Optimal Data Analysis
Vol. 2, Release 2 (October 30, 2013), 120-142

2155-0182/10/$3.00

Trajectory of Crude Mortality Rate in North Dakota Counties

Paul R. Yarnold, Ph.D.
Optimal Data Analysis, LLC

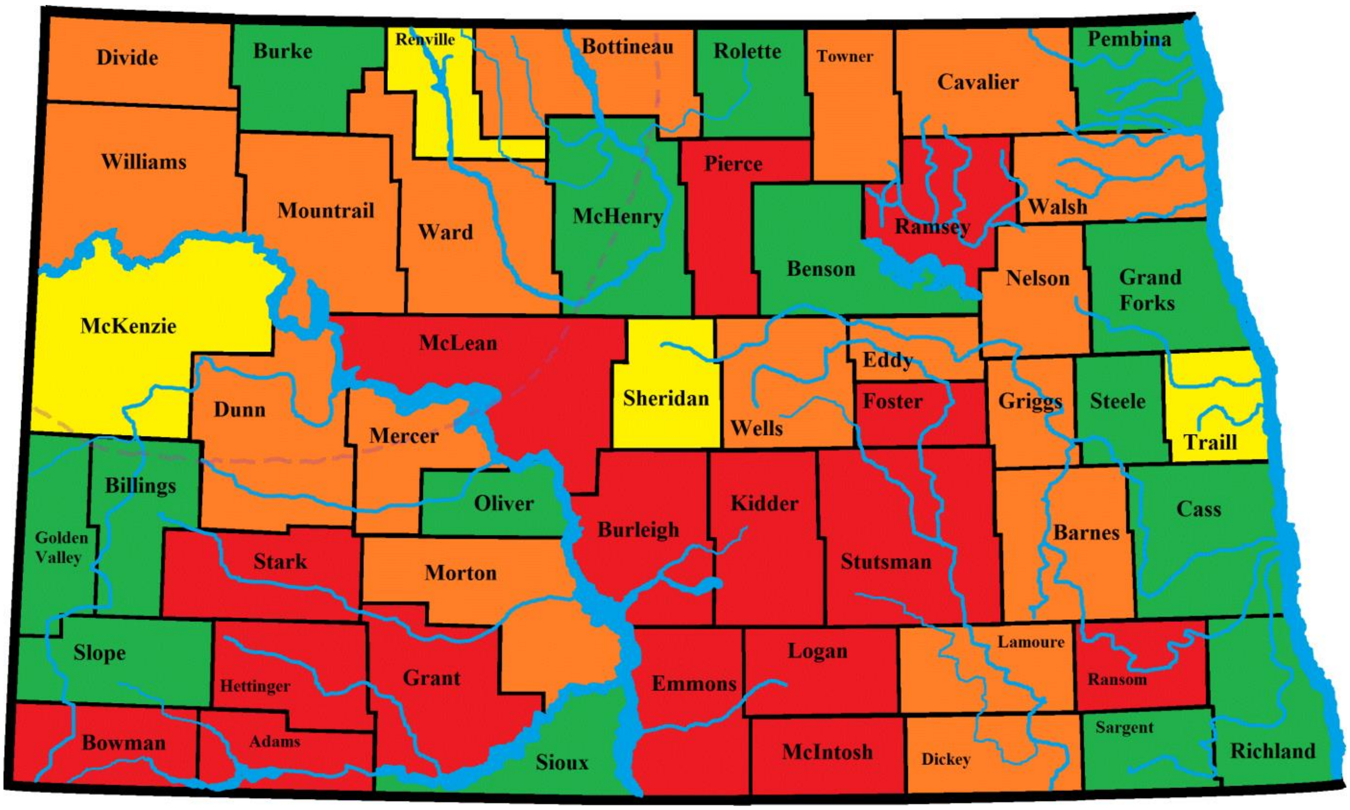
Recent research reported support for the *a priori* hypothesis that annual crude mortality rate (ACMR) was higher after widespread commercial usage of toxic chemicals and biocides began in the environment in North Dakota in 1998.[1] UniODA[2] was used to compare ACMR in 1934-1997 versus 1998-2005 (the most current data available[3]). Different county types identified included those for which ACMR increased significantly ($p<0.05$) at the experimentwise criterion; increased significantly ($p<0.05$) at the generalized criterion; had a statistically marginal increase ($p<0.10$) at the generalized criterion; or did *not* have a statistically significant increase ($p>0.10$) at the generalized criterion. Prior research is extended by investigating the path of ACRM over time for these four county groupings, and for each county considered separately. The pattern of mean ipsative ACMR across time for counties experiencing a *statistically significant increase* in ACMR after 1998, revealed the means fell into three almost perfectly discriminable levels: (a) *low* mean scores (mean $z<0$) seen early (1965 or earlier); (b) *medium* scores ($0<$mean $z\leq 0.5$) in the middle of the series (1966-1985); and (c) *high* scores (mean $z>0$) late in the series (1986 or later). Series achieved mean $z>1$ in 1998, and mean scores since 1998 have been among the highest on record.

In Figure 1 the counties in which ACMR had a statistically significant ($p<0.05$) assessed by the experimentwise criterion[2] are shown in *red*; those in which ACMR increased significantly ($p<0.05$) by the generalized criterion are indicated in *orange*; counties with a statistically marginal increase ($p<0.10$) by the generalized criterion are given in *yellow*; and counties *not showing* a significant increase ($p>0.10$) in ACRM by the generalized criterion are displayed in green.

This study sheds more light on this issue by reporting the path of ACRM across time for these groups, and for each county individually. All data are not simply combined to conduct an overall analysis, in order to avoid the induction of paradoxical confounding.[4-6] For example, if data for a county having ACMR increasing over time were combined with data for a county with ACMR decreasing over time, the combined data may indicate no temporal change in ACMR.

2155-0182/10/$3.00

Figure 1: Findings of *a priori* Test of the Hypothesis that ACMR Increased Significantly After Commercial Usage of Toxic Chemicals and Biocides Began in 1998 in the Environment (Waterways Drawn by Artist, and are Not Shown to Scale)[1]



Expressed in raw score form the ACMR data exhibit significant base-rate differences, so data were ipsatively standardized separately by county to eliminate variability attributable to the base-rate, and to express the data using a scale that enables direct comparison between different time series.[7] The first group of analyses examine the course of ipsative ACMR scores across time for each group identified in prior research.[1]

For counties having significantly greater ACMR (experimentwise criterion) after 1998, Figure 2 is a scatterplot representing all 16 red counties seen in Figure 1, showing a polynomial regression model[8] for the data and its R^2 value. As seen, the regression model explains less than 54% of the variation in ipsative ACRM scores, indicative of poor predictive accuracy.[9] The model crossed from negative to positive *z* scores (mean of *z* is zero) in 1976, and reached a value of 1.0 in 1998. Several notable outliers reflect inherent variability in the ACMR statistic for counties having small populations.

Figure 3 provides mean ipsative ACMR scores over time, indicating the 8-year forward moving average—selected since eight years of ACMR data are available after 1998, inclusive[3]. Note that all means prior to 1965 were negative (below mean) but one; between 1965 and 1985 all means exceeded zero and were less than 0.5, and all means since 1985 but one exceeded 0.5.

2155-0182/10/$3.00

Figure 2: Ipsative ACMR Data for Counties Having *Significantly Greater* ACMR (*Experimentwise* Criterion) After 1998, Indicating Polynomial Regression Model

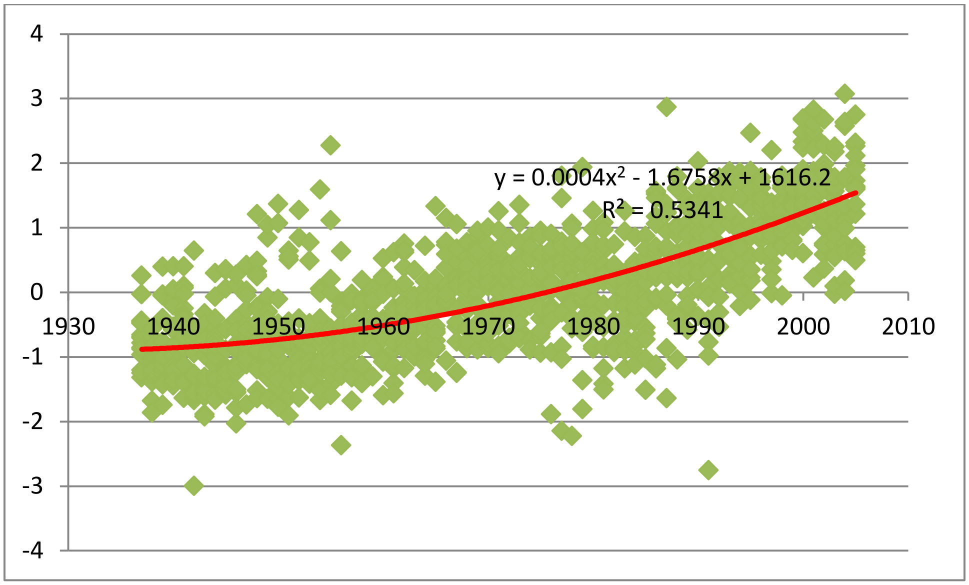
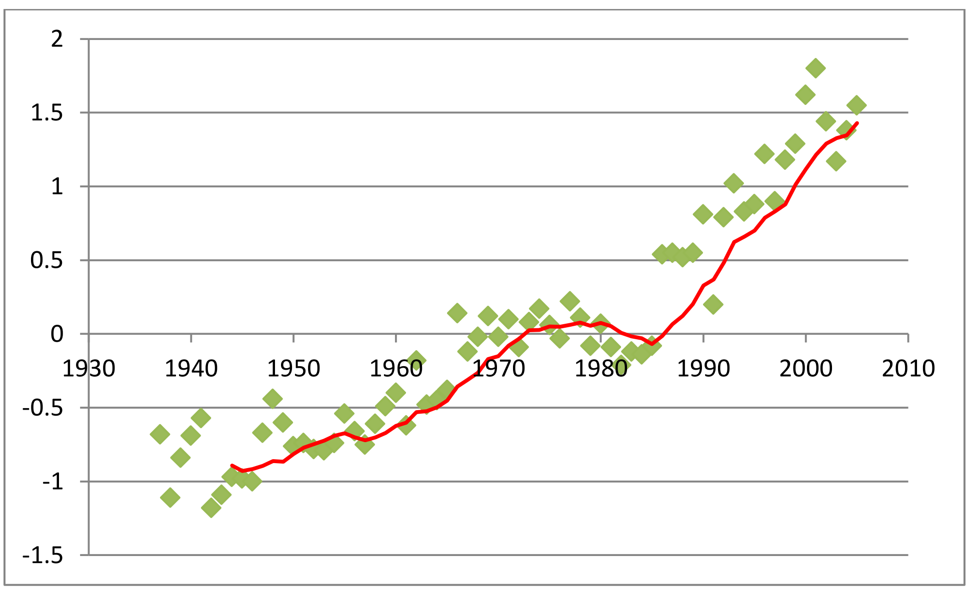


Figure 3: Mean Ipsative ACMR Data for Counties Having *Significantly Greater* ACMR (*Experimentwise* Criterion) After 1998, Indicating 8-Year Forward Moving Average



The accuracy achieved in discriminating mean ipsative *z* score levels on the basis of year as just described is virtually perfect (ESS=95.4), as is predictive value (ESP=95.8). A qualitative analysis of the scatterplot reveals that the means increased slowly between 1934 and 1965; then abruptly experienced an increase of one-half a standard deviation (SD) magnitude in 1965 and then remained stable until 1985; and since 1985 mean *z* scores assumed uninterrupted linear rise. Mean ipsative *z* surpassed 1.0 for good in 1998, and since that time the six highest-ever-recorded means have been observed.

For counties having significantly greater ACMR (generalized criterion) after 1998, Figure 4 is a scatterplot for 18 orange counties seen in Figure 1. The regression model explains 41% of the variation in ipsative ACRM scores: it crossed through negative *z* scores in 1971, and it reached a value of 1.0 in 2004.

Figure 4: Ipsative ACMR Data for Counties Having *Significantly Greater* ACMR (*Generalized* Criterion) After 1998, Indicating Polynomial Regression Model

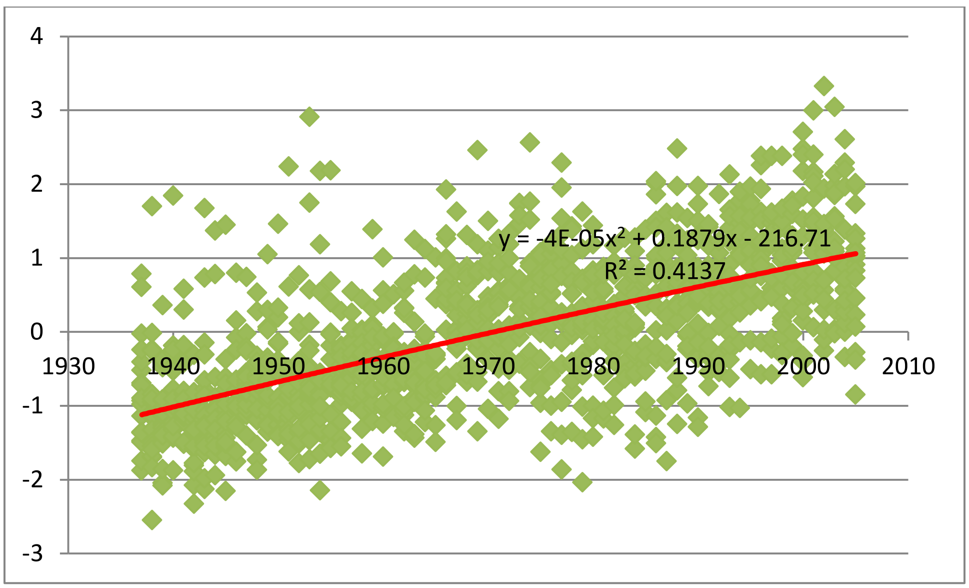


Figure 5 provides mean ipsative ACMR scores over time, and closely resembles findings in Figure 3 obtained for the strongest models. All means prior to 1965 were negative; between 1965 to 1985 all means exceeded zero and none exceeded 0.5; and all means except three since 1985 exceeded 0.5 (ESS=92.5; ESP=93.8).

A qualitative analysis of the scatterplot suggests means increased slowly between 1934 and 1965; had an abrupt increase of 0.5 to 1 SD in magnitude in 1965 and slowly degraded until 1985; and since 1985 assumed an uninterrupted linear rise until regressing in most recent years. The mean ipsative *z* score reached the value of 1.0 in 2000, and the two highest recorded means were seen in 2000 and in 2002.

Figure 6 presents data for counties with marginally higher ACMR ($p<0.10$, generalized criterion) after 1998 (4 yellow counties, Figure 1). Notably different than the prior findings, the regression model explained 31% of the variation in ipsative ACRM scores; crossed into positive *z* scores in 1964; never achieved a value of 1.0; and has been declining since 1990.

2155-0182/10/$3.00

Figure 5: Mean Ipsative ACMR Data for Counties Having *Significantly Greater* ACMR (*Generalized* Criterion) After 1998, Indicating 8-Year Forward Moving Average

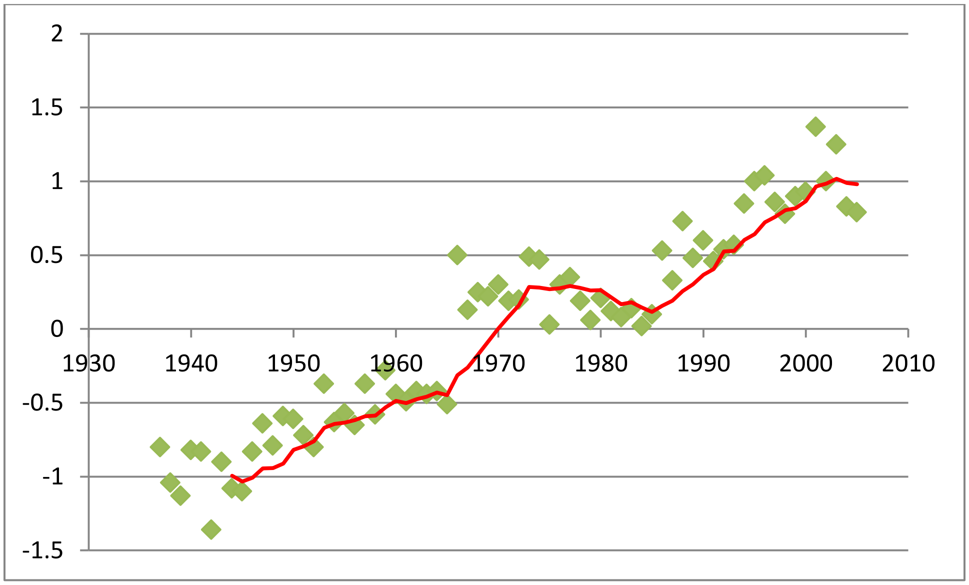


Figure 6: Ipsative ACMR Data for Counties Having *Marginally Greater* ACMR (*Generalized* Criterion) After 1998, Indicating Polynomial Regression Model

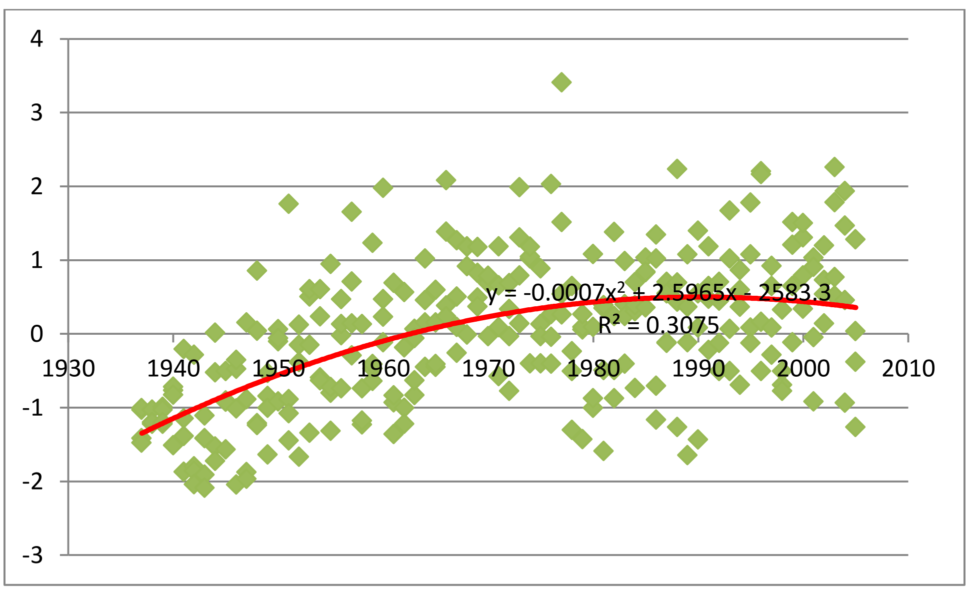
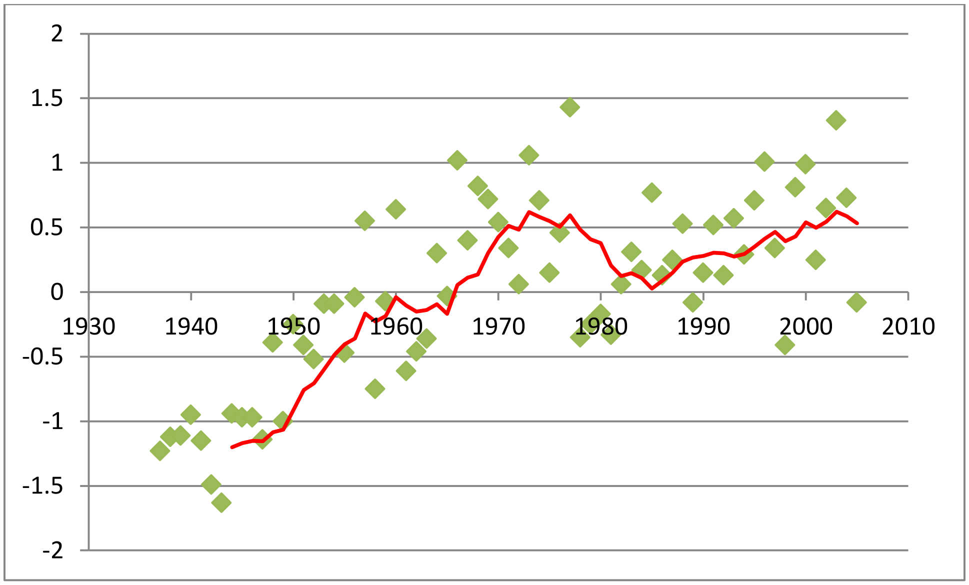


Figure 7 presents mean ipsative ACMR data over time, and is reminiscent of findings in Figure 5 obtained for stronger models, but with a weaker increase occurring since 1985. Means prior to 1954 were negative; high variability over time after 1954 persists to today; and there is no additional obvious partioning of the mean ipsative ACMR on the basis of year.

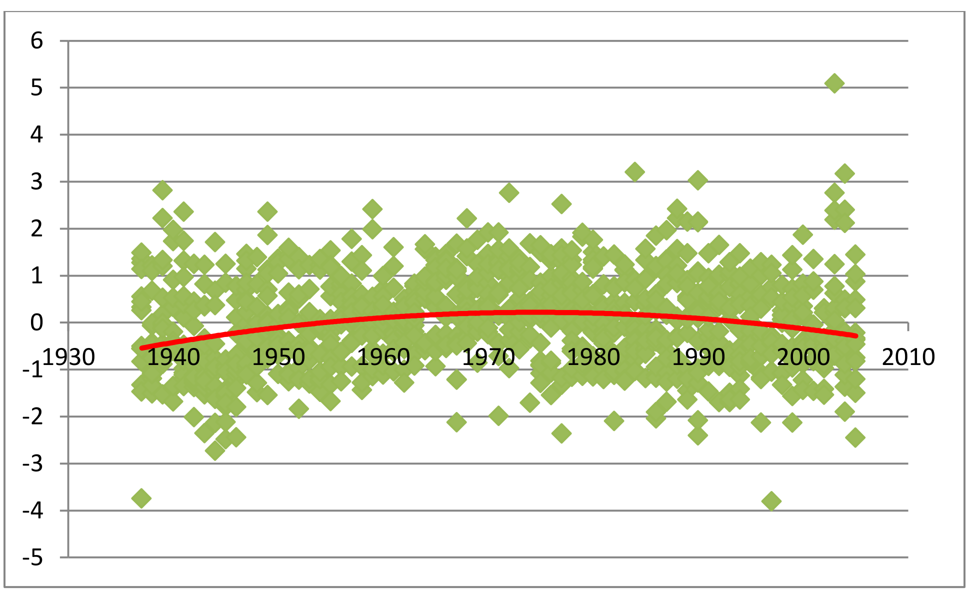
2155-0182/10/$3.00

Figure 7: Mean Ipsative ACMR Data for Counties Having *Marginally Greater* ACMR (*Generalized* Criterion) After 1998, Indicating 8-Year Forward Moving Average



Qualitative analysis of the scatterplot suggests means increased steadily between 1934 and 1975; declined 0.5 SD in magnitude over the next five years; and since 1985 had a steady rise of 0.5 SD, regressing most recently. Mean ipsative *z* score never reached 1.0. Three of the six greatest observed means occurred in 1997 and thereafter.

Figure 8: Ipsative ACMR Data for Counties *Not Having Greater* ACMR (*Generalized* Criterion) After 1998, Indicating Polynomial Regression Model

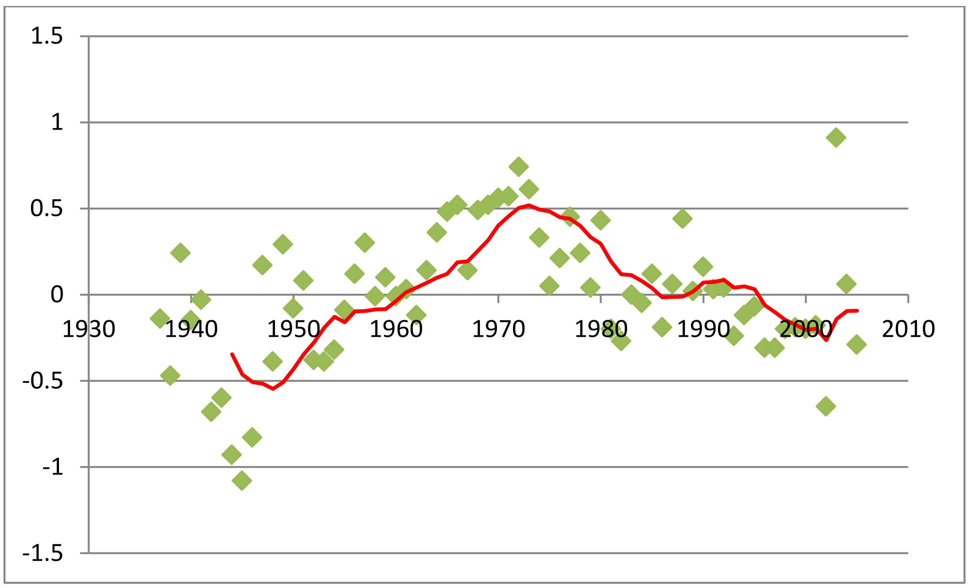


2155-0182/10/$3.00

Figure 8 gives data for counties that did *not* have a higher ACMR (p>0.10, generalized criterion) after 1998 (green counties in Figure 1). The regression model explained less than 5% of the variance in ipsative ACRM scores; stayed within a 0.5 SD range across time; never reached a value of 1.0; crossed into positive z scores in 1954; and has been declining since 1978.

Finally, Figure 9 presents mean ipsative ACMR data over time: it is reminiscent of the findings in Figure 7 for marginal models, but without an increase occurring since 1985. There is high variability across time, and no obvious partioning of mean ipsative ACMR on the basis of year. The greatest mean ever observed was recorded in 2002.

Figure 9: Mean Ipsative ACMR Data for Counties *Not Having Greater* ACMR (*Generalized* Criterion) After 1998, Indicating 8-Year Forward Moving Average



Findings obtained by examining the trajectory of ACMR across time separately for the four groups of counties identified in prior research are revealing. Analysis of the pattern of mean ipsative ACMR across time, performed for counties experiencing a *statistically significant increase* (whether the experimentwise or generalized criterion was applied) in ACMR after 1998, revealed the means fell into three almost perfectly discriminable groups: (a) *low* mean ipsative ACMR scores (mean z<0) were seen early in the series (1965 or earlier); (b) *medium* scores (0<mean $z\leq$0.5) were seen in the middle of the series (1966-1985); and (c) *high* scores (mean z>0) were seen late in the series (1986 or later). Series for both groups achieved mean z>1 in 1998 or soon after, and mean scores since 1998 have been among the highest on record. Findings for the counties experiencing a *statistically marginal increase* in ACMR after 1998 followed a reminiscent pattern, but with a greater increase in mean ipsative ACMR score in the early and middle portions of the series, and a more muted increase later in the series. In contrast, findings for the counties *not experiencing* a significant increase in ACMR after 1998 indicated that a peak mean ipsative score occurred in the middle of the series, with regression toward higher mean ipsative ACMR score occurring in recent years.

2155-0182/10/$3.00

Series for Individual Counties

Inspection of the scatter plots for each of the four county groupings, and consideration of meager performance achieved by the regression models, indicates substantial variability between counties and across time within each of the four groups. Accordingly, separate scatter plots are provided for each county, illustrating the 8-year forward moving average for ipsative ACMR.

Series showing ipsative ACMR scores over time are illustrated in Figure 10 for the counties having a *significantly greater* (experimentwise criterion) ACMR after 1998.[1] As seen, the series for McLean (A), Pierce (F), Bowman (G) Ramsey (H), Foster (I), Kidder (K), Logan (M), Ransom (N) and Grant (P) counties were generally consistent with the group series shown in Figures 2 and 3. Series for Stark (B), Burleigh (C), and Stutsman (J) counties rose slightly and then declined through the early and middle years, rising in the later years. And, series for Hettinger (D), Adams (E), Emmons (L), and McIntosh (O) counties gained continuously over years.

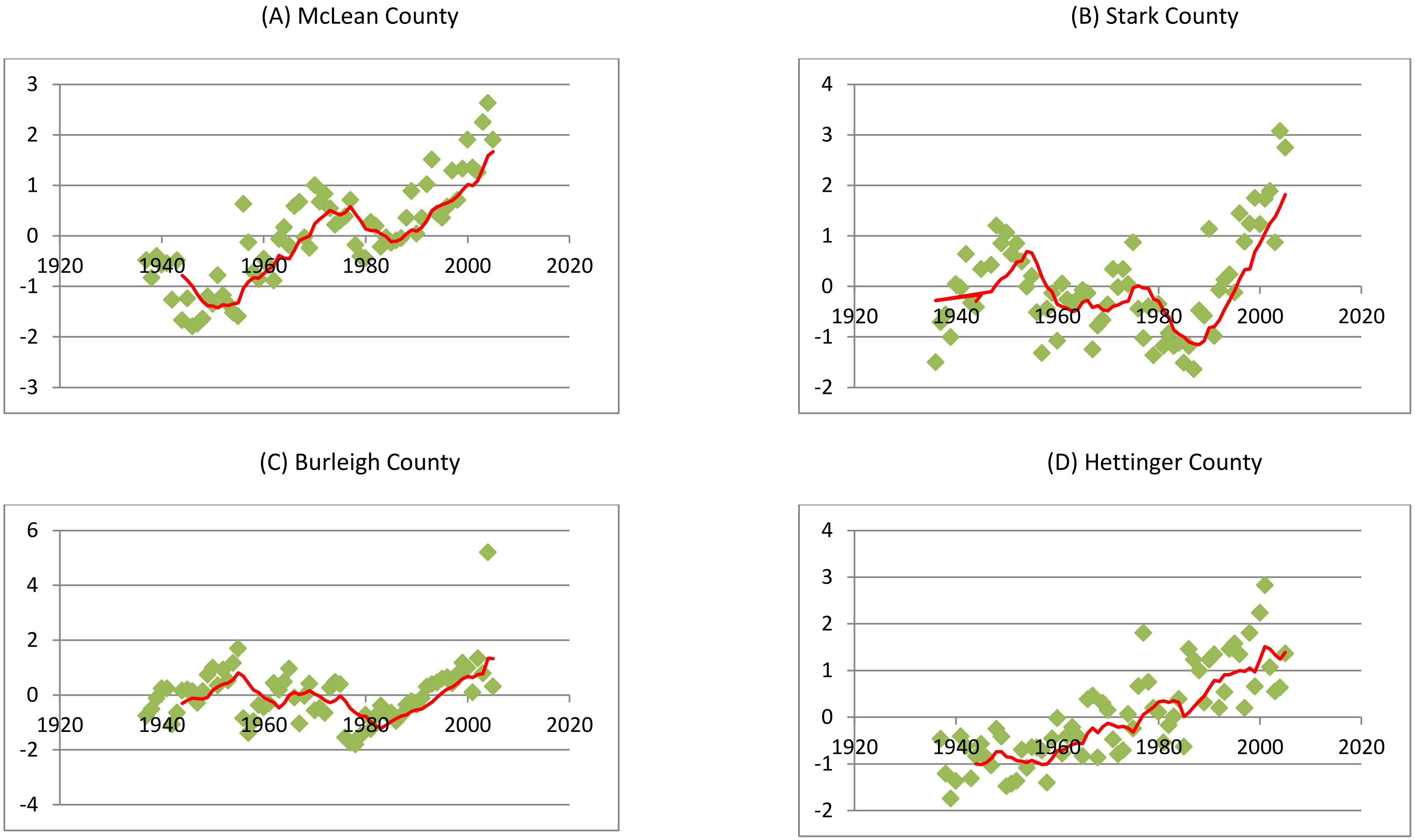
Series showing ipsative ACMR scores over time are illustrated in Figure 11 for counties having a *significantly greater* (generalized criterion) ACMR after 1998.[1] As seen, the series for Divide (A), Dunn (B), Cavalier (H), Nelson (K), Griggs (L), and Lamoure (Q) counties were generally consistent with the group series shown in Figures 4 and 5, as were the series for Morton (N) and Mercer (O) counties, although scores in these series decreased modestly in the middle years. Series for Mountrail (D), Towner (F), Bottineau (G), Walsh (I), Eddy (J), Wells (M), Dickey (P), and Barnes (R) counties rose continuously over years. The series for Ward (E) county was U-shaped, and the series for Williams (C) county was reminiscent of a cosine wave across years.

Series illustrating ipsative ACMR scores across time are given in Figure 12 for counties with *marginally greater* (generalized criterion) ACMR after 1998.[1] As seen, series for Renville (A), McKenzie (B), and Sheridan (C) counties were generally consistent with the group series presented in Figures 6 and 7. In contrast, the series for Traill (D) county gained continuously across years.

Finally, series showing ipsative ACMR scores across time are illustrated in Figure 13 for counties *not having* a significantly greater (generalized criterion) ACMR after 1998.[1] As seen, series for Golden Valley (A), Slope (B), Oliver (E), Benson (I), Pembina (K), Sargent (N) and Richland (O) counties were generally consistent with the group series shown in Figures 8 and 9, as did the series for Burke (D) county, although the decline in the middle period was delayed. In contrast, series for Billings (C), Sioux (G), Rolette (H), Grand Forks (J), and Cass (M) counties fell across years, regressing to higher scores recently. And, the series for McHenry (F) and Steele (L) counties rose across years, with recent regression to lower levels.

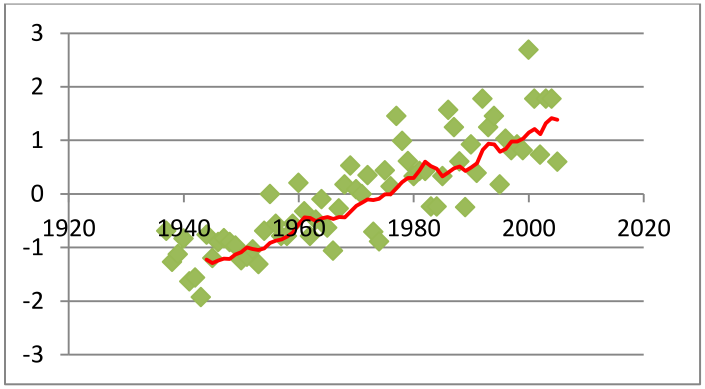
2155-0182/10/$3.00

Figure 10: Ipsative ACMR Data for Counties Having *Significantly Greater* ACMR (*Experimentwise* Criterion) After 1998, Indicating 8-Year Forward Moving Average

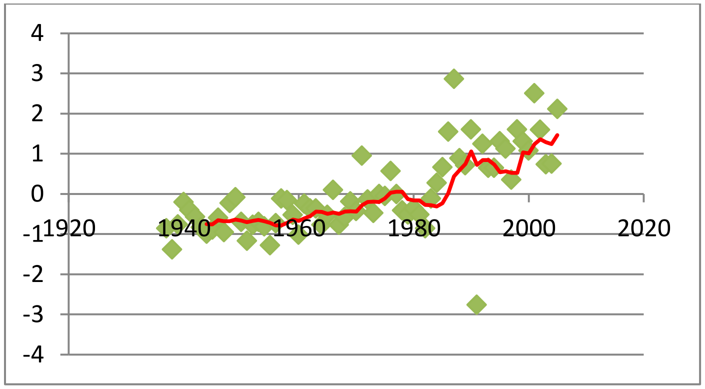


2155-0182/10/$3.00

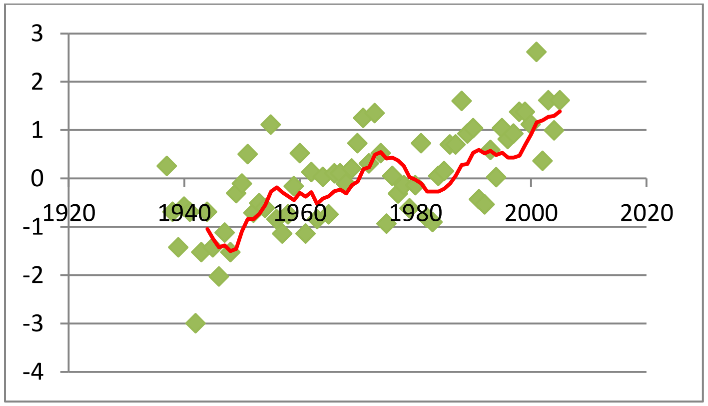
(E) Adams County



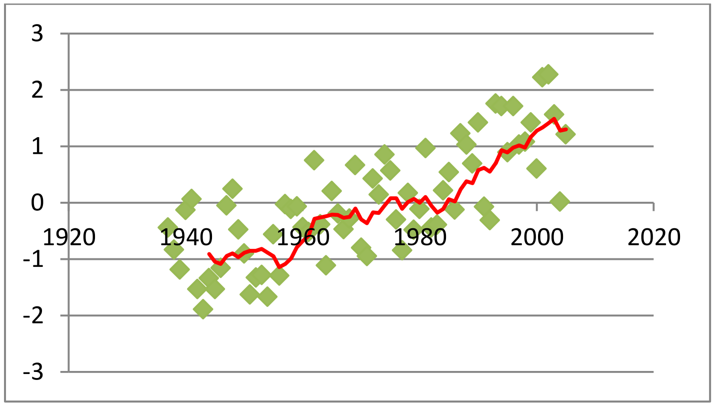
(F) Pierce County



(G) Bowman County

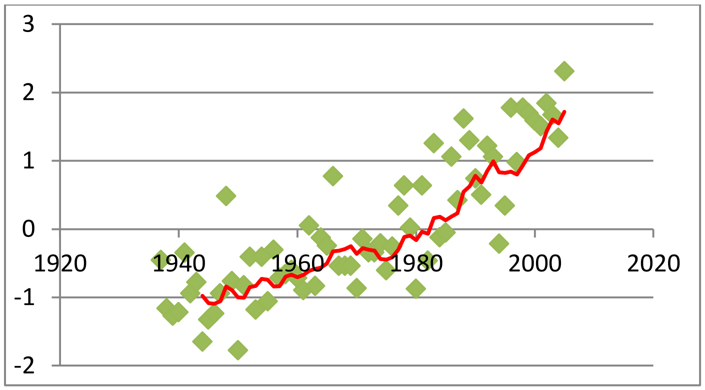


(H) Ramsey County

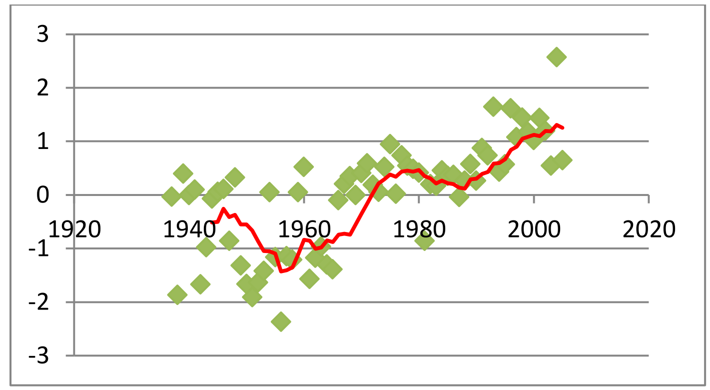


2155-0182/10/$3.00

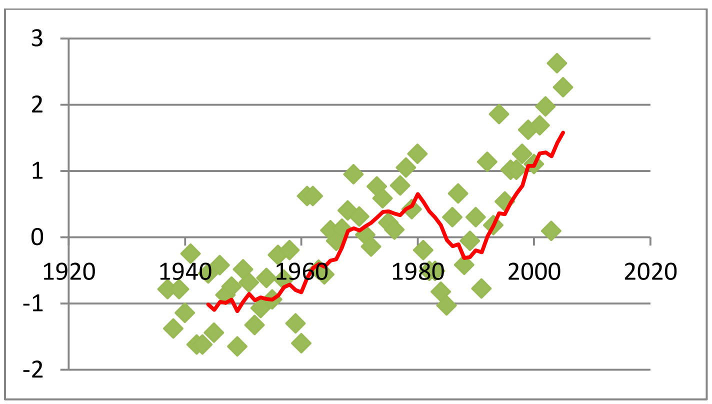
(I) Foster County



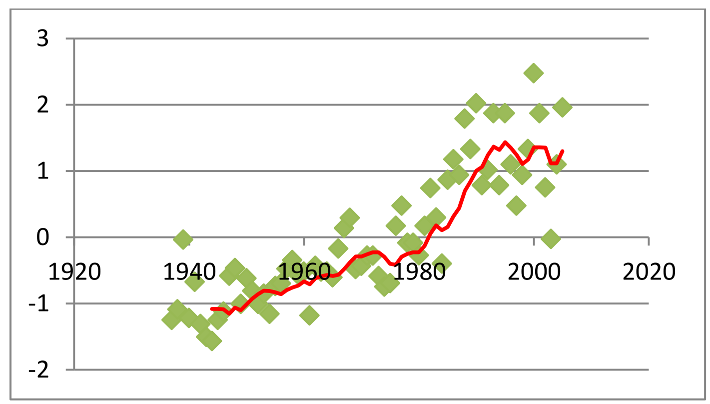
(J) Stutsman County



(K) Kidder County



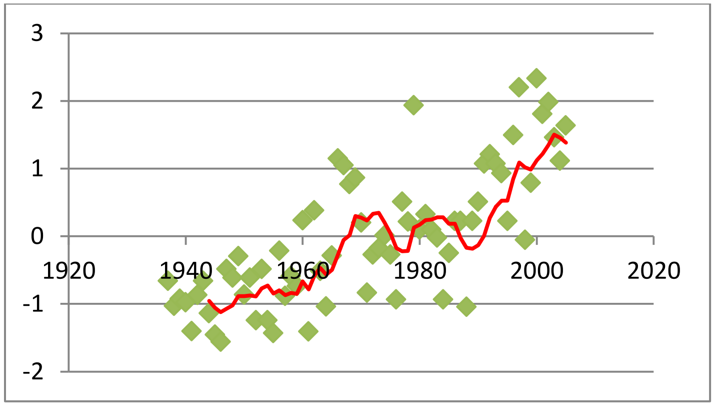
(L) Emmons County



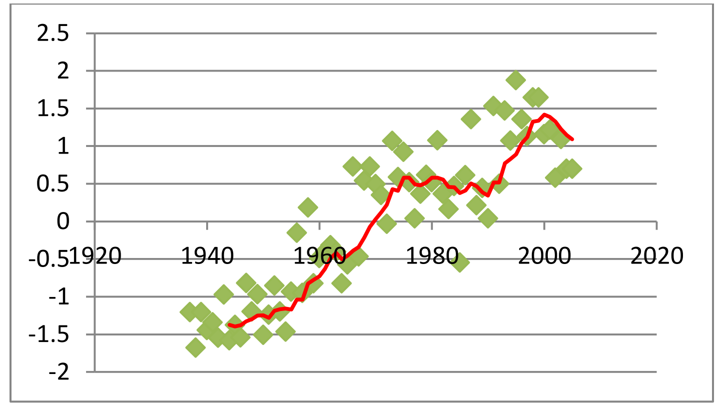
Optimal Data Analysis
Vol. 2, Release 2 (October 30, 2013), 120-142

2155-0182/10/$3.00

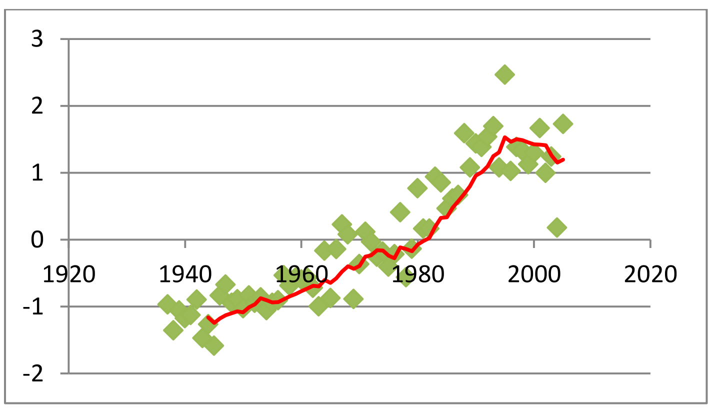
(M) Logan County



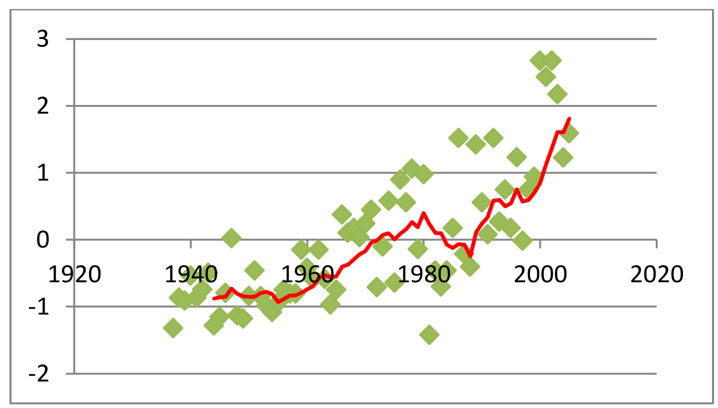
(N) Ransom County



(O) McIntosh County

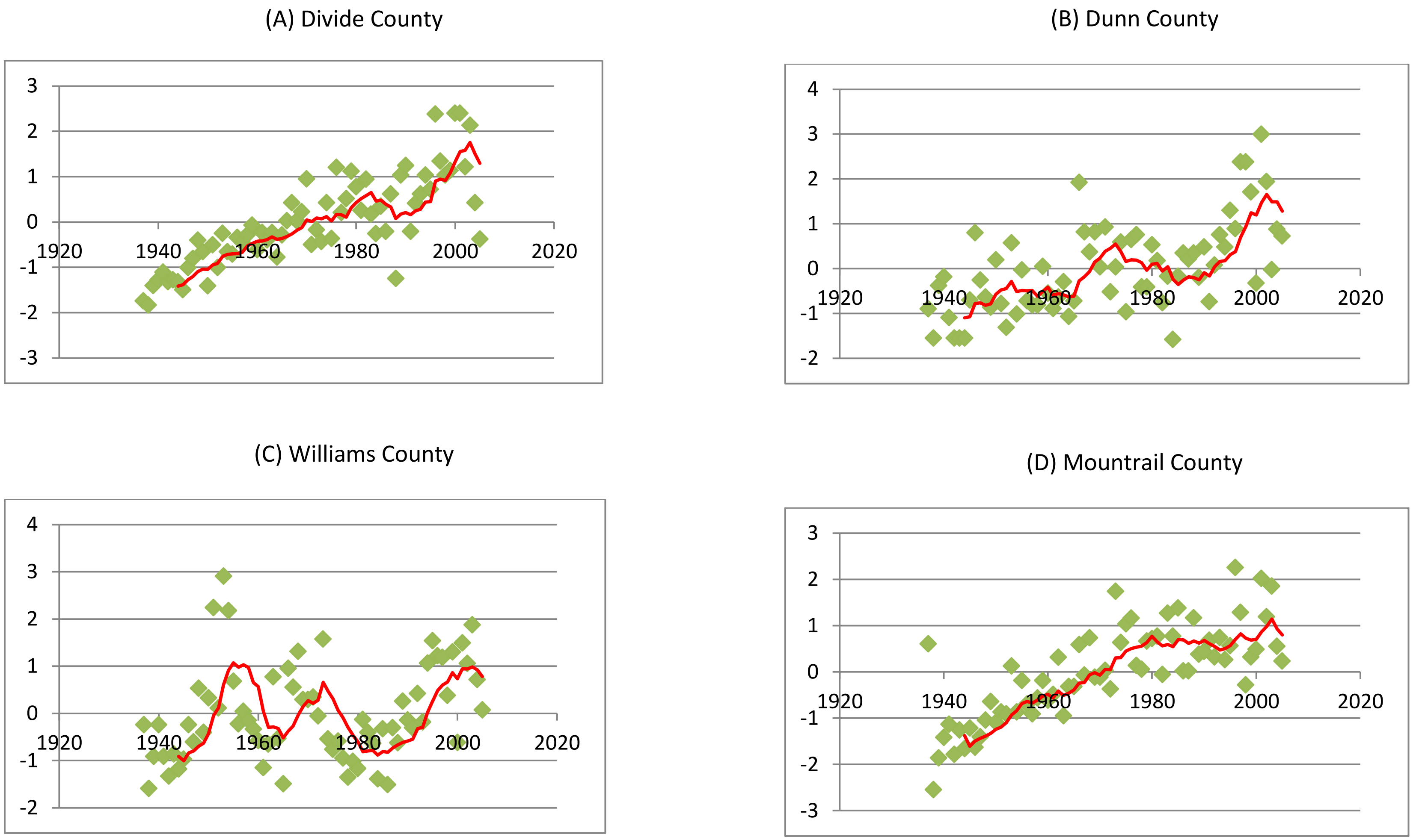


(P) Grant County



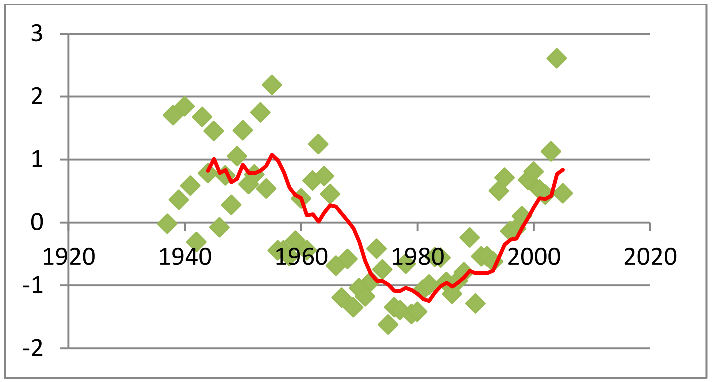
2155-0182/10/$3.00

Figure 11: Ipsative ACMR Data for Counties Having *Significantly Greater* ACMR (*Generalized* Criterion) After 1998, Indicating 8-Year Forward Moving Average

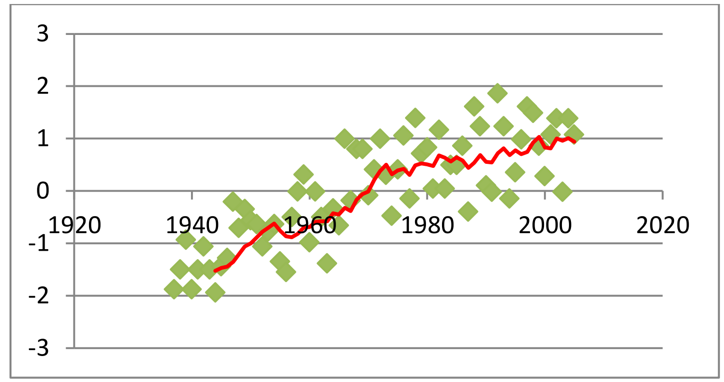


2155-0182/10/$3.00

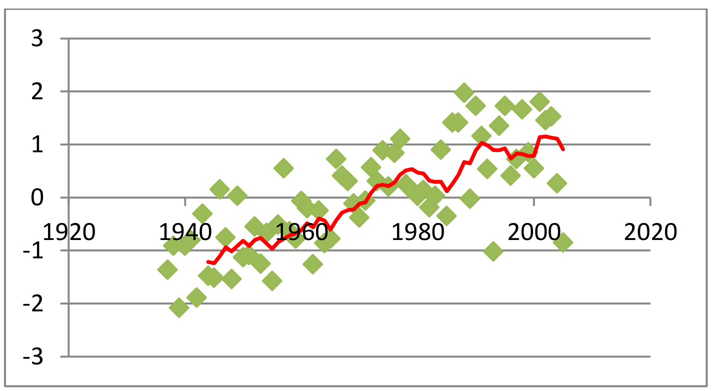
(E) Ward County



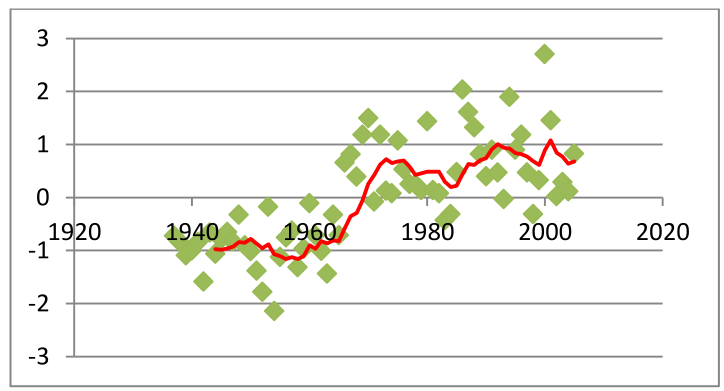
(F) Towner County



(G) Bottineau County

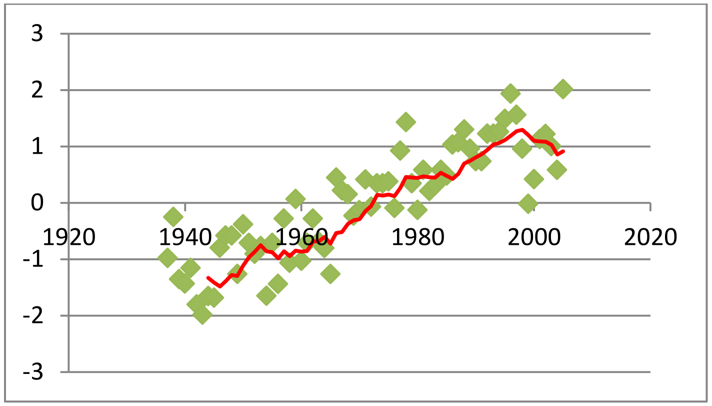


(H) Cavalier County

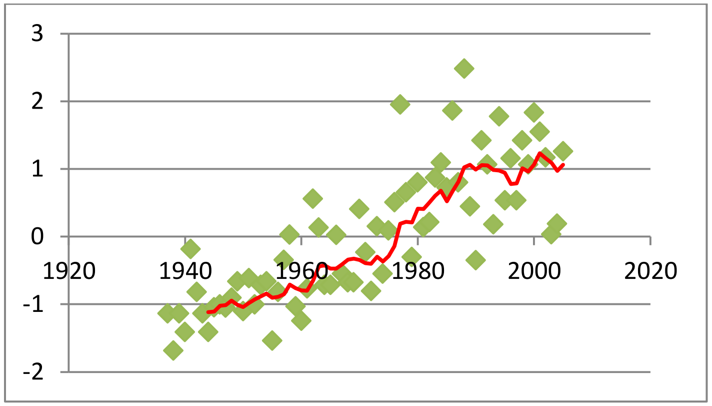


2155-0182/10/$3.00

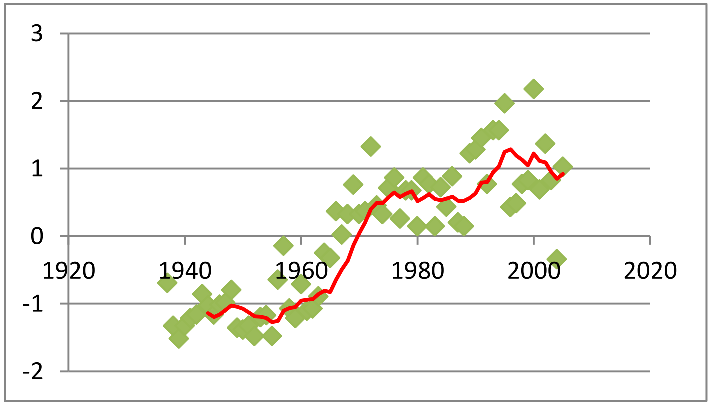
(I) Walsh County



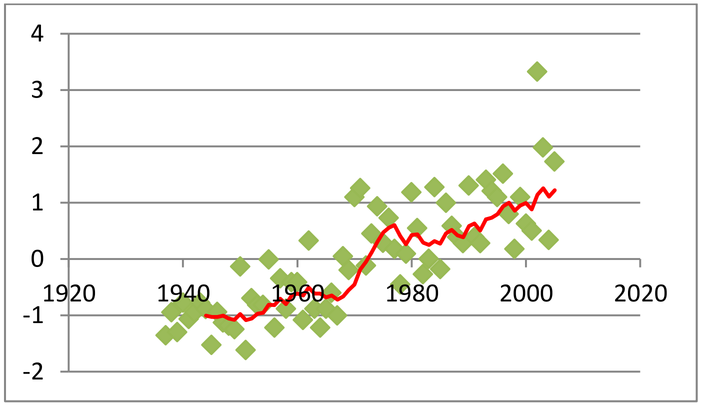
(J) Eddy County



(K) Nelson County

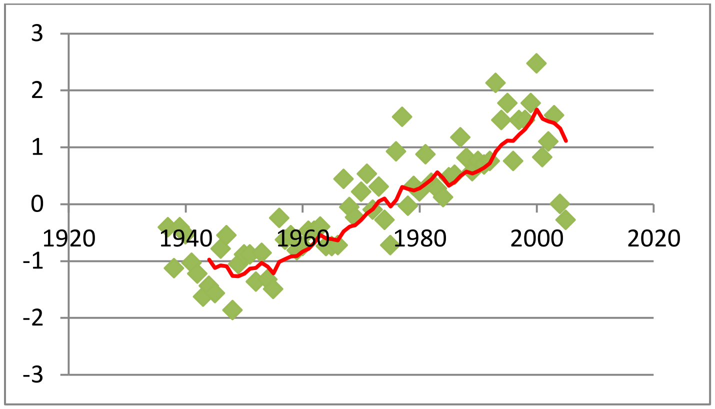


(L) Griggs County

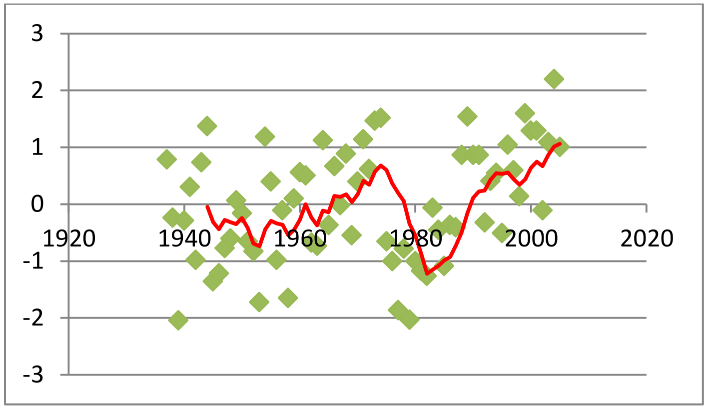


2155-0182/10/$3.00

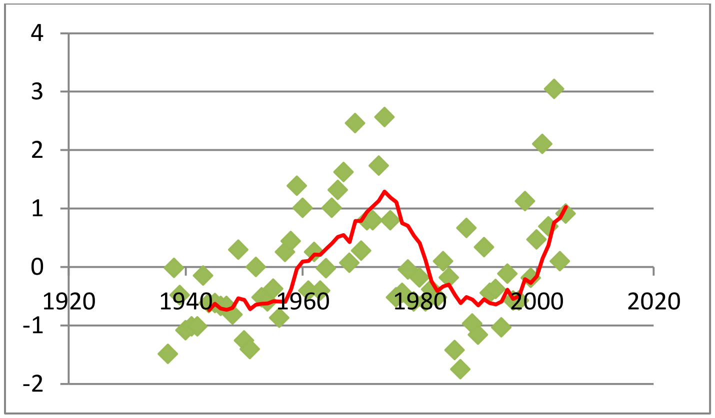
(M) Wells County



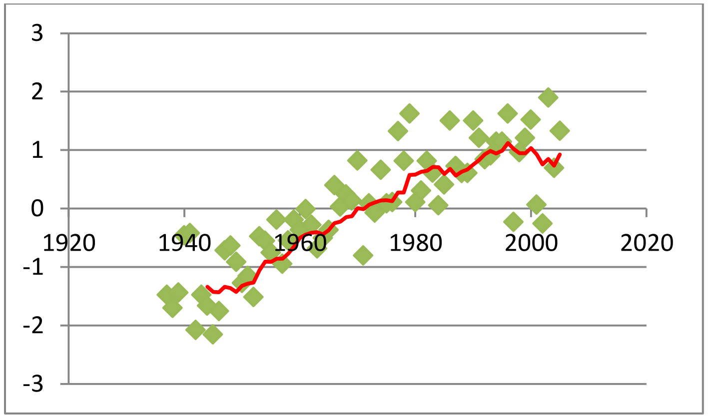
(N) Morton County



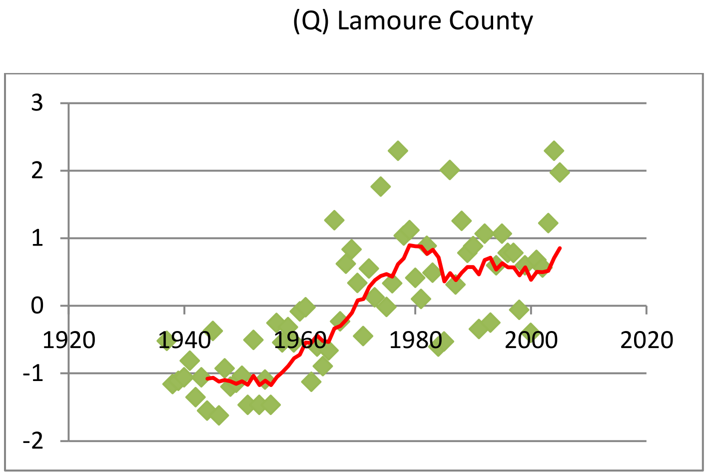
(O) Mercer County

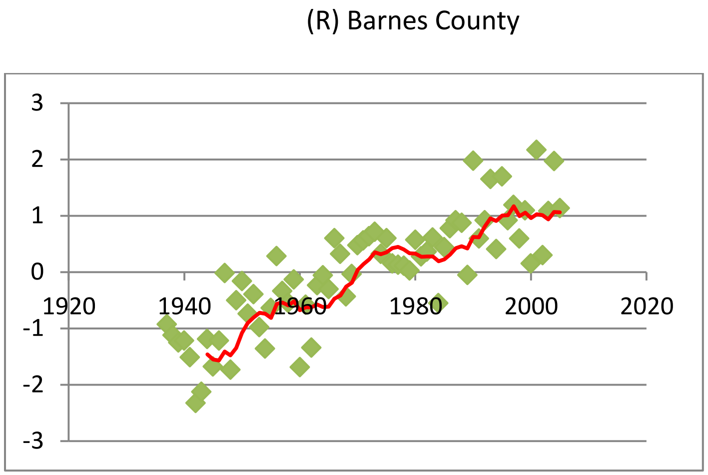


(P) Dickey County



2155-0182/10/$3.00



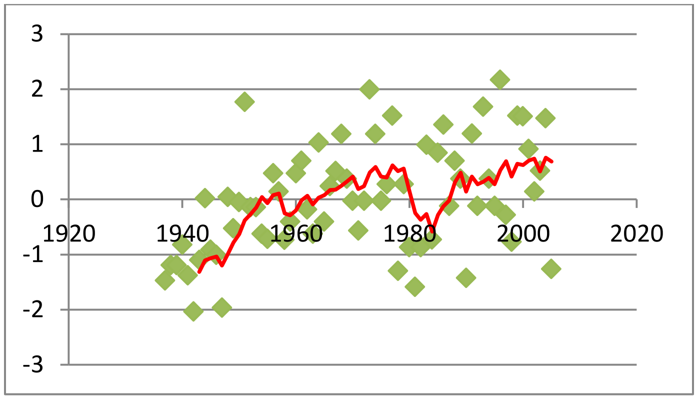


Optimal Data Analysis
Vol. 2, Release 2 (October 30, 2013), 120-142

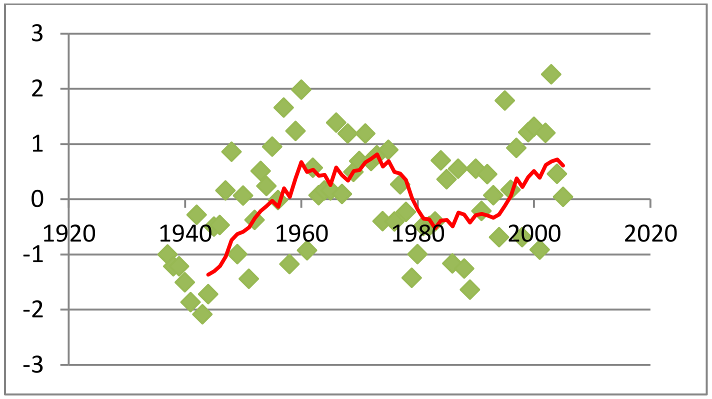
2155-0182/10/$3.00

Figure 12: Ipsative ACMR Data for Counties Having *Marginally Greater* ACMR (*Generalized* Criterion) After 1998, Indicating 8-Year Forward Moving Average

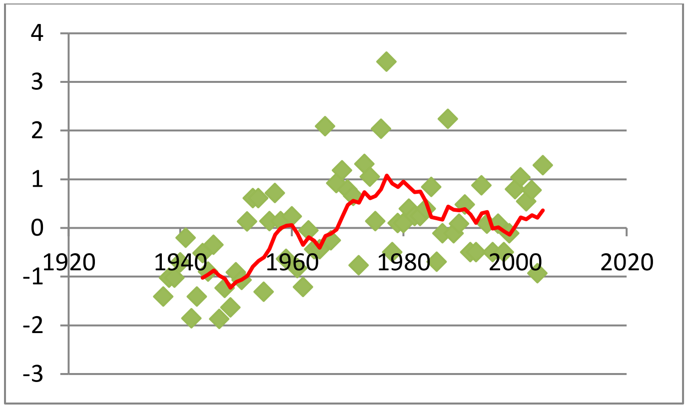
(A) Renville County



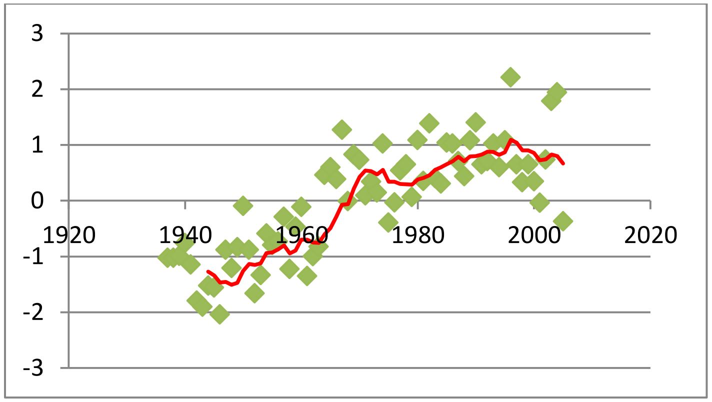
(B) McKenzie County



(C) Sheridan County



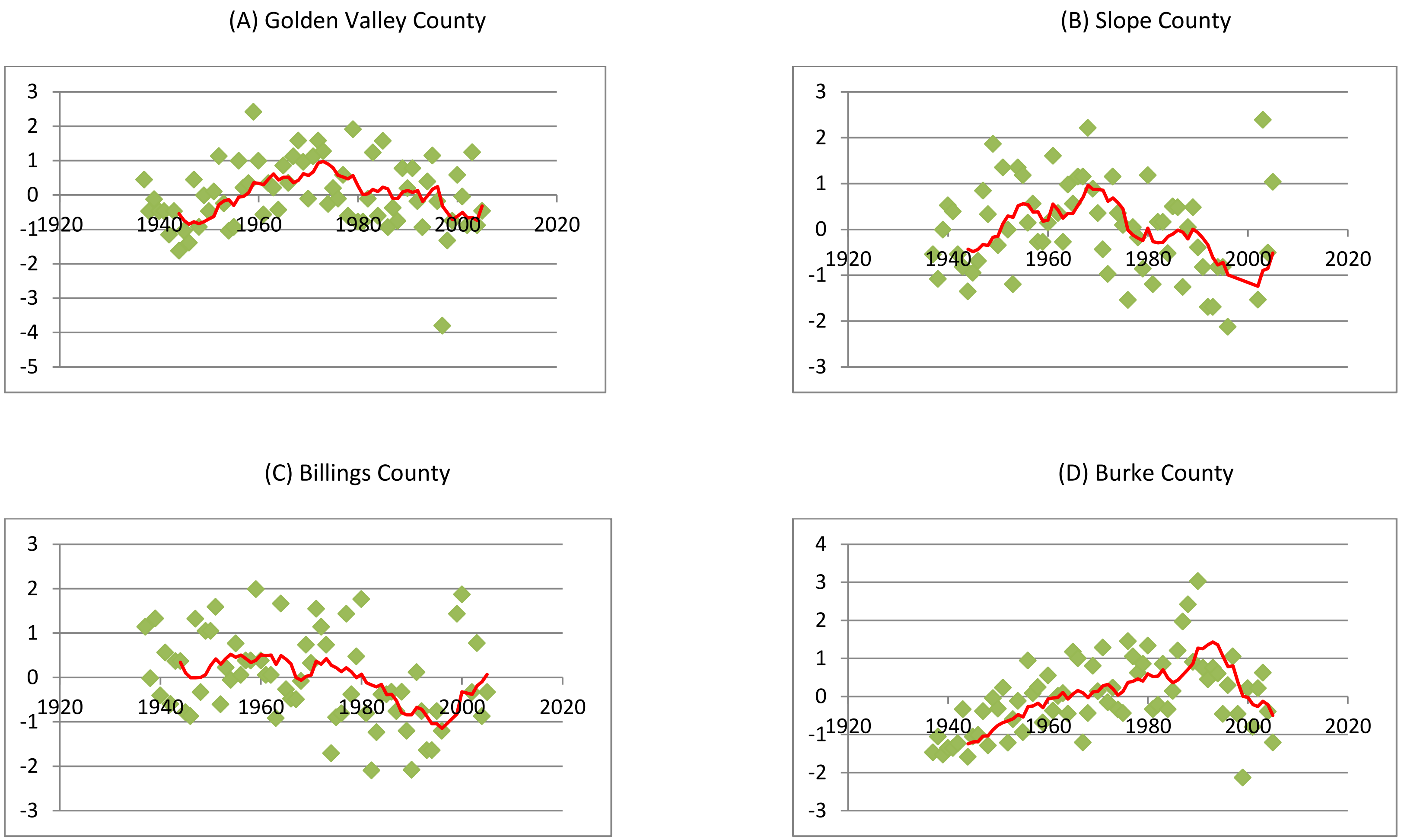
(D) Traill County



Optimal Data Analysis
Vol. 2, Release 2 (October 30, 2013), 120-142

2155-0182/10/$3.00

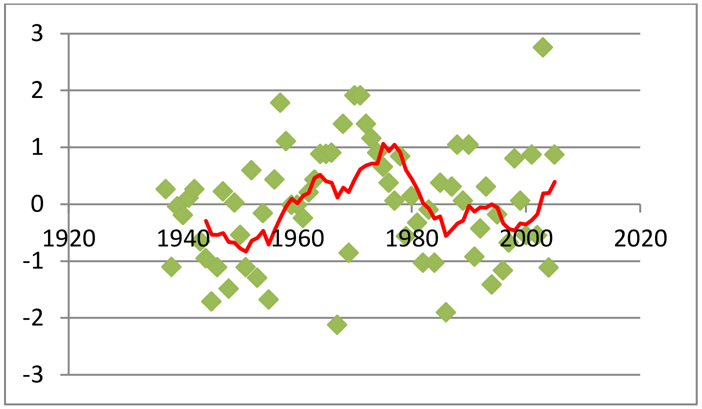
Figure 13: Ipsative ACMR Data for Counties *Not Having* Greater ACMR (*Generalized* Criterion) After 1998, Indicating 8-Year Forward Moving Average



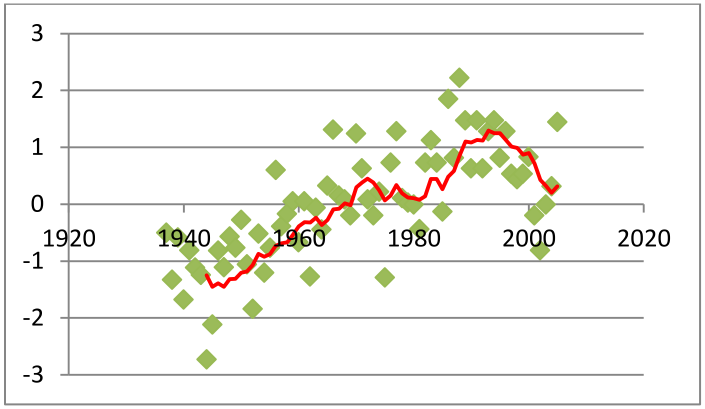
Optimal Data Analysis
Vol. 2, Release 2 (October 30, 2013), 120-142

2155-0182/10/$3.00

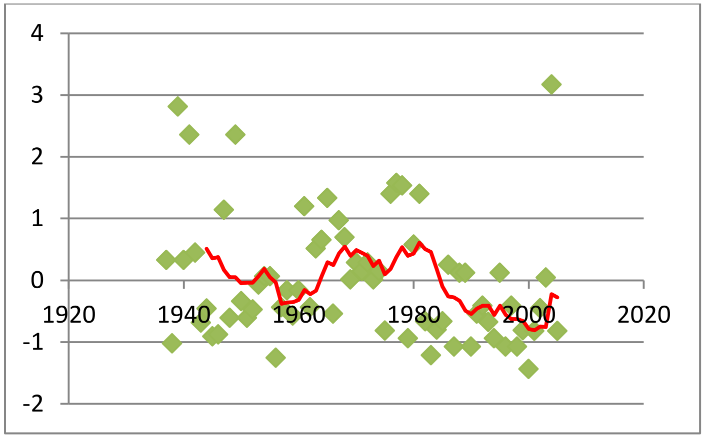
(E) Oliver County



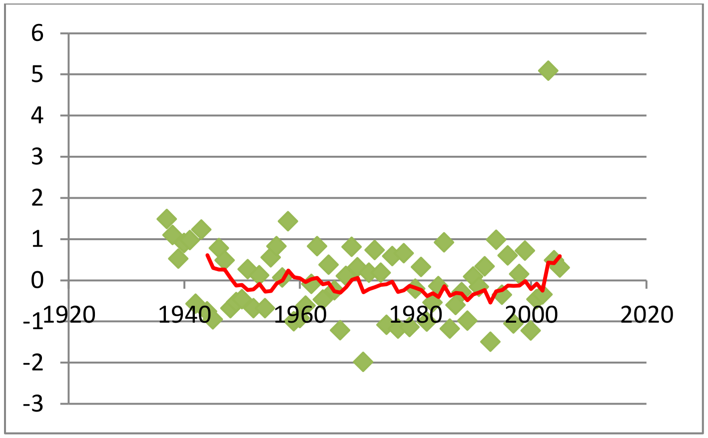
(F) McHenry County



(G) Sioux County



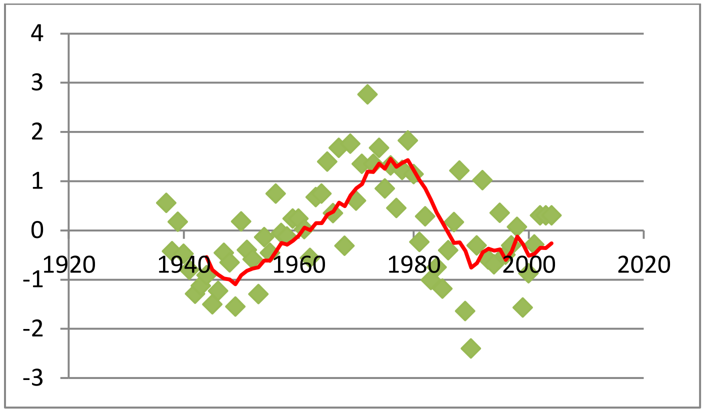
(H) Rolette County



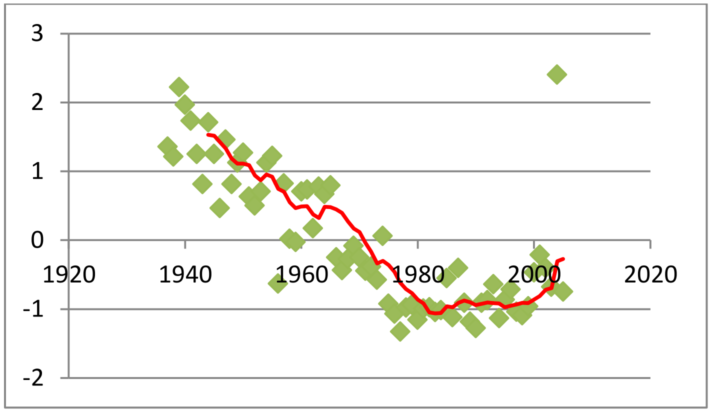
Optimal Data Analysis
Vol. 2, Release 2 (October 30, 2013), 120-142

2155-0182/10/$3.00

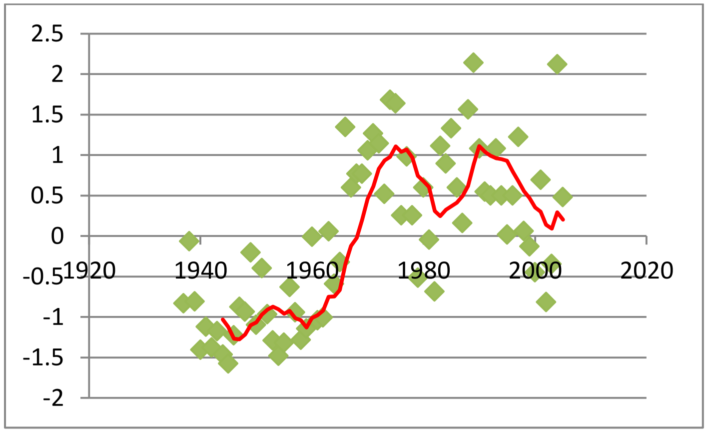
(I) Benson County



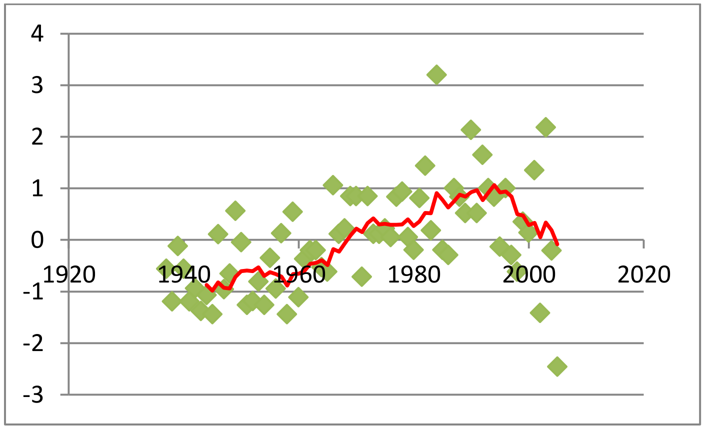
(J) Grand Forks County



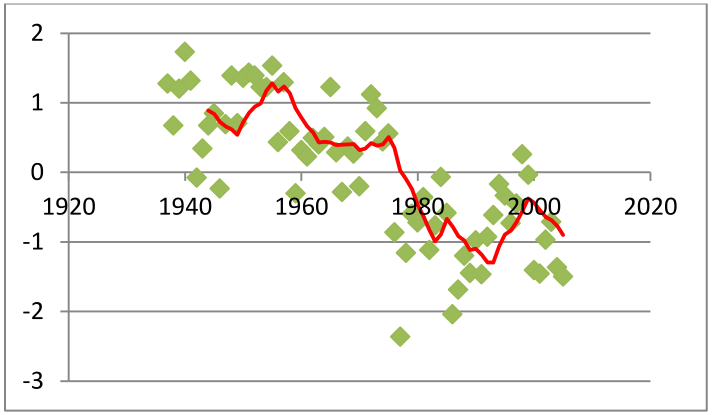
(K) Pembina County



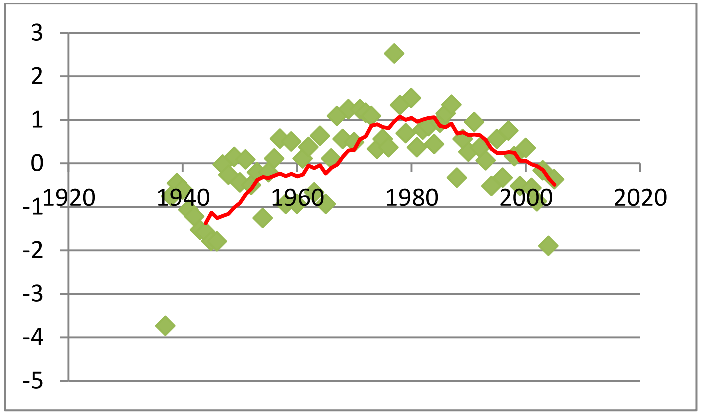
(L) Steele County



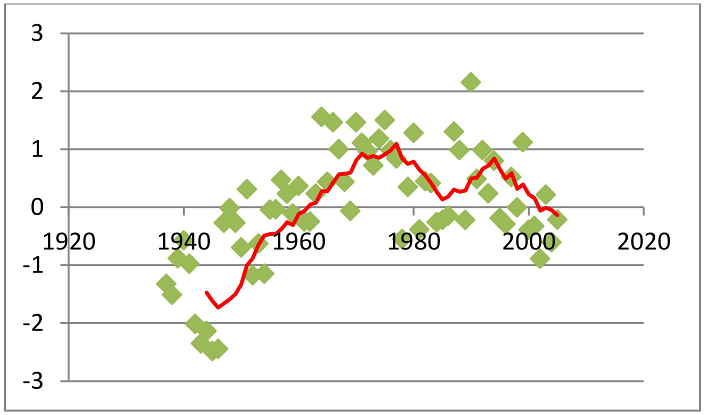
(M) Cass County



(N) Sargent County



(O) Richland County



2155-0182/10/$3.00

Summary

Results identify many counties in North Dakota having ipsative ACMR mean scores rapidly increasing in time. Means began to exceed 1.0 in value in 1998, and many of the highest-ever recorded values of this statistic have been observed since then.

County series presented herein serve as the class (or "dependent") variable for planned research, the first thrust of which depends for success upon being able to obtain theoretically motivated data to serve as *attributes* (or "independent variables") for multivariable statistical analysis. Data that may be used as attributes are thus far ellusive.[1] For example, statistics on age by county are typically given for categories such as "percent of population less than 18 years of age," and often data are available only for a limited number of years. As the hunt for age and other theoretically-motivated attributes continues, the ODA laboratory is constructing new variables, such as proximity to production areas—including distance, wind, and water table-based measures, which may help to explain the changes in ipsative ACMR scores over time observed in each counties series. Thorough retrospective analysis of the causes of mortality occurring in each county over its series is warranted, as is prospective validation of the present research.

The second thrust of the planned research involves exporting these methods to study these phenomena in all conceptually similar, heavily-mined areas of the world for which data are available. The objective is assessment of cross-generalizability of the phenomena: it is important to understand which factors are consistent across samples, and which factors are sample-specific, in the past, now, and in the future.

References

[1]Yarnold PR (2013). Statistically significant increases in crude mortality rate of North Dakota counties occurring after massive environmental usage of toxic chemicals and biocides began there in 1997: An optimal *static* statistical map. *Optimal Data Analysis*, 2, 98-105.

[2]Yarnold PR, Soltysik RC (2005). *Optimal data analysis: Guidebook with software for Windows*. Washington, D.C.: APA Books.

[3]http://ndhealth.gov/vital/stats.htm

[4]Yarnold PR (1996). Characterizing and circumventing Simpson's paradox for ordered bivariate data. *Educational and Psychological Measurement*, *56*, 430-442.

[5]Bryant FB, Siegel EKB (2010). Junk science, test validity, and the Uniform Guidelines for Personnel Selection Procedures: The case of *Melendez v. Illinois Bell*.*Optimal Data Analysis*, *1*, 176-198.

[6]Soltysik RC, Yarnold PR (2010).The use of unconfounded climatic data improves atmospheric prediction. *Optimal Data Analysis*, *1*, 67-100.

[7]Yarnold PR, Soltysik RC (2013). Ipsative transformations are *essential* in analysis of serial data. *Optimal Data Analysis*, 2, 94-97.

[8]Licht MH. Multiple regression and correlation. In: Reading and understanding multivariate statistics. APA Books, Washington, DC, 1995, pp. 19-64.

[9]Yarnold PR, Bryant FB, Soltysik RC (2013). Maximizing the accuracy of multiple regression models via UniODA: Regression *away from* the mean. *Optimal Data Analysis*, *2*, 19-25.