results of the county- and district-level models are aggregated up to the region level so that their accuracies in predicting aggregate measures can be compared.

The model from Clark (2008) predicts land and water allocation for eight crops (irrigated wheat, nonirrigated wheat, irrigated corn, nonirrigated corn, irrigated sorghum, nonirrigated sorghum, irrigated soybean, and irrigated alfalfa).

The famer land and water allocation problem is represented as follows[5]:

$$max \sum_i p_i f_i(w_i)x_i - c_i(w_i, x_i; \alpha_i, \gamma_i, \delta_i)x_i \qquad (1)$$

s.t. $\sum_i x_i \leq b$

where b = size of the farm,

$f_i(w_i)$ = production function for crop *i*,

p_i = output price for crop *i*,

x_i = land area planted to crop *i*,

[5] The complete version of the model can be finding in Clark (2008).

w_i = water use per crop i,

$c_i(w_i, x_i; \alpha_i, \gamma_{i,}\delta_i)$ = cost function for crop i, and

$\alpha_i, \gamma_{i,}\delta_i$ = parameters of the cost function.

The production function that is used in this paper was developed by Martin, Watts, and Gilley (1984). The cost function is linear in both land allocations and water use, and it was specified by Clark (2008). The cost parameters are obtained from the calibration process and represent part of the necessary input for the simulation part. The model that is simulated for a period of t = 1… 9 years can be represented as follow:

$$\max \sum_i p_{i,t} f_i\left(w_{i,t}\right) x_{i,t} - \left[c_i\left(w_{i,t}, x_{i,t}, \widehat{\alpha_i}, \widehat{\gamma_i}, \widehat{\delta_i}\right) + k_t w_{i,t}\right] x_{i,t} \qquad (2)$$

$$\text{s.t.} \sum_i x_{i,t} \leq b$$

$$\sum_i x_{i,t} \leq b_a$$

$$w_i \leq m_t, \qquad i \in Q$$

where Q = set of indices of irrigated crops,

m_t = maximum pumping given aquifer conditions in year t

b_a = legally authorized irrigated acreage,

k_t = extra cost of pumping per year,

$p_{i,t}$ = output price per crop and per year.

Several measures can be calculated to assess a model's prediction accuracy. Willmott et al. (1985) mentioned different measures that can be estimated to determinate both model accuracy and precision. In the set of different measures they emphasize the approaches based in the difference between the elements and values of a model predicted and observed. These measures include the root-mean-square error (RMSE), the systematic root-mean-square error (RMSEs), the unsystematic root-mean-square error (RMSEu), and the index of agreement (d2). They also present additional indices such as the mean absolute error (MAE) and a modified version of the index of agreement (d1). They also argued that bootstrapping can be used to evaluate both the confidence and significance associated with each of the difference indices. They showed that if the difference measures mentioned above are used in combination with the correct statistics and data-display graphics the evaluation of the performance of models can be accomplished.

Let Y denote some endogenous variable of interest that is simulated by the model. In this analysis we calculate and report the root-mean-square simulation error using the following formula (Pindyck and Rubinfeld, 1998):

$$rms\ error = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^S - Y_t^a)^2} \qquad (3)$$

where Y_t^S = simulated value of Y_t

Y_t^a = actual value

T = number of periods in the simulation

The root-mean-square-simulation error (RMSE) measures the differences between the values predicted by a simulation model and the actual value of the studied variable. The RMSE

is measured in the same units as the data, and is representative of the size of a typical error.

Additionally, the root-mean-square percent simulation error (RMSPE) is calculated following:

$$rms\ percent\ error = \sqrt{\frac{1}{T}\sum_{t=1}\left(\frac{Y_t^s - Y_t^a}{Y_t^a}\right)^2} \qquad (4)$$

The root-mean-square percent error provides the same properties as the root-mean-squares error, but expressed as percents.

Data

This paper uses county- level data from different sources. Crops acres and yields from 2002 to 2008, for the thirty-one counties located in the northwest, west-central and southwest of Kansas, are from the National Agriculture Statistics Service (NASS). The base period for the calibration and simulation models is the average of this nine-year period 2000-08.

The crop prices and the cost variables are from Kansas State University Extension budgets. The crops price are expected prices and the cost variables, which are included in the model are irrigation costs, variable expenses, fertilizer and seed costs, and harvest and hauling cost. The aquifer level data are from the Kansas Geological Survey. The variables which are considered from the later source are lift, saturated thickness, specific yield, land area above aquifer, hydraulic conductivity, annual recharge rate, and depth. The number of wells is from the Water Information Management and Analysis System (WIMAS).

Another important group of data is the agronomic data, which are obtained from several

sources. Agronomic variables required for the model include actual irrigation water use, irrigation required for fully watered yield, precipitation level, gross and net irrigation requirement, fully watered yield, and dryland yield. These data were obtained from the Water Information Management and Analysis System, NOAA National climatic Data Center, Kansas Weather Data Library, The National Engineering Handbook, Stone et. al; and O'Brien et. al.

Results

The models calibrate and simulate crop allocation and water use for eight different crops. As the main objective of this paper is to evaluate which of the three simulation models fits better with the actual value of the variables, irrigated corn acreage is selected for comparison purposes. The following graph shows the simulation results obtained from the three models and the actual value of the irrigated corn from 2000 to 20008.

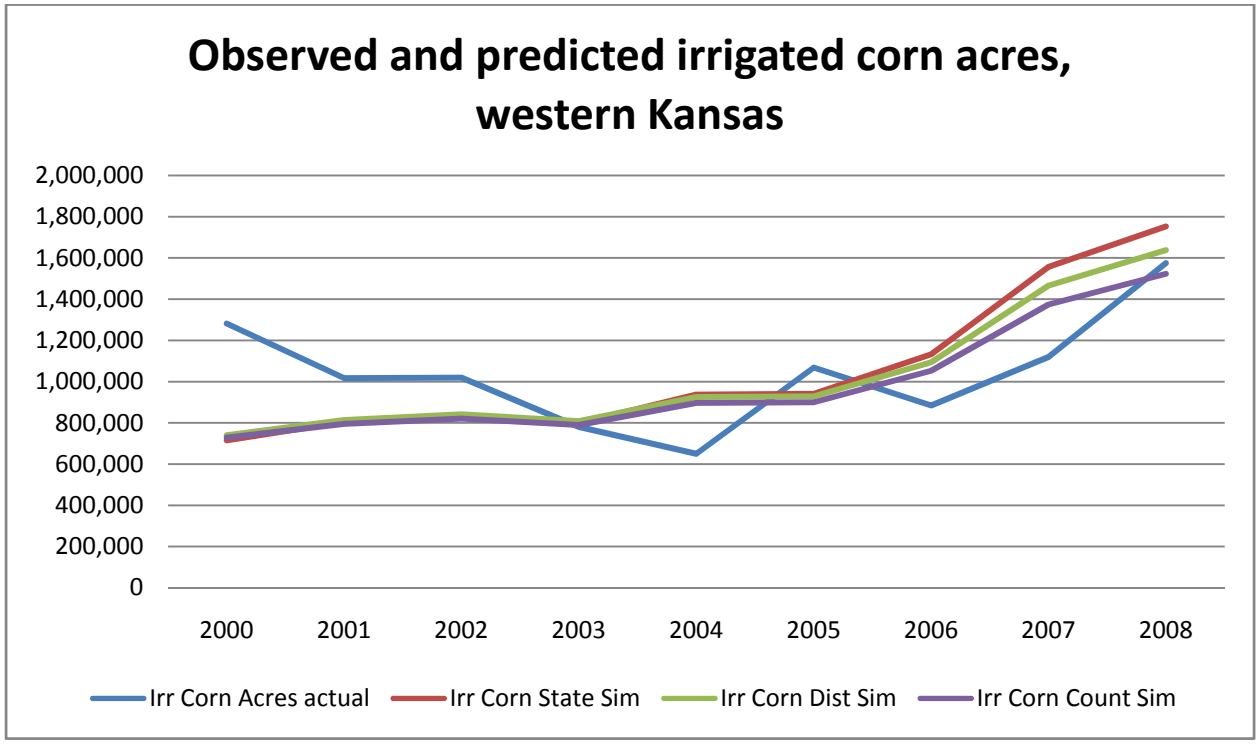

Figure 1

The blue line reflects the actual value of irrigated corn acreage corresponding to the 31-county region obtained from the National Agriculture Statistics Service (NASS) for the period of 2000 to 2008. The red line represents the simulated results of irrigated corn using the most aggregated data. This model is called state level simulation. The green line reflects the result obtained from the model that used the district level aggregation. The value was obtained adding the irrigate corn acreages from each district. This model represents the first level of data aggregation and it is called district simulation model. Finally, the purple line represent the results of the third model called county level simulation. The value of this line is the summation of irrigate corn acreages obtained in the simulation for each of the thirty-one counties.

To make a more precise evaluation about the fit of three models, the root-mean-square simulation errors and root-mean-square-percent simulation error are presented in the following table.

Table 1. Model Prediction Errors for Region-Level Irrigated Corn Acreage, 2000-08

Model	RMSE (acres)	RMSPE (percent)
State Level Aggregation	294,911	28.37
District Level Aggregation	264,777	25.83
County Level	254,129	24.28

The root-mean-square-simulation-error (RMSE) is a measure of the differences between values predicted by a simulation model and the actually values observed from the variables modeled. It is a measure of precision. The results of RMSE presented in the above table show

that the model that used the less level of data aggregation fits better with the actual value of the variable than the others two models. In the same sense, the district level aggregation model fits better than the state level model. These results are consistent with the second measure of precision represented by the root-mean-square percent error. Both the RMSE and the RMSPE are consistent with the expected results under the argument that the more aggregated the data, the probability of getting sample errors are higher.

Conclusion

The main objective of this paper was to determinate the level of data aggregation that can give us a better fit of a simulation model to actual value of the variable studied. This paper analyzed the data aggregation issue using a Positive Mathematical Programming (PMP) model constructed by Clark (2008). The data to calibrate and simulate the models were obtained from NASS, Kansas State Research and Extension, the Kansas Geological Survey, the Water Information Management and Analysis System, and Kansas Weather Data Library.

The models were calibrated with data aggregated to the county, crop reporting district, and region level for comparison. This allowed us to compare the approaches of running disaggregated models and then aggregating the results, and running the models on aggregate data. To analyze the results, the root-mean-square simulation error for each model was calculated. For simplicity, just the irrigate corn acreages was chosen as variable of comparison.

The results of the tests suggest that model that is simulated with most disaggregated data gives us the smallest value of the RMS simulation error, and the smallest percentage error, implying that for this specific variable (Irrigate corn acreages) the data aggregation can generate

a big difference between the values predicted by the model and the values actually observed. Consistent with this result, a graphical analysis shows that the results obtained from models run with disaggregated data follow more closely the actual value pattern of the selected crop.

This paper can be considered as initial study for future researches which can provide information about the ideal level of aggregation for different types of data such as dryland crops and water use.

References

Clark, M. (2008), “Effects of High Commodity Prices on Western Kansas Crop Patterns and the Ogallala Aquifer”, M.S. Thesis, Department of Agricultural Economics, Kansas State University.

Brouwer, R. and M Hofkes, “Integrated hydro-economic modeling: Approaches, key issues and future research directions” *Ecol. Econ*., 66(1), 16–22, 2008.

Harou, J. J., Pulido-Velazquez, M., Rosenberg, D. E., Medellin-Azuara, J., Lund, J. R., and Howitt, R. E.: Hydro-economic models: Concepts, design, applications and future prospects, *Journal of Hydrology*, 375(3–4), 627–643, 2009.

Howitt, Richard E (1995), "Positive Mathematical Programming." *American Journal of Agricultural Economics* 77.2: 329-42.

Kansas Geological Survey (KGS). “High Plains Aquifer Section-Level Database.” Available at http://hercules.kgs.ku.edu/geohydro/section_data/hp_step1.cfm

Martin D.L., D.G. Watts, J.R. Gilley. “Model and Production Function for Irrigation Management.” *Journal of Irrigation and Drainage Engineering.* Vol 110, No 2, June 1984, pp 149-164

McGuire, V.L. (2009), “Water-level changes in the High Plains aquifer, predevelopment to 2007, 2005–06, and 2006–07”: U.S. Geological Survey Scientific Investigations Report 2009–5019, 9 p., available at: http://pubs.usgs.gov/sir/2009/5019/.

Pindyck R.S., D.L. Rubinfeld, “Econometric Models and Economic Forecast,” Fourth Edition The McGraw-Hill, 1998

Willmott, C.J., S.G. Ackeleson, R.E. Davis, J.J. Feddema, K.M. Klink, D.R. Legates, J. O’Donnell, and C.M. Rowe, “Statistics for the Evaluation and Comparison of Models”, *Journal of Geophysical Research*, 90 (1985): 8995-9005.

U.S. Geological Survey (2006), “Water use in the United States”: U.S. Geological Survey data, accessed December 2006 at http://water.usgs.gov/watuse/.