JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 94, NO. B6, PAGES 7673–7677, JUNE 10, 1989

# Reply to "Comment on 'Saugus-Palmdale, California, Field Test for Refraction Error in Historical Leveling Surveys' by R. S. Stein, C. T. Whalen, S. R. Holdahl, W. Strange, and W. Thatcher" by Michael R. Craymer and Peter Vaníček and Comment on "Further Analysis of the 1981 Southern California Field Test for Leveling Refraction" by M. R. Craymer and P. Vaníček

ROSS S. STEIN

*U.S. Geological Survey, Menlo Park, California*

CHARLES T. WHALEN, SANFORD R. HOLDAHL, AND WILLIAM E. STRANGE

*National Geodetic Survey, Charting and Geodetic Services, Rockville, Maryland*

WAYNE THATCHER

*U.S. Geological Survey, Menlo Park, California*

## INTRODUCTION

In 1981 we conducted a field test of atmospheric refraction error in historical leveling surveys and found that refraction was 6 times larger than random error and could be modeled and removed [*Stein et al.*, 1986]. *Craymer and Vaníček* [1986] performed a multiple linear regression on the 1981 data; in addition to confirming the refraction error, they reported equally large experimental errors associated with the number of turning points (or rod supports) and the height difference between bench marks. In response, we pointed out [*Stein et al.*, 1986] that their regression was not robust; by design, the number of turning points was positively correlated with height difference. When two out of the 60 observations are removed from the sample, the two independent variables are correlated with each other at the 99.9% confidence level. In their comment, *Craymer and Vaníček* [this issue] argue that no observations may be removed from the sample and that in any case the correlation between the independent variables is too small to invalidate their results. We now agree that the correlation does not prohibit the multiple linear regression, but parameters that depend on two points are suspect, prompting us to examine their regression further. Here we report what we believe to be the fundamental defect in Craymer and Vaníček's analysis, that the dependent variable (divergence) is uncorrelated with either of the two disputed independent variables (turning points and height), resulting in meaningless regression coefficients. When the two influential points are removed to test the stability of the regression, the height dependence disappears. When the intercept in the regression is not constrained to pass through the origin, the dependence on the number of turning points also vanishes, leaving only the refraction.



Paper number 89JB00355.

## ANALYSIS

### *Regression of Uncorrelated Data*

Although *Craymer and Vaníček*'s [1986] regression coefficients $b$ appear highly significant, the partial correlation coefficients for the turning-point and height-difference arguments are miniscule ($r = -0.06$ and $-0.08$, respectively; see our Table 1, which shows values listed by *Craymer and Vaníček* [1986, p. 9503] in their Table 2, column 3). This means that neither the number of turning points nor the height differences are correlated with the divergence. In a simple regression, the slope $b$ and correlation $r$ are related by

$$r = b\,\frac{S_x}{S_y} \qquad (1)$$

where $S_x$ and $S_y$ are the standard deviations in $x$ and $y$, respectively [*Draper and Smith*, 1981, p. 45]. Thus $b$ is a version of $r$ scaled by the spread in the data. The correlation coefficient measures the linear association between $x$ and $y$, while the regression coefficient measures the predicted change in $y$ for unit change in $x$. From (1) it can be seen that the coefficients $r$ and $b$ should, in general, display the same sign and a similar level of significance. Such a relation, however, holds only for refraction, which explains why regression on the number of turning points and height difference explains 9% of the variance in the data [*Craymer and Vaníček*, 1986, Table 2], whereas when refraction is included 60% of the variance is explained (Table 1). A simple regression on refraction alone explains 56% of the variance, which means that the other two arguments reduce the variance by only 4% in the regression on refraction, turning points, and height. This confirms the central error in their analysis, as regression of variables which have no linear association is illusory. The only remaining question is, what went wrong?

TABLE 1. Multiple Linear Regression on 1981 Saugus-Palmdale Leveling Data

| | | | *Craymer and Vaníček* [1986, this issue] $n = 60$; Intercept $\equiv 0$ | | | | This Reply $n = 58$; Free Intercept | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Partial Regression $b$ | | Partial Correlation $r$ | | Partial Regression $b$ | | Partial Correlation $r$ | |
| | Units | Cumulative Magnitude, mm | Coefficient | Significance, % | Coefficient | Significance, % | Coefficient | Significance, % | Coefficient | Significance, % |
| Argument | | | | | | | | | | |
| Refraction | °C $m^3 \times 10^{-6}$ | +16 | 44 ± 5 | 99.99 | 0.74 | 99.99 | 43 ± 5 | 99.99 | 0.75 | 99.99 |
| Turning points | mm$\times 10^{-3}$ | +21 | **14 ± 6** | **96** | **−0.06** | **35** | −18 ± 19 | 66 | −0.08 | 55 |
| Height difference | $\times 10^{-6}$ | −15 | **−28 ± 12** | **97** | **−0.08** | **48** | −12 ± 22 | 41 | −0.10 | 66 |
| Variance explained by model, % | | | | 59.3 | | | | 59.5 | | |

Bold numbers highlight the principle inconsistency of *Craymer and Vaníček*'s [1986, this issue] analysis: regression coefficients that appear highly significant associated with insignificant correlation coefficients. When the intercept is not constrained to pass through the origin and two influential points are removed, the regression and correlation coefficients for turning points and height are seen to have the same sign and low level of significance, with the refraction coefficients unchanged (right columns). The unstable arguments do not increase the variance explained by the model, further confirmation that refraction is the only robust argument. Values for Craymer and Vaníček differ slightly from those listed in their paper but were calculated using a nearly identical data set.

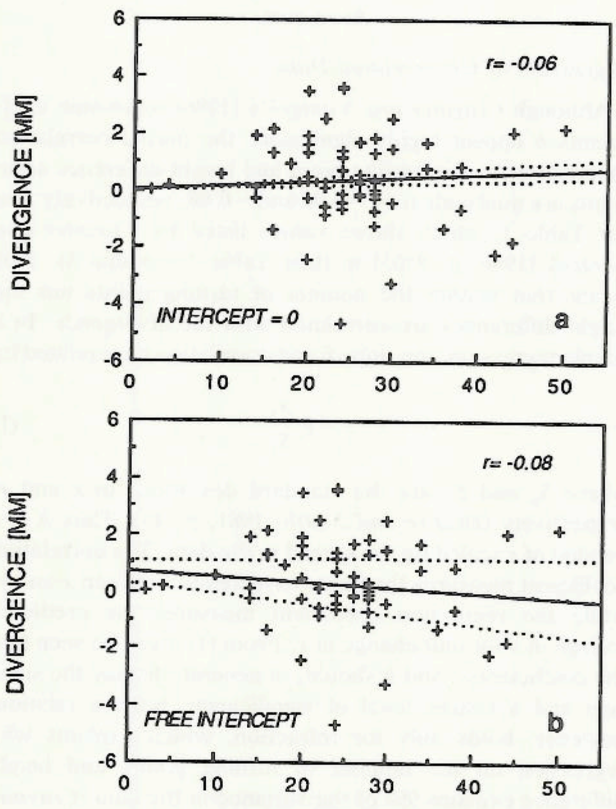


Fig. 1. Regression of observed divergence per section (or pair of bench marks), that is, the difference between the height measured on the forward run minus that measured on the backward run, on the number of turning points per section, showing slope $b$ (solid line) and $\pm 1\sigma$ envelope (dotted). (*a*) *Craymer and Vaníček* [1986] fit the data with line constrained to pass through the origin. (*b*) When the intercept is unconstrained, the slope is not significantly different from zero. Note that, irrespective of the choice of intercept, the data are uncorrelated ($r$).

### *Turning Point Argument*

*Craymer and Vaníček* [1986] set the intercept in their regression equal to zero. This forces the regression line to pass through the origin and, approximately, the mean of the data, rather than letting the regression fit a trend through the observations. Craymer and Vaníček have thus assumed that the number of turning points correlates with divergence, rather than testing such a correlation. This constraint would have little impact if the data clustered near the origin; however, the mean number of turning points per section (25 ± 9) lies three standard deviations from zero. The effect of constraining the intercept can be seen in the plots of the divergence against the number of turning points (Figure 1). When Craymer and Vaníček's proffered regression line and its uncertainty are used (Figure 1*a*), only 7% of the observations lie within the dotted $\pm 1\sigma$ envelope, an impermissible result. Figure 1*b* shows the regression on the same data when the intercept is not imposed; neither the slope nor the intercept differs significantly from zero. That the data clustering is not an artifact of projection on the divergence–turning point plane can be seen in the perspective plot of divergence on turning point and height (Figure 2*a*).

*Craymer and Vaníček* [1986] report an intercept ("bias") that is significant at the 94–99.5% level of confidence in their preliminary "discrepancy series analysis" [*Craymer and Vaníček*, 1986 Table 1, column 3, p. 9050], but they inexplicably drop it from their multiple linear regression with this remark.

> Note that the above **B** matrix does not provide for bias estimation even though it appears significant from the analyses of individual effects. If this were desired, there would be an extra column of ones. Because we have (as yet) no physical explanation for this constant effect, we have somewhat arbitrarily omitted it in our analyses.

The intercept is simply the mean divergence per section not attributable to the tested error sources; it is poorly deter-

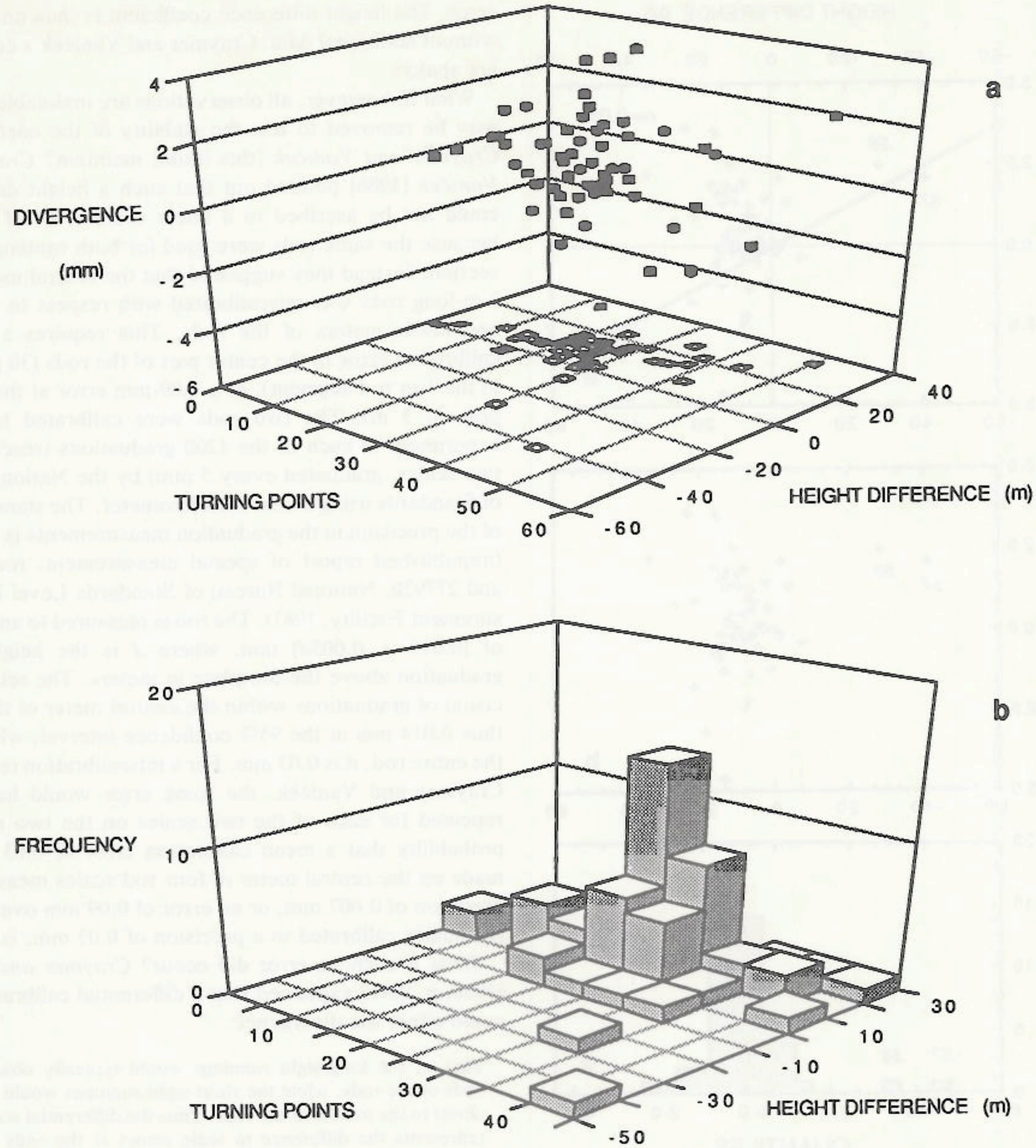


Fig. 2. Perspective diagrams of (*a*) the observed divergence and (*b*) frequency, or number of observations on turning points and height difference. In Figure 2*a* the points (solid squares) cast shadows (open squares) vertically downward on the turning point–height plane.

mined but large ($0.7 \pm 0.4$ mm) relative to the refraction error and thus not negligible. When the intercept was included in their preliminary analysis, *Craymer and Vaníček* [1986] reported an insignificant turning point coefficient (35% confidence of being nonzero [*Craymer and Vaníček*, 1986 Table 1, p. 9050]. Craymer and Vaníček's claim in their comment that the matrices for their coefficients are properly conditioned hinges on their omission of the intercept; when the intercept is restored, the matrices become ill-conditioned. Their claim that the intercept should not be included because they have no explanation for it is also unsound: In fact the intercept does not differ from zero at the 95% level of confidence. What Craymer and Vaníček have done is to assume that the intercept is determined to be identically zero. This is why, as they show in Table 1 of their comment, either the intercept or the turning point argument can be used (but not both). What is not shown in their Table 1 is that the correlation coefficient for the turning point argument is insignificant under all circumstances ($r < 0.08$; Table 1). Thus the turning point argument explains less than 0.7% of the variance in the data.

### *Height Difference Argument*

The height difference coefficient and its significance depend on two out of the 60 observations. Contrary to *Craymer and Vaníček*'s [this issue] assertions, removal of influential points is standard practice to test whether a statistic is robust. Beneath the heading, "Detection of influential observations," *Draper and Smith* [1981, p. 170] write

> In any data set where the estimation of one or more parameters depends heavily on a very small number of observations, problems can arise. One way to tackle this problem is to check whether the deletion of one or two critical observations greatly affects the fit of the model and the subsequent observations. If it does, the conclusions are shaky and more data are needed.

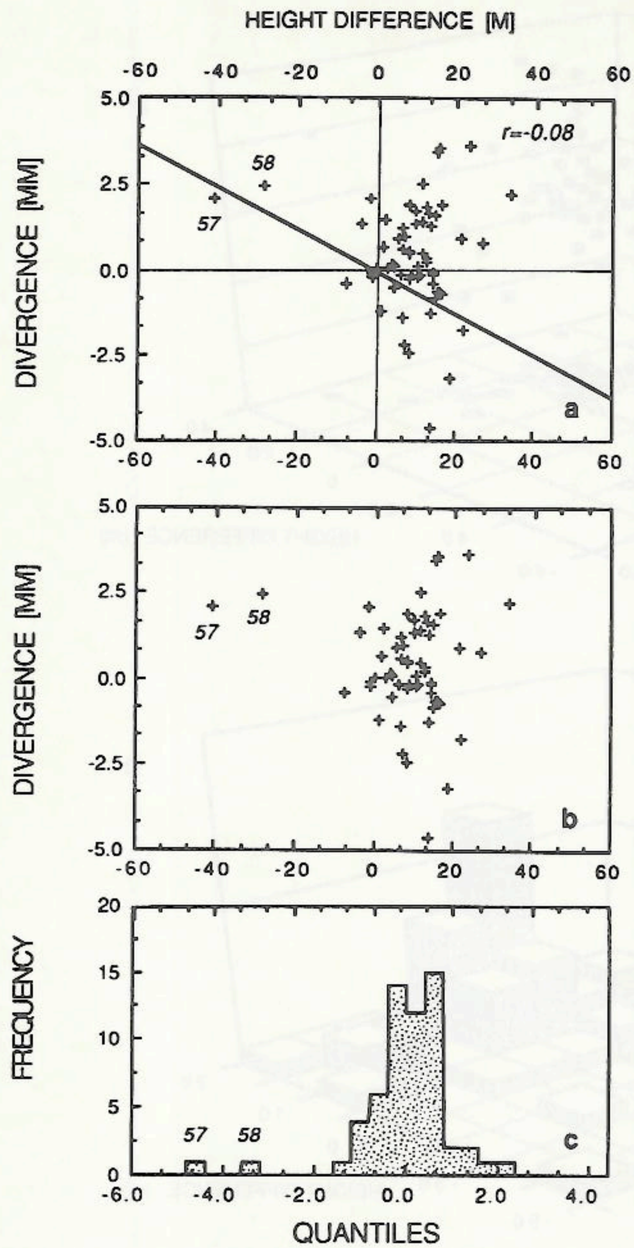


Fig. 3. Significance of the two outliers in the regression of divergence on height. (*a*) Observed divergence per section against height between bench marks. *Craymer and Vaníček*'s [1986] regression line passes through the origin and the outliers (57 and 58). (*b*) Same data without a regression line: (*c*) Frequency distribution of the height difference in meters, upper scale, and in quantiles (standard deviations from the mean), lower scale.

Divergence is shown as a function of height difference in Figure 3*a*. That sections 57 and 58 are outliers is perhaps more easily seen in Figures 3*b* and 3*c*: the points locate 3.4 and 4.5 standard deviations from the mean. When these two points are left out of the sample, the coefficient is no longer significant. That these points are outliers in the turning point–height plane can be seen in Figure 2*b*. *Craymer and Vaníček* [this issue] argue that since each observation contains many individual setups, no point may be removed. This argument, too, is flawed because the setups in a section are summed, not averaged. Thus a blunder in any individual instrument setup, or a disturbance of an individual turning point or bench mark, will cause the observation to be in error. The height difference coefficient is thus unstable and without additional data, Craymer and Vaníček's conclusions are shaky.

What if, however, all observations are inviolable and none may be removed to test the stability of the coefficient, as *Craymer and Vaníček* [this issue] maintain? *Craymer and Vaníček* [1986] pointed out that such a height dependence could not be ascribed to a linear scale error of the rods, because the same rods were used for both runnings of each section. Instead they suggested that the central meter of the 3-m-long rods was miscalibrated with respect to the upper and lower meters of the rods. This requires a 0.03-mm calibration error in the center part of the rods (30 ppm error in the 1-m rod segment), or a 0.09-mm error at the ends (30 ppm in 3 m). The two rods were calibrated before the experiment at each of the 1200 graduations (reach rod has two scales, graduated every 5 mm) by the National Bureau of Standards using a laser interferometer. The standard error of the precision in the graduation measurements is 0.007 mm (unpublished report of special measurement, rods 270718 and 277920, National Bureau of Standards Level Rod Measurement Facility, 1981). The rod is measured to an accuracy of $[0.015 + 0.005d]$ mm, where $d$ is the height of the graduation above the footplate in meters. The relative precision of graduations within the central meter of the rods is thus 0.014 mm at the 95% confidence interval, whereas for the entire rod, it is 0.03 mm. For a miscalibration required by Craymer and Vaníček, the same error would have to be repeated for each of the two scales on the two rods. The probability that a mean calibration error of 0.03 mm was made on the central meter of four rod scales measured to a precision of 0.007 mm, or an error of 0.09 mm over the four rod scales calibrated to a precision of 0.03 mm, is small.

What if such an error did occur? *Craymer and Vaníček* [1986, p. 9048] explained that a differential calibration error could affect the divergence:

> That is, the long sight runnings would typically observe the ends of the rods, while the short sight runnings would observe closer to the middle of the rods. Thus the differential scale error represents the difference in scale errors at the ends and the middle of the rods.

What they had overlooked and now recognize, however, is that the long and short sights are equally mixed in the forward-backward divergence, and therefore the differences of rod-center versus rod-end exactly cancel in their analysis. Such a scale error would appear only by regression of the long-sight divergence minus short-sight divergence after correction for refraction, a regression performed by *Stein et al.* [1986]. Our results showed that neither the regression nor the correlation coefficients differ from zero at the >75% confidence interval for this dependence ($b = -15 \pm 13$ ppm, $r = 0.16$, for all data; $b = -12 \pm 19$ ppm, $r = 0.09$, for 58 observations). We thus regard a rod scale error as implausible. *Craymer and Vaníček* [this issue] argue that though they have no explanation for the source of the height dependence, it must be maintained. This statement oddly contrasts with their refusal to include an intercept term because they have no explanation for its source. In fact, Craymer and Vaníček's height dependence owes its existence exclusively to the two outliers.

## CONCLUSIONS

Neither the number of turning points nor the height is correlated with the divergence in the 1981 Saugus-Palmdale leveling experiment ($r < 0.1$). Regression on variables that have no linear association is meaningless, explaining little, if any, variance in the data. When the intercept is unconstrained, the turning point coefficient is not significantly different form zero; when two outliers are also removed to test the stability of the height difference coefficient, the height dependence disappears. Furthermore, no height dependence can be attributed to the leveling rods when it is properly tested using all data, as Craymer and Vaníček now concede. In contrast, the refraction coefficient is statistically robust, 8 times larger than its standard error, and contributes nearly all of the variance reduction in the data.

The 1981 test showed refraction to be a large leveling error. The cumulative short-sight minus long-sight divergence, which amplifies rather than dampens refraction, is 51 mm. After correction for refraction error, the summed divergence is $8 \pm 7$ mm [*Stein et al.*, 1986]. We agree with *Craymer and Vaníček* [this issue] that systematic errors due to turning point settlement, height dependence, and other sources inevitably are present, but they are smaller than refraction and have not been reliably detected above random errors.

*Acknowledgments.* We would like to thank Michael Craymer for generously sharing all data, programs, and results with us and for his correspondence on our points of departure. Sergio Barrientos ably assisted us in programming the regressions. Malcolm Johnston, Mark Mathews, Art McGarr, and Tom Holzer furnished thoughtful reviews. We also thank anyone who has the patience to read this exchange.

## REFERENCES

Craymer, M. R., and P. Vaníček, Further analysis of the 1981 southern California field test for leveling refraction, *J. Geophys. Res.*, *91*, 9045–9055, 1986.

Craymer, M. R., and P. Vaníček, Comment on "Saugus-Palmdale, California, field test for refraction error in historical leveling surveys" by R. S. Stein, C. T. Whalen, S. R. Holdahl, W. E. Strange, and W. Thatcher, and Reply to "Comment on 'Further analysis of the 1981 southern California field test for leveling refraction' by M. R. Craymer and P. Vaníček" by R. S. Stein, C. T. Whalen, S. R. Holdahl, W. E. Strange, and W. Thatcher, *J. Geophys. Res.*, this issue.

Draper, N. R., and H. Smith, *Applied Regression Analysis*, 709 pp., John Wiley, New York, 1981.

Stein, R. S., C. T. Whalen, S. R. Holdahl, W. E. Strange, and W. Thatcher, Saugus-Palmdale, California, field test for refraction error in historical leveling surveys, *J. Geophys. Res.*, *91*, 9031–9044, 1986.

S. R. Holdahl and W. E. Strange, National Geodetic Survey, Charting and Geodetic Services, NOS/NOAA, Rockville, MD 20852.

R. S. Stein and W. Thatcher, U.S. Geological Survey, 345 Middlefield Road, MS 977, Menlo Park, CA 94025.

C. T. Whalen, 18211 Almond Terrace, Olney, MD 10832.

(Received April 16, 1987;
revised December 22, 1988;
accepted February 15, 1989.)