Predicting Large-Area Corn Yield with a Weighted Palmer *Z*-Index

SCOTT A. ISARD[†]

Department of Geography, University of Illinois, Urbana, Illinois

WILLIAM E. EASTERLING[*]

Climate and Meteorology Section, Illinois State Water Survey, Champaign, Illinois

(Manuscript received 15 November 1987, in final form 4 October 1988)

ABSTRACT

Palmer's *z*-index, calculated to reflect only the planting-emergence and anthesis-grainfill stages of the growing season, is related with detrended corn yields to produce a predictive model for Illinois corn production. The model is evaluated to see how well it can predict mean bushel per acre corn yields for large areas (state of Illinois). Results suggest the *z*-index, if calculated to emphasize moisture-sensitive periods in corn production, is a reliable predictor of yields, and, moreover, this predictive ability improves with more extreme moisture conditions.

1. Introduction

In the future, detailed crop simulation models (e.g., CERES-MAIZE, SOYGRO, COMAX) will provide the best area-wide assessments of impacts of weather anomalies such as drought on crop production. At present, however, restrictive data requirements and lack of extensive field verification somewhat limit these approaches as tools for large-area agricultural drought monitoring. Though not developed with conventional agroclimatic indices (e.g., growing degree days and radiation), standard water budget-based drought detection and monitoring indices, such as the Palmer Drought Severity index (PDSI or X), its precursor (Moisture Anomaly Index or z), and its derivative (Crop Moisture Index), have some value in assessing the affect of varying moisture conditions on agricultural production. Despite criticism of their underlying assumptions (see Alley 1984) and their lack of explicit conjunctive use with known moisture-sensitive crop stages and management decisions (see Wilhite and Glantz 1985), Palmer-related indices have been successfully correlated with crop yields (e.g., Sakamoto's 1978 use of the z-index to estimate South Australian wheat yields). Indeed, Easterling et al. (1988) have used results of an Illinois corn-climate sensitivity analysis to weight Palmer X- and z-indices to reflect how different types of moisture conditions interact with crop phenology and crop management decisions. These weighted indices were shown to correlate with final yields more closely than use of X- or z-indices that are simply averaged over the growing season.

The purpose of this paper is to extend the use of weighted z-indices (as described in Easterling et al. 1988) to develop and evaluate a predictive model of Illinois corn yields. Specifically, the z-index, calculated only to reflect the critical moisture-sensitive periods encompassing spring planting-emergence and midsummer reproduction-grainfill, is related with corn yields to parameterize a simple linear regression model from which area-wide yield predictions can be derived using z-index values.

2. Data and methodology

a. Data

As in the Easterling et al. (1988) study, 24 yr (1960–83) of climatological and crop yield data are obtained from 12 National Weather Service stations and associated counties across Illinois. Although both datasets extend prior to 1960, generally there is little reliability in the accuracy of pre-1960 crop yield estimates. Thus, these earlier data were omitted.

General examination of the Illinois climate record (1960–84) raises no serious concern over the appropriateness of this period for developing empirical weather-yield relationships for predictive purposes.

[†] Also affiliated with the Climate and Meteorology Section, Illinois State Water Survey, Champaign.

[*] Also affiliated with the Department of Geography, University of Illinois, Urbana.

Corresponding author address: Dr. Scott Isard, Dept. of Geography, 220 Davenport Hall, University of Illinois, Urbana, Illinois 61801.

Richman and Easterling (1988) have examined recent growing season climate trends in the Midwest, as related to corn production. Though periods of persistent climate fluctuations covered portions of Illinois at different times during the 1960–84 period, it would not be possible to characterize the period as being singularly good or bad for Illinois corn production. Rather, the large mix of corn growing conditions found by Richman and Easterling suggests that model estimates would be valid over a wide range of climatic conditions.

Corn yields, in bushels per acre, are detrended in a conventional manner (e.g., McQuigg et al. 1973; Sakamoto 1978) using a simple linear regression model (with time as the independent predictor of yield) to remove the technology influence on production. Thus, the values of corn yield used in this study are raw residuals from the regression model and are expressed as interannual departures from the linear trend. Although far from optimal, this technique does permit the interannual yield data to reflect weather influences more clearly than unmodified corn yields. Durick (1986) and Croason (1986) believe present rates of yield increases attributed to technological progress will continue for some time into the future. Durick (1986) proposes that advances in plant breeding alone will result in 1% to 2% increase of annual yields (based on a percentage of yields over time). This would seem to support the validity of our detrending technique, shortcomings notwithstanding, for making future projections.

b. Methodology

The Palmer z-index is calculated using standard computational procedures (see Palmer 1965) and validated with datasets provided by Palmer (1965, Table 12) and Alley (1984, Table 2). Karl (1986) suggested that the z-index is a better indicator of agricultural drought than the accumulated PDSI since the z-index, by virtue of its discrete monthly calculation, is more reflective than the PDSI of short-term moisture shifts during sensitive crop development stages at a given location. Although Easterling et al. (1988) concluded that for midwestern corn production there is no clear advantage to using one index over the other, the z-index can be calculated more easily than the PDSI and thus is used exclusively in this study to model corn yield deviations.

The z-index is weighted in such a manner to reflect the generally positive effect on corn production of dry springs and generally negative effect of dry mid- to late summers. A dry spring permits early (or normal) planting activities to take place, gives full-season varieties maximum opportunity to mature, maximizes benefits from fertilizer application, and, provided moisture is adequate for germination (as it almost always is in Illinois, drought or not), encourages vertical development of root systems. A dry mid- to late summer denies corn plants moisture during the critical silking and tasseling time (anthesis), and the grainfill stage; both of these periods are when moisture and nutrient uptake demands are high, and insufficient amounts of either are detrimental to yields. The converse, wet springs and wet mid- to late summers, would achieve the opposite effect on yields to the conditions discussed above; i.e., wet springs would tend to be yield-depressing and wet mid- to late summers would be yield-enhancing.

The z-index weighting scheme (discussed more fully in Easterling et al. 1988) is developed in the following manner. First, for each of the 12 stations shown in Fig 1, we give a planting and emergence period index p by

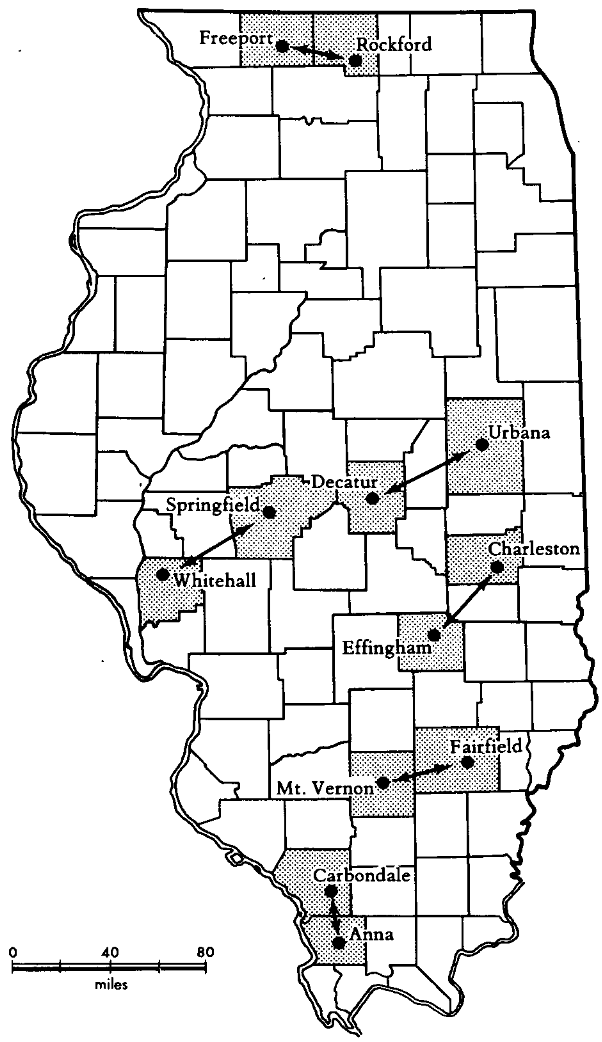

FIG. 1. Pairs of Illinois counties used in the analysis of the relationship between the y'-index and corn yield.

$$p_s = \sum_{i=14}^{22} z_i/9.0,$$
$$p_n = \sum_{i=16}^{22} z_i/7.0, \qquad (1)$$

where s is the southern half of Illinois; n, the northern half; i, the week of the year; and z, the weekly value of the z-index.

Second, we obtain an anthesis and grainfill index (a) by

$$a_s = \sum_{i=27}^{30} z_i/4.0,$$
$$a_n = \sum_{i=28}^{31} z_i/4.0. \qquad (2)$$

Our final composite z-index (y'), reflective of the appropriate planting-emergence and anthesis-grainfill moisture sensitivities, is given as

$$y' = a - p. \qquad (3)$$

c. *Predictive model*

The main question posed in this paper was, How well does the y'-index, which is developed from point specific data (NWS climate stations) predict corn yield deviations for the immediate county? A corollary to this is the added question of how well the y'-index predicts areally averaged corn yield deviations (i.e., corn yield deviations in neighboring counties). This corollary question becomes meaningful if a predictive model based on the y'-index is to be relevant for large areas such as the state of Illinois. It is possible to address both of these questions concurrently in the following modeling protocol.

The first step is to pair spatially nearest neighbor stations resulting in a total of six pairs (12 total stations). These pairings are shown in Fig. 1. The next step is to arbitrarily choose one station out of each of the six spatial pairs and calibrate a linear model regressing the y'-index values for the first six stations on the crop yield deviations for the corresponding six counties. The intention, then, is to predict crop yield deviations for the second six neighboring counties using parameter estimates from the first six counties and y'-index values for the second six stations.

Two underlying issues must be addressed in this approach. First, the y'-index values range from negative (yield-depressing conditions) to positive (yield-enhancing conditions) and as y'-index values approach zero, it is quite likely that other factors, (e.g., pests, disease) may assume a greater importance than moisture availability in accounting for yield deviation. On the other hand, it is equally likely that large y'-index values will accentuate the effect of moisture availability and yield deviations. Therefore, it is argued that the model be calibrated on three sets of grouped data: (i) all available observations for the six stations, (ii) all observations in which $|y'| \geq 1$, and (iii) all observations in which $|y'| \geq 2$.

Second, it is recognized that the choice of counties to include in model calibration versus the counties used for validation from any given spatial pair will present an artifact in the parameter estimates. As a result, the predictive model was recalculated for each of the 2^6 unique combinations of spatially paired counties. Results of the analysis are reported below in summary form.

3. Results and discussion

a. *Results*

Table 1 shows the mean parameter estimates (and their standard deviations) for the models developed with the 2^6 possible county and station combinations. Two major characteristics of these results are noted. First, the standard deviations of the parameter estimates are small. This suggests that regardless of which 6 of the 12 counties are included in the calibration of the model, the parameter estimates are not meaningfully affected; i.e., the parameter estimates are stable and, implicitly, reliable throughout the 2^6 combinations of counties.

Second, in stratifying the data into the three groups according to severity of the y'-index (as discussed above), it is clear that the model strength improves as the absolute value of the y'-index increases. The mean r^2 value for the 2^6 models, calibrated on all observations is 0.36. The elimination of those y'-index values closest to 0 ($|y'| < 1$) results in an improvement in the mean r^2 value to 0.69. Further elimination of all y'-index values between 2 and -2 results in a mean r^2 of 0.71.

TABLE 1. Mean values of least-squares parameter estimates for the linear relationship between y'-index values and corn yield deviations for the 2^6 possible combination of Illinois counties.†

Range of y' index values	n	b_0	b_1	r^2
All y' values	138	−0.09 (0.07)	5.03 (0.52)	0.36 (0.02)
$\|y'\| \geq 1$	59	−2.79 (0.45)	4.83 (0.51)	0.60 (0.03)
$\|y'\| \geq 2$	28	−2.31 (0.83)	4.84 (0.49)	0.71 (0.04)

† Here n is the average number of observations in each half of the dataset, b_0 the regression intercept, b_1 the regression coefficient, and r^2 the coefficient of determination.

Numbers in parentheses represent the standard deviation of the mean; H_0: $b_1 = 0$ rejected at $\alpha = 0.10$ for all 2^6 combinations of stations.

These findings suggest that planting-emergence and anthesis-grainfill moisture conditions, as expressed in the y'-index, are effective in accounting for corn yield deviations in the immediately surrounding county. Moreover, when the data is stratified to include only the more extreme y'-index values, explanation of crop yield deviations improves considerably. This would support the argument that nonclimatic-related factors are at least equally important as climate when y' values are close to 0.

The subsequent step is to examine how well parameter estimates for the relationship between y'-index and corn yield deviations from the six counties predict the yield deviations in the six counties reserved for model validation. The key issue here is determining how accurately parameter estimates developed with data from one station–county set can account for crop yield deviations in an adjacent (yet different) geographic area.

Root-mean-square error (RMSE) and mean bias error (MBE) values are calculated using predicted and observed yield deviations for the counties reserved for model validation. Table 2 shows the mean values of the RMSE and MBE (and their associated standard deviations) over the 2^6 model runs for each of the three categories of y'-index values. In both cases (MBE and RMSE), the exhibited units are in bushels of corn per acre.

The MBE, an indicator of degree to which the model estimates over- or underpredict the observed yield deviations, are extremely small. In fact, the only conclusion to be drawn from these values is that there is no consistent bias among the model runs toward either underprediction or overprediction of the yield deviations. Moreover, there is no appreciable impact on estimates of corn yield deviations of stratifying the dataset on the basis of extreme y'-index values.

The mean RMSE, a measure of the mean absolute error between predicted and observed yield deviations, over all the model runs is also relatively small. In addition, based on standard deviations, these values are remarkably consistent across the 2^6 datasets. The largest RMSE value of 13.08 bushels per acre is, as expected, associated with model runs including all values of the y'-index. Also as expected, the mean RMSE values decrease when the model is run on datasets containing only the more extreme y'-index values.

These results suggest that the parameter estimates for predicting corn yield deviations from one-half of the spatially paired stations can reliably predict the observed corn yield deviations from the remaining half in Illinois. Moreover, there does not seem to be any appreciable impact on the model parameter estimates of the choice of counties upon which the model is calibrated. Nor does this choice seem to have significant bearing on the skill of the model in predicting adjacent county corn yield deviations.

TABLE 2. Mean values of MBE and RMSE of the corn yield estimates for the 2^6 possible combinations of Illinois counties.†

Range of y' index values	n	MBE	RMSE
All y' values	138	0.01 (0.14)	13.08 (0.33)
$\|y'\| \geq 1$	59	−0.02 (0.40)	7.80 (0.35)
$\|y'\| \geq 2$	28	−0.03 (0.32)	5.73 (0.64)

† Numbers in parentheses represent the standard deviation of the mean. Here n is the average number of observations in each half of the dataset.

b. *Discussion*

These results support the argument that knowledge of moisture conditions in the spring planting-emergence and mid- to late summer anthesis-grainfill periods, as represented by the y'-index (weighted z-index), can be used in a growing season to forecast, with some skill, season-ending corn yield deviations from the time trend. Specifically, the use of y'-index values calculated as early as mid-August (end of grainfill period) during the course of a growing season in conjunction with the mean parameter estimates from the 2^6 combinations of stations can give a prediction of final corn yields in the corresponding season for Illinois. This would permit 1 to 2 months lead time for agricultural decision making.

Moreover, the reliability of the predictions of crop yield deviation increases for periods of benign or stressful climate. Obviously, it is under climatic conditions that are unusually favorable or unfavorable to crop production that prior knowledge of crop response can be most useful. However, it must be noted that *the regression equations are not physically based, and thus they cannot be expected to yield reliable predictions outside the range of conditions upon which they were calibrated.* For example, no rain during the spring following a dry fall and winter combined with torrential precipitation during anthesis would produce large y' values but poor corn yields. Similarly, favorable weather conditions during planting and anthesis could be offset by extremely hot and dry weather during the intervening period.

4. Conclusion

Agricultural decision making could benefit from reliable predictions of crop yield response to moisture conditions during a growing season, especially under anomalous climatic conditions. In this research, it is shown that the Palmer z-index, calculated to reflect the appropriate crop response to specific types of moisture conditions, can be of use in predicting crop yields for large geographic areas. It should be noted that the specific relationships between the accumulated z-index during spring and midsummer as related to corn yield presented in this study are only valid for Illinois. For example, the equations suggest that dry springs are conducive to high yields. This is probably not the case in the western corn belt where moisture may be a more limiting factor for growth than in Illinois. The applicability of the Palmer drought index outside of humid areas is also limited. For example, the use of the Thornthwaite method of calculating evapotranspiration can produce an underestimate of the Palmer z-index in subhumid and semiarid areas. Finally, since the model predicts deviations from a corn yield–time trend line, forecast for future yields are valid only as long as the trend line remains the same.

Notwithstanding these limitations, the y'-index (weighted Palmer z-index) reflecting planting-emergence and anthesis-grainfill exhibits significant predictive capabilities for corn yield deviations throughout Illinois. A model developed with data from six stations across Illinois is shown to provide considerable skill in relating with actual crop yield deviations for spatially adjacent areas. With the model, it would be possible to predict crop yield deviations for the harvest based on weighted z-index values available from the National Weather Service in late summer.

Who might benefit from such predictions and how would they be utilized? Examples come from virtually all aspects of corn production. If available to farmers, such predictions could help them choose the optimal harvest time, schedule equipment and labor for the harvest, and abandon acreage (nonharvest of planted acreage). Agribusiness, such as food processing manufacturers who are dependent on farm outputs, could find such yield predictions useful in planning commodity purchases. Alternative sources of commodity supplies could be sought if weather-induced crop yield disruptions are projected to be severe in certain production areas. Institutions that provide support services to the corn producers might also benefit from such predictions. Banks and commodities traders could find this information useful in evaluating the extent to which yield disruptions may affect short-term price stability. Of course, major support services for corn producers are federal, state, and local governments. Yield predictions can be routinely distributed by these governments to farmers and agribusiness. Moreover, such information could be useful to governments in targeting areas for special program assistance.

These examples are only illustrative of the kinds of uses that could be made of such yield predictions. Details on these uses could be the subject of another paper. However, for the interested reader, there is a growing volume of literature on the use of predictive information in agricultural decision making (e.g., Sonka et al. 1982; Easterling and Mjelde 1987).

Acknowledgments. This work was partially supported by NSF Grant ATM 84-13043. The authors wish to thank Stanley A. Changnon, Jr. and Peter J. Lamb for advice on the effort. Also, Patrick Guinan, Pamela Warren, and Mark Shafer provided valuable assistance in the data handling and analysis.

REFERENCES

Alley, W. M., 1984: The Palmer Drought Severity Index: Limitations and assumptions. *J. Climate. Appl. Meteor.*, **23**, 1100–1109.

Crosson, P., 1986: Agricultural-development: Looking to the future. *Sustainable Development of the Biosphere,* W. C. Clark and R. E. Munn, Eds., Cambridge University Press, 104–136.

Duvick, D. N., 1986: North American grain production: Biotechnology research and the private sector. *The Future of the North American Granary: Politics, Economics and Resource Constraints in North American Agriculture,* C. F. Runge, Ed., Iowa State University Press.

Easterling, W. E., and J. W. Mjelde, 1987: The importance of seasonal climate prediction leadtime in agriculture decision making. *Agric. Meteor.*, **40**, 37–50.

——, S. A. Isard, P. Warren, P. Guinan and M. Shafer, 1988: Improving the detection of agricultural drought: A case study of Illinois corn production. *Agric. Meteor.*, **43**, 37–47.

Karl, T. R., 1986: The sensitivity of the Palmer Drought Severity Index and Palmer's Z-index to their calibration coefficients including potential evapotranspiration. *J. Climate. Appl. Meteor.*, **25**(1), 77–86.

McQuigg, J. D., L. Thompson, S. Le Duc, M. Lockard and G. McKay, 1973: The Influence of Weather and Climate on United States Grain Yields: Bumper Crops or Drought. Rep. to the Associate Administrator for Environmental Monitoring and Prediction, NOAA/U.S. Department of Commerce.

Palmer, W. C., 1965: Meteorological drought. Res. Pap. 45, U.S. Weather Bureau, Washington, D.C. [Available through NOAA Library and Information Services Div., Washington, D.C. 20852.]

Richman, M. B., and W. E. Easterling, 1988: Procustes Target Analysis: A multivariate tool for identification of climate fluctuations. *J. Geophys. Res.*, **93**(D9), 10 989–11 003.

Sakamoto, C. M., 1978: The Z-index as a variable for crop yield estimation. *Agric. Meteor.*, 305–314.

Sonka, S. T., P. J. Lamb, S. A. Changnon Jr. and A. Wiboonpongse, 1982: Can climate forecasts for the growing season be valuable to crop producers: Some general considerations and an Illinois pilot study. *J. Appl. Meteor.* **21**(4), 471–476.

Wilhite, D. A., and M. H. Glantz, 1985: Understanding the drought phenomenon: The role of definitions. *Water. Inter.*, **10**, 111–120.