An Empirical Assessment of Refraction Error in Leveling as a Function of Survey Order and Environment

U.S. GEOLOGICAL SURVEY BULLETIN 2114

AVAILABILITY OF BOOKS AND MAPS OF THE U.S. GEOLOGICAL SURVEY

Instructions on ordering publications of the U.S. Geological Survey, along with prices of the last offerings, are given in the current-year issues of the monthly catalog "New Publications of the U.S. Geological Survey." Prices of available U.S. Geological Survey publications released prior to the current year are listed in the most recent annual "Price and Availability List." Publications that are listed in various U.S. Geological Survey catalogs (**see back inside cover**) but not listed in the most recent annual "Price and Availability List" are no longer available.

Reports released through the NTIS may be obtained by writing to the National Technical Information Service, U.S. Department of Commerce, Springfield, VA 22161; please include NTIS report number with inquiry.

Order U.S. Geological Survey publications **by mail** or **over the counter** from the offices given below.

BY MAIL

Books

Professional Papers, Bulletins, Water-Supply Papers, Techniques of Water-Resources Investigations, Circulars, publications of general interest (such as leaflets, pamphlets, booklets), single copies of Earthquakes & Volcanoes, Preliminary Determination of Epicenters, and some miscellaneous reports, including some of the foregoing series that have gone out of print at the Superintendent of Documents, are obtainable by mail from

U.S. Geological Survey, Map Distribution
Box 25286, MS 306, Federal Center
Denver, CO 80225

Subscriptions to periodicals (Earthquakes & Volcanoes and Preliminary Determination of Epicenters) can be obtained ONLY from the

Superintendent of Documents
Government Printing Office
Washington, DC 20402

(Check or money order must be payable to Superintendent of Documents.)

Maps

For maps, address mail orders to

U.S. Geological Survey, Map Distribution
Box 25286, Bldg. 810, Federal Center
Denver, CO 80225

Residents of Alaska may order maps from

U.S. Geological Survey, Earth Science Information Center
101 Twelfth Ave., Box 12
Fairbanks, AK 99701

OVER THE COUNTER

Books and Maps

Books and maps of the U.S. Geological Survey are available over the counter at the following U.S. Geological Survey offices, all of which are authorized agents of the Superintendent of Documents.

- **ANCHORAGE, Alaska**—4230 University Dr., Rm. 101
- **LAKEWOOD, Colorado**—Federal Center, Bldg. 810
- **MENLO PARK, California**—Bldg. 3, Rm. 3128, 345 Middlefield Rd.
- **RESTON, Virginia**—National Center, Rm. 1C402, 12201 Sunrise Valley Dr.
- **SALT LAKE CITY, Utah**—Federal Bldg., Rm. 8105, 125 South State St.
- **SPOKANE, Washington**—U.S. Post Office Bldg., Rm. 135, W. 904 Riverside Ave.
- **WASHINGTON, D.C.**—Main Interior Bldg., Rm. 2650, 18th and C Sts., NW.

Maps Only

Maps may be purchased over the counter at the U.S. Geological Survey offices:

- **FAIRBANKS, Alaska**—New Federal Building, 101 Twelfth Ave.
- **ROLLA, Missouri**—1400 Independence Rd.
- **STENNIS SPACE CENTER, Mississippi**—Bldg. 3101

An Empirical Assessment of Refraction Error in Leveling as a Function of Survey Order and Environment

By Robert O. Castle, Robert K. Mark, *and* Roger H. Shaw

U.S. GEOLOGICAL SURVEY BULLETIN 2114

An analysis of the results of three North American field experiments designed to measure the systematic component of the atmospheric refraction error in leveling

UNITED STATES GOVERNMENT PRINTING OFFICE, WASHINGTON : 1994

U.S. DEPARTMENT OF THE INTERIOR

BRUCE BABBITT, *Secretary*

U.S. GEOLOGICAL SURVEY

Gordon P. Eaton, *Director*

For sale by
U.S. Geological Survey, Map Distribution
Box 25286, MS 306, Federal Center
Denver, CO 80225

Text and illustrations edited by and book designed by Derrick D. Hirsch
Graphics prepared by Kathryn P. Nimz

Library of Congress Cataloging-in-Publication Data

Castle, Robert O.
An empirical assessment of refraction error in leveling as a function of survey order and environment : an analysis of the results of three North American field experiments designed to measure the systematic component of the atmospheric refraction error in leveling / by Robert O. Castle, Robert K. Mark, and Roger H. Shaw.
p. cm. — (U.S. Geological Survey bulletin ; 2114)
Includes bibliographical references.
1. Leveling. 2. Refraction. 3. Surveying. I. Mark, Robert K. II. Shaw, Roger H. III. Title. IV. Series.
TA606.C38 1994
526.3'6—dc20 94–36782
CIP

CONTENTS

FIGURES

TABLES

An Empirical Assessment of Refraction Error in Leveling as a Function of Survey Order and Environment

By Robert O. Castle, Robert K. Mark, *and* Roger H. Shaw[1]

ABSTRACT

Analyses of the results obtained from three field tests intended to measure the accumulation of the unequal-refraction error (URE) in leveling demonstrate that this error can be quantified as a function of survey order and, less certainly, factors controlled by the local environment. The three experiments are (1) the Gaithersburg-Tucson refraction tests of the National Geodetic Survey, (2) the Saugus-Palmdale field test conducted jointly by the U. S. Geological Survey and the National Geodetic Survey, and (3) the Canadian or Eastern Townships experiment of the Geodetic Survey of Canada. The critical design feature of the Gaithersburg-Tucson refraction tests was the attempted creation of error-free datums against which refraction-contaminated measurements could be compared. The other two experiments were designed to measure the divergence between two essentially contemporaneous surveys over the same route, where both were assumed to be contaminated in significant degree by URE. Owing to the theoretical dependency of the URE on the square of the sight length, the latter two experiments compared single-run, long-sight-length surveys against surveys in which the sight lengths were either nearly exactly or approximately half those of the long-sight-length surveys. The Saugus-Palmdale and Canadian experiments differed chiefly in the imposition of contrasting setup-rejection criteria (0.30 mm for the short-sight-length survey and 0.75 mm for the long-sight-length survey) on the Saugus-Palmdale levelings, as compared with the absence of any setup-rejection criterion imposed on either of the Canadian test levelings.

Temperature measurements made in conjunction with each of the experiments, supplemented by an additional set of temperature measurements made at Corbin, Virginia, demonstrate that a variety of environmental factors may either enhance or suppress the URE. Because any tendency toward the accumulation of the URE ultimately depends on the curvature in the thermal gradient in the atmospheric boundary layer, these temperature measurements provide an independent assessment of the likely occurrence of this error. Our analyses of the results of the temperature measurements indicate that until $\Delta t'$ (the temperature at 0.5 m minus that at 2.5 m) exceeds 0.4°C to 0.5°C, the curvature in the thermal gradient approaches insignificance over the full 0.5- to 2.5-m range. Because $\Delta t'$ depends in turn on several poorly quantified but important environmental factors, these factors clearly affect the magnitude of any URE.

Recognized threshold values in the ambient air temperature, $\Delta t'$, and the curvature in the thermal gradient, below which the URE does not accumulate, are ultimately attributable to the fact that temperature proxies for the sensible heat flux. These threshold values are thought to be functions of both the setup-elevation difference (Δh) and the total environment. As Δh decreases, the inequality of the refraction error in foresight and backsight observations decreases as well, and the threshold temperatures tend to increase. Because it is very unlikely that the URE is a simple multiple of Δh, we can presently identify no more than representative and limiting threshold values as functions of the total environment. Specifically, in a classically arid regime (Tucson) devoid of soil moisture, significant vegetation, or cloud cover, and where Δh matches the maximum procedurally permitted value of 2 m, the threshold temperature may be as low as 6°C to 8°C. As moisture, vegetation, and cloud cover increase to levels appropriate to a semi-arid regime (Saugus-Palmdale) and Δh is reduced to ~1.2 m, the threshold temperature increases to 22°C to 24°C. With increasing cloud cover and soil moisture and surfaces characterized by a thick, actively transpiring turf representative of temperate to humid regimes (Gaithersburg-Corbin), and where Δh is again increased to 2 m, the threshold temperature probably exceeds 26°C to 28°C and the URE drops to near zero. These increasing threshold temperatures are due to increasing fractions of the solar radiation blocked by

[1]Department of Land, Air and Water Resources, University of California, Davis.

Manuscript approved for publication August 24, 1994.

cloud cover and the extraction of much of the residual thermal input by evaporation phenomena, whereby the system is robbed of heat that could otherwise contribute to the sensible heat and, hence, increased curvature in the thermal gradient.

The use of double-scale rods in each of the three experiments provides a means by which the results of all three can be integrated through analysis of the high-scale low-scale differences. Considered together, these experiments show that throughout the 20th century levelings conducted to first-order standards rarely, if ever, accumulated URE in excess of the predicted random error. Similarly, where the sectional temperatures held at or below some locale- and Δh-dependent threshold temperature or where the observer-ensured threshold scintillation (equivalent to a standard deviation about the mean of the high-scale low-scale difference of ~0.18 mm) was not exceeded, the URE associated with geodetic leveling of any order probably was about the same as that which could be expected to have contaminated first-order work. Extremely unlikely worst-case estimates of maximum URE contamination of any 20th-century second-order leveling range up to 50 mm over a 500-m height difference and perhaps as much as 100 mm for third-order leveling over the same height difference.

Modeled values of the URE consider only a few of the variables that control the occurrence of this error; they explicitly exclude, for example, the impact of ambient air temperature. As a result, the modeled corrections become progressively less useful with increasingly constrained leveling procedures. Accordingly, empirically based estimates of the URE probably are far more reliable. Thus, additional experimental observations based on a modified Gaithersburg-Tucson design should provide a sound basis for estimating the URE in those situations intermediate between worst-case accumulations and those in which $\Delta h \leq 1.2$ m and the temperature rarely rose above 22°C to 24°C.

INTRODUCTION

One of the most enigmatic sources of systematic error in leveling is that associated with atmospheric refraction. Owing to a theoretical proclivity toward greater refraction errors in the upslope direction, this error should be asymmetrical with respect to a balanced-sight setup and is thought to accumulate in an at least semi-systematic fashion as an unequal-refraction error (URE). The potential for the accumulation of an atmospheric refraction error has been recognized by geodesists for the better part of a century (Hayford, 1904, p. 213; Bowie and Avers, 1914, p. 18–19, 21–22). The cumulative impact of atmospheric refraction on leveling observations, whether systematic or not, traditionally has been dealt with through the imposition of procedural constraints intended to minimize any resultant error to no more than trivial values. Much more recently, attempts have been made to model the URE and thereby provide a means for correcting the field observations (see, for example, Whalen, 1981; Strange, 1981; Holdahl, 1982).

The combined results of three modern North American field experiments, supplemented by data obtained from a series of routinely produced double-scale rod levelings, suggest an essentially empirical approach to determining the refraction error content in any geodetic leveling. Analysis of these three experiments indicate that we can now place fairly tight, worst-case limits on the URE as a function of survey order. Moreover, even a rudimentary knowledge of the survey characteristics, such as average slope or temperature, may provide additional bases for estimating the actual rather than the worst-case error identified with a particular leveling.

An empirical determination of the magnitude of the URE in any procedurally constrained leveling can be expected to serve the needs of several groups within the geophysical community. For example, the inclusive cost of the new North American Vertical Datum is estimated to be about $50 million (M.G. Grunthal, 1992, written commun.). Our empirical estimates of the URE may provide a useful, independent means of evaluating the extent to which the results of this enormous effort are either enhanced or degraded with the addition of modeled refraction corrections. Similarly, the historical vertical-control record currently constitutes a grossly underutilized data set containing almost a century of observations that should lend themselves to a variety of crustal-deformation investigations. The failure to exploit more fully this historical record is attributable in large measure to the widely held belief that the results of older geodetic levelings in particular are highly contaminated by refraction error. An explicit, empirically derived determination of the URE for any given leveling, worst case or otherwise, can only augment the applicability of the historical vertical-control record to studies of continuing crustal deformation.

THE NATURE OF THE URE

The theoretical basis for an expected accumulation of the URE, and thus the underpinnings for the modeled corrections, has been described in detail by a number of investigators (see, for example, Whalen, 1981; Strange, 1981; Shaw and Smietana, 1983). Accordingly, we consider here only the essentials of both the problem and those procedures that have been used in the past in attempting to minimize its occurrence and estimate its magnitude.

Because the refractive index of air in any given locale is almost exclusively a function of temperature (Shaw and Smietana, 1983, p. 10,485), any expectation for the occurrence of an unequal atmospheric-refraction error over sloping terrain is rooted in the near-surface thermal gradient, where $d^2t/dz^2 \neq 0$ (fig. 1). Accepting this premise, the upslope refraction error (Δu) should be greater than the downslope error (Δl) during daylight hours, thereby creating a likelihood for the systematic accumulation of refraction error. The magnitude of the URE (the negative or modeled refraction correction) commonly is approximated as:

$$\text{URE} = -4 \times 10^{-9} \gamma \, \Delta h \Delta t' L^2 \qquad (1)$$

where

γ = a parameter that depends on latitude, month of the year, and the time of day and is believed to vary within a range of 50 to 80, but generally is close to 80;

$\Delta h = z_l - z_u$ (fig. 1);

$\Delta t' = t_{0.5} - t_{2.5}$ (the difference in temperature in degrees Celsius, at 0.5 m and 2.5 m above the ground); and

L = sight length (modified from Strange, 1981, p. 2810, 2812).

Accepting the relative invariance of γ, equation (1) is structured such that the URE depends in theory almost entirely on three variables: Δh, $\Delta t'$, and L. Moreover, because the URE is proportional to the square of the sight length, it is this variable that emerges as what is thought to be the chief control on the magnitude of the error (Strange, 1981, p. 2814) and has, in fact, governed the design of all of the field experiments on which we base this study.

Although the three variables given in equation (1) are independent quantities, they collectively control the magnitude of the URE. Because Δh is limited by both the length of the rod and the height of the level (fig. 1), increasing slopes tend to limit sight lengths; where the slope rises above some instrumentally controlled value, sight lengths are automatically shortened. Similarly, because scintillation tends to increase with increasing temperature (or $\Delta t'$), the maximum allowable sight length in any procedurally constrained leveling may be further shortened in order to satisfy the setup-, section- or circuit-rejection criteria (Rappleye, 1948, p. 2–3, 7; Federal Geodetic Control Committee, 1980, p. 28), and Δh will be diminished as well. Thus, owing to the interdependence among these three variables, the URE is path dependent in both space and time and not simply a function of the height difference between two points.

ACKNOWLEDGMENTS

We thank Rodney Jack Lee and Robert C. Wilson for advice and background on a variety of survey practices. Roy M. Sachs offered thoughtful counsel on evapotranspiration effects that impact on the sensible heat content of the boundary layer. We are especially indebted to Emery I. Balazs for computation of the thread interval for the three-wire Fischer level reticule, for acquisition of the Corbin temperature measurements, and for the recovery of important archival data. Finally, we thank Kenneth R. Lajoie, David S. McCulloch, Robert E. Powell, and Arthur G. Sylvester for their constructive and insightful reviews of earlier versions of this report.

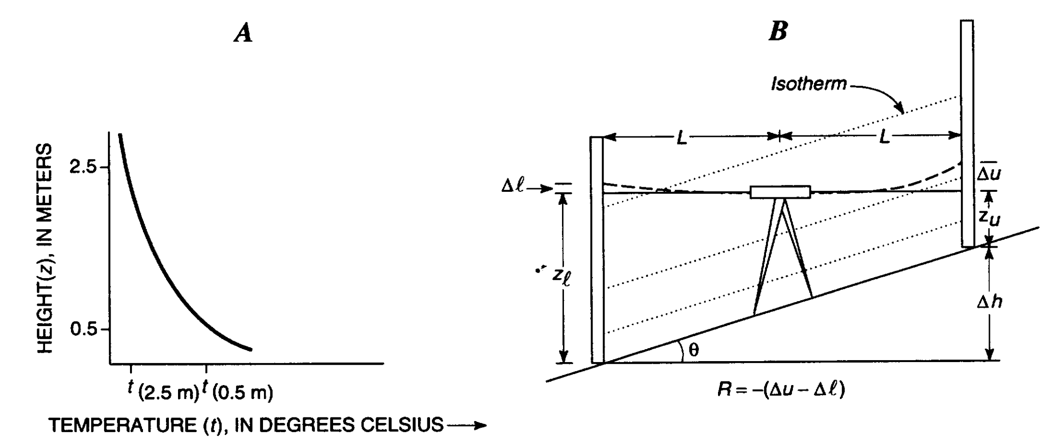


Figure 1. Schematic diagrams illustrating conceptual basis for the occurrence of unequal-refraction error (URE). *A*, Idealized, averaged thermal gradient during normal leveling day. *B*, Balanced sights leveling setup, where separations between isotherms are of equal temperature difference and R (URE) is $-(\Delta u - \Delta l)$; curvature of ray path greatly exaggerated. Δu, upslope refraction error; Δl, downslope refraction error; L, sight length; $\Delta h = (z_l - z_u)$; R, refraction error; θ, slope angle from horizontal.

THE EXPERIMENTS

The three experiments on which we base this investigation are: (1) The Gaithersburg-Tucson refraction tests of the National Geodetic Survey (Whalen, 1980; 1981); (2) the Saugus-Palmdale field test for refraction error, jointly conducted by the U.S. Geological Survey and the National Geodetic Survey (Whalen and Strange, 1983; Stein and others, 1986); and (3) the Canadian or Eastern Townships field test for the systematic effects of refraction conducted by the Geodetic Survey of Canada (Heroux and others, 1985). Insofar as they impact the objectives of each of the several investigations, all of these experiments incorporated both innovative and constructive design features. However, these same experiments were also characterized by modest deficiencies in design and (or) implementation. Nevertheless, if the experimental results are examined in concert, these deficiencies tend to be nullified. This is not to say that further and much more comprehensive experimental investigation could not define more firmly the conclusions developed in this report.

THE GAITHERSBURG-TUCSON TESTS

Perhaps the most critical of the experimental results on which we base this investigation are drawn from the Gaithersburg-Tucson refraction tests. While these geographically separate parts of this experiment were conceptually identical, the Gaithersburg measurements were carried out within the temperate, generally moist, and heavily canopied environment of central Maryland. The Tucson measurements, on the other hand, were conducted over a classically arid desert surface devoid of significant moisture or plant cover during a fairly warm part of the year (April—see below). The defining feature of this set of experiments was the attempted creation of near error-free datums, against which refraction-contaminated observations might be compared. The experimental layout permitted readings at nominal heights on the rods of 0.5 m, 1.5 m, and 2.5 m above the ground surface at progressively greater sight lengths (fig. 2). Thus, repeated readings on the rods at 0.5 m (upslope) and at 2.5 m (downslope) simulated levelings over sloping terrain, whereas those at 1.5 m simulated levelings over a line identified with zero slope. The simulated cumulative line lengths ranged between 14.6 km and 28.6 km; the simulated height differences ranged upward from 436.5 m to 545.8 m. The measured cumulative errors over the simulated sloping lines, errors which ideally equate with the cumulative URE, fell between –9.1 mm and –149.7 mm (table 1). Therefore, the results of the Gaithersburg-Tucson tests are thought to provide relatively unequivocal demonstrations of the potential for the accumulation of a large URE—where the observations are constrained by no other criteria than the levelman's ability to read the graduations on the rod. These same tests were thought to be equally demonstrative of the efficacy of the Kukkamäki (1938) modeled refraction corrections where the errors were unusually large; the modeled corrections proved much less effective where the errors were relatively small (Whalen, 1981).

Our chief concern regarding the Gaithersburg-Tucson test results stems from the less meticulously determined datums (or standards) developed for the Gaithersburg observations than those produced in connection with the Tucson experiment (Whalen, 1981, p. 5). A recent and much more comprehensive treatment of the Gaithersburg data than that attempted by Castle and others (1985) suggests the presence of a large systematic error (or errors) unrelated to refraction and probably related to either the elevation standards or, less likely, the subsequent and presumably refraction-contaminated observations (see section on analysis of Gaithersburg-Tucson test results). Indeed, that the Gaithersburg standards may have been contaminated by relatively large, unassessed errors is independently implied by the relative instability of these standards, as contrasted with the apparent stability of those at Tucson (table 2). Moreover, there is no reason why the elevation standards at the 30-, 50- (or 45), and 60-m

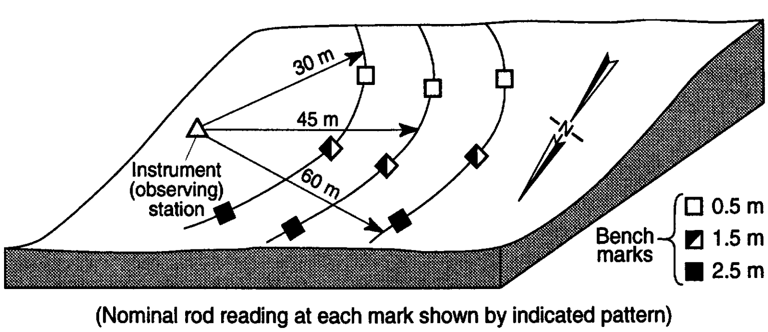


Figure 2. Layout of Tucson test range. Gaithersburg layout is generally similar except for substitution of marks at 50 rather than 45 m distant from instrument station. Modified from Whalen (1981, p. 2).

Table 1. Summary of daytime balanced-sight results from Gaithersburg-Tucson refraction tests.[a]

(1) Sight length (m)	(2) Number of setups	(3) Simulated line length (km)	(4) Standard elevation difference per setup on nominal 0.5- and 2.5-m marks (m)	(5) Simulated height difference based on 0.5- to 2.5-m standard elevation difference (m)	(6) O–S[b] (mm)	(7) O–S[c] (mm)	(8) σ/(O–S)[d]	(9) O–S[c], normalized to Gaithersburg 30-m line height difference of 491.4 m (mm)	(10) σ_{DDH}, based on readings at nominal 1.5 m (zero-slope) height (mm)	(11) σ_{DDH}, based on readings at nominal 0.5 m and 2.5 m (sloping-line) heights (mm)
					GAITHERSBURG					
30	243	14.6	2.02212	491.4	0.3	−9.1	1.54	−9.1	0.163	0.133
50	221	22.1	1.97516	436.5	4.4	−38.4	.33	−43.2	.247	.337
60	225	27.0	2.02672	456.0	9.2	−77.5	.17	−83.5	.321	.381
					TUCSON					
30	287	17.3	1.79558	515.3	−10.8	−29.4	0.20	−28.0	0.180	0.190
45	284	25.6	1.92180	545.8	−8.7	−52.1	.17	−46.9	.283	.256
60	239	28.6	1.95401	467.0	−7.4	−149.7	.38	−157.5	.355	.478

[a]Chiefly from tables 3, 4, A-1, A-3, B-1, and B-3 of Whalen (1981). Standard deviation about the mean high-scale minus low-scale discrepancy (DDH), σ_{DDH}, computed from field observations (Castle and others, 1985).
[b]Observed elevation difference minus standard elevation difference over simulated zero-slope line.
[c]Observed elevation difference minus standard elevation difference over simulated sloping line.
[d]σ equals standard error of O–S sums over simulated sloping line.

sight-length distances should have been progressively less accurately determined, as seems to have been the case at Gaithersburg but obviously was not at Tucson (table 2). Accordingly, because of the more rigorously defined and more frequently determined standards (Whalen, 1981, p. 5) (table 2), systematic contamination of the Tucson data seems much less likely, although this likelihood cannot be confirmed owing to the disappearance of the Tucson records from (or within) the National Geodetic Survey archives since our earlier examination of these data in 1985. Regardless, independent support of the accuracy of the Tucson standards is shown by the sharp reversal in both the size and magnitude in the observed minus standard nighttime observations, a reversal consistent with the normal nighttime reversal in the sign of $\Delta t'$ (equation 1) (Whalen, 1981, p. 12–13).

THE SAUGUS-PALMDALE FIELD TEST

The second set of experimental results that we have drawn on here are those of the Saugus-Palmdale field test for refraction error. This experiment, which was carried out over a gently sloping 50-km line (fig. 3), is unique among the three that we have considered here in that it is the only one in which procedural (rejection) constraints were placed on the observations. The Saugus-Palmdale regime much more closely matched that of the Tucson test site than that at Gaithersburg; nevertheless, the Saugus-Palmdale environment probably is better described as semi-arid rather than classically arid.

The Saugus-Palmdale field test was designed to compare two temporally and spatially coincident surveys that differed chiefly in sight length, setup-elevation difference, and the procedural constraints incorporated in the separate levelings. The experimental procedure was governed by the generally accepted premise that refraction error varies as a function of the square of the sight length. Specifically, "the experimental leveling was double [*sic*] run, in which one leg of each section was composed of long-sight-length [LSL] observations and the other was made with sights of approximately half the length [SSL]" (Stein and others, 1986, p. 9032). In order to dispel whatever ambiguity might exist in the preceding statement, it should be noted that neither the LSL nor the SSL leveling was itself double run, but rather that the single-run SSL leveling closed on single-run LSL leveling over each section; it is only in this sense that these surveys were double run. An especially prudent design feature included in this experiment required that the same equipment be used in both the LSL and SSL surveys, thereby minimizing the likelihood that any instrumentally related sources of systematic error might be introduced into the experiment (Stein and others, 1986, p. 9032, 9041).

Two procedural constraints were imposed on the Saugus-Palmdale test leveling: (1) Closures of the SSL on the LSL survey were required to meet a section-rejection limit of 4.00 mm/km$^{1/2}$ (Stein and others, 1986, p. 9032, 9035). (2) Because double-scale rods were used in the experiment (see fig. 4), each of the comparative surveys could be assigned its own setup-rejection limit. The maximum allowable high-scale

Table 2. Standard elevation differences for sloping-line rod stations at 30-, 50- (or 45) and 60-m sight-length distances at Gaithersburg and Tucson test ranges.[a]

Period or day	Sight length (m)	Adjusted elevation difference (m)	Mean (m)	Standard deviation (mm)
		GAITHERSBURG		
8/9–10	30	2.02216		
9/7–14	30	2.02211		
9/24–26	30	2.02210		
	30		2.02212	0.032
9/7–14	50	1.97521		
9/24–26	50	1.97510		
	50		1.97516	0.078
8/9–10	60	2.02627		
9/7–14	60	2.02680		
9/24–26	60	2.02663		
	60		2.02672	0.120
		TUCSON		
4/8	30	1.79556		
4/9	30	1.79558		
4/10	30	1.79562		
4/11	30	1.79557		
4/14	30	1.79561		
4/15	30	1.79557		
4/17	30	1.79559		
4/18	30	1.79562		
4/21	30	1.79558		
4/23	30	1.79548		
	30		1.79558	0.040
4/8	45	1.92171		
4/9	45	1.92176		
4/10	45	1.92180		
4/11	45	1.92181		
4/14	45	1.92178		
4/15	45	1.92179		
4/17	45	1.92178		
4/18	45	1.92176		
4/21	45	1.92189		
4/23	45	1.92188		
	45		1.92180	0.054
4/8	60	1.95397		
4/9	60	1.95399		
4/10	60	1.95409		
4/11	60	1.95405		
4/14	60	1.95401		
4/15	60	1.95398		
4/17	60	1.95398		
4/18	60	1.95395		
4/21	60	1.95406		
4/23	60	1.95424		
	60		1.95401	0.047

[a]Modified from tables A-3 and B-3 of Whalen (1981).

minus low-scale difference (the DDH or, in the terminology of Stein and others, δ) per setup for the SSL survey was set at ±0.30 mm, whereas that for the LSL survey was set at ±0.75 mm. The setup-rejection limit for the SSL survey (±0.30 mm) was deliberately chosen to match that currently in force for class II first-order leveling (Stein and others, 1986, p. 9038–9039)[2]. Although Stein and others (1986, p. 9039) indicate that the limiting DDH was purposely relaxed for the LSL survey, they offer no explicit a priori basis for the choice of the ±0.75-mm value. A rationale for this particular choice has been provided, however, by the former Director of the National Geodetic Survey. According to J.D. Bossler (1982, written commun.):

> Based on our refraction test observations at Gaithersburg, Maryland, and Tucson, Arizona, the tolerance on high- and low-scale elevation differences (DDH) at each setup should be, based on a 5-percent test level:
>
> $$\text{Tol. on } |\text{DDH}| = 0.00392\sqrt{2800+0.683s^2+0.197s^3}, \text{ mm} \qquad (1)$$
>
> where s is the sight distance in meters. * * * The average sight distances for short and long sights [actually obtained] for the [Saugus-Palmdale] test were 25 and 42 m respectively. Based on equation (1), the corresponding tolerance limits should have been ± 0.30 mm and ± 0.55 mm. When the test was planned, the average short- and long-sight distances were unknown. The long-sight distances were anticipated to be of the order of 55 m, which would result in a tolerance limit of 0.75 mm, based on equation (1).

This statement clearly indicates that the anticipated sight lengths governed the choice of the experimental LSL setup-rejection criterion, whereas in actual practice

[2]Although Stein and others (1986, p. 9039) cite the Federal Geodetic Control Committee (1984) as their authority for this setup-rejection limit, the 1984 setup-rejection limit is, in fact, given as ±1.00 mm for reversible-compensator instruments (Federal Geodetic Control Committee, 1984, p. 3–7)—such as the Jena NI 002 which was used in the Saugus-Palmdale experiment (Stein and others, 1986, p. 9035). Before 1984 no distinction was made between reversible-compensator levels and any other level in the setup-rejection criteria (Federal Geodetic Control Committee, 1980, p. 28). This equality in the rejection criteria was achieved through a programmed correction in the field calculator that removed any instrument contribution to the DDH that could be attributed to the use of a reversible compensator instrument. Thus, the observer read a low-scale high-scale difference in which this programmed correction effectively cancelled any instrument-dependent DDH. While it is uncertain that this programmed correction was in fact incorporated in the field calculator used in connection with the Saugus-Palmdale field test, the Geological Survey's field consultant's best recollection is that the actually utilized field calculator did provide for this correction (R.C. Wilson, 1994, oral commun.). Regardless, because any reversible-compensator instrumental contribution vanishes if both foresight and backsight observations are precisely balanced, and because the maximum permissible sight-length imbalance for this experiment is given as 2 m (Stein and others, 1986, p. 9035), we have treated the Saugus-Palmdale low-scale high-scale differences as the equivalents of those that would have been obtained had this instrument been either a non-compensator instrument or an NI 002 that was at all times free of any instrumental contribution to the DDH owing to both the highly contrained sight-length imbalance and regular checks and adjustments for collimation error (see also table 6).

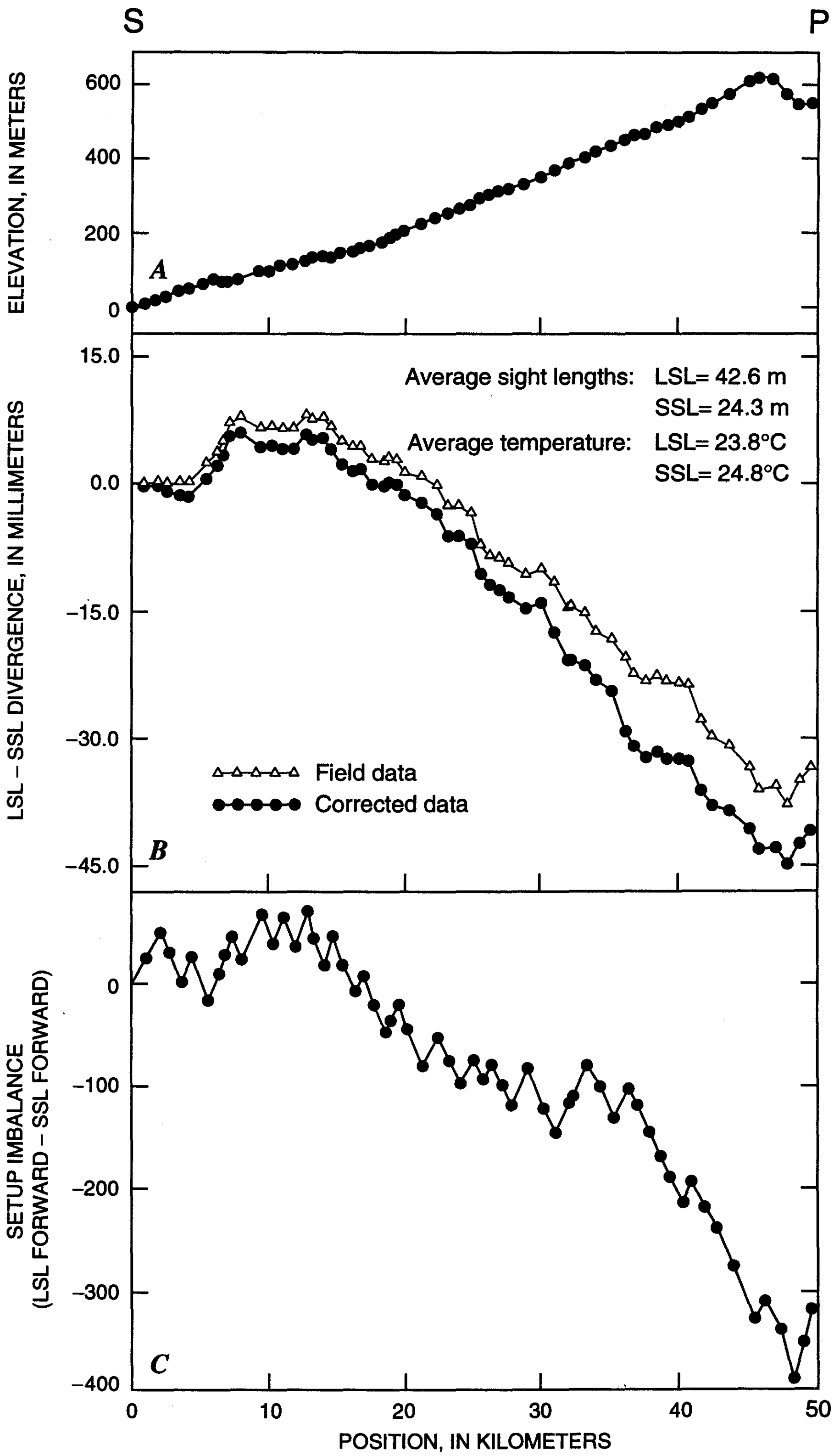


Figure 3. Significant elements of the Saugus-Palmdale field test. *A*, Terrain profile between Saugus (S) and Palmdale (P). *B*, Cumulative divergence between long-sight-length (LSL) and short-sight-length (SSL) surveys as a function of position along line. *C*, Cumulative sum of total number of setups for LSL forward minus SSL forward sections (setup imbalance) as a function of position along line. Field data from Teledyne Geotronics (1981); corrected data from R.S. Stein (1982, written commun.). Modified from Castle and others (1985).

the rejection criteria in any geodetic leveling, whether setup, section, or circuit, control the sight lengths. Thus, the selection of the limiting DDH for the LSL survey obviously departed from that which led to the selection of the limiting DDH for the SSL survey—where the latter value was tied to an established, experience-based procedural constraint that is no more than indirectly related to sight-length considerations. We focus on this issue here, at least in part, because it impacts controversial aspects of our analysis of the experimental data.

We have made use of two conceptually separable but very similar data sets in our presentation and analysis of the Saugus-Palmdale test results. The first set consists simply of the unsullied field observations (Teledyne Geotronics, 1981). The second set consists of the same measurements as the first, but corrected for (1) collimation errors, (2) invar-band- (rod-) scale errors, (3) thermal expansion of the invar bands, and (4) the effects of solid-earth tides (R.S. Stein, 1982, written commun.). Any collimation errors should have been vanishingly small (see Stein and others, 1986, p. 9035; Schomaker and Berry, 1981, p. (3-20)–(3-22)). Moreover, because the same rod pair was used in both the LSL and SSL surveys, any variation in the rod excess in the necessarily different parts of the rods actually sampled by the short-sight and long-sight observations

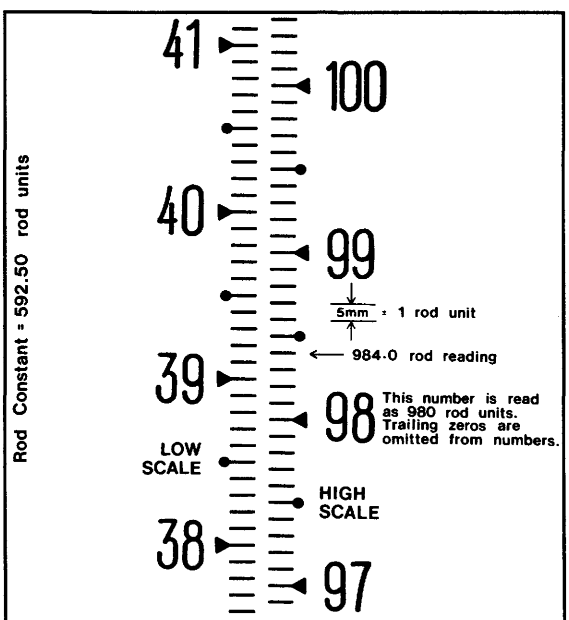


Figure 4. Significant features of a double-scale rod with half-centimeter graduations. An optical micrometer attached to level permits readings to 0.01 rod units (0.05 mm). If the left-hand (low-scale) band is read as 391.50 rod units against 984.00 rod units on the right-hand (high-scale) band, the high-scale low-scale difference, or DDH, would be 0.00 mm. Adapted from Schomaker and Berry (1981, p. (3-22), (3-23), (3-42).

could only be due to the nonlinear distribution of error along the length of the rod. However, based on a 100-rod sample drawn from the National Geodetic Survey rod and instrument file, Mark and others (1981, p 2787–2790) found the distribution of error to be strikingly linear. Accordingly, we infer that rod-scale error made no more than a trivial contribution to the LSL-SSL divergence based on the field observations. Although the ambient air temperature ranged widely along the route traversed by the Saugus-Palmdale experiment (Teledyne Geotronics, 1981), the mixing of LSL and SSL sectional runs as a function of temperature should have nearly nullified any impact on the divergence of the thermal expansion of the invar bands—especially since the coefficient of thermal expansion for the Kern rod invar is given as only about 8×10^{-7}/°C (see Stein and others, 1986, p. 9035; National Geodetic Survey rod and instrument file). Similarly, because the maximum amplitude and wavelength of any earth tide are such that the error contribution to an individual section would be at the micron (or submicron) level, a correction for its occurrence contributes virtually nothing to the goals of this experiment. Finally, corrections made to the LSL-SSL field values in only two sections account for two thirds of the discrepancy in the cumulative divergence based on the field values versus that based on the corrected data—an observation that not only challenges the validity of the corrected data but requires that we view this entire data set as suspect.

Although we recognize a clearly defined preference, we have generally presented the Saugus-Palmdale results obtained from *both* the field data and the corrected values (see, for example, fig. 3*B*). It could be argued that the corrected data minimize any contribution to the LSL-SSL divergence unrelated to refraction and should be the focus of our analysis. However, all of the corrections border—or should border—on the miniscule, and there is, as we have indicated in the preceding paragraph, some basis for questioning both their value and their accuracy. Moreover, no attempt was made to accommodate what may be the most serious of the non-refraction related systematic errors—namely, that associated with staff (rod) and (or) instrument settlement. Most importantly, perhaps, because it is the field observations that determine the acceptability of any procedurally constrained measurement or set of measurements, the field data may be the more useful in attempting to understand the implications of this experiment. Regardless, because we have generally presented the results in both formats, the reader is free to excercize his own preference.

The results of the Saugus-Palmdale field test are generally consistent with theory and again demonstrate the potential for the accumulation of a large URE. That is, as a function of position along line, the LSL-SSL divergence sums to a maximum of about −38 mm (or −45 mm based on the corrected data) over a distance of less than 50 km (fig. 3*B*), and hence accords with that which could be predicted from equation (1). There is little doubt, moreover, that the indicated divergence is almost entirely attributable to atmospheric refraction. This conclusion should be qualified in part, however, owing to statistically demonstrable contributions of staff- and (or) instrument-settlement error to one or both of the comparative surveys (Castle and others, 1983; Craymer and Vaníček, 1986). Staff settlement may lead to the generation of a directionally dependent systematic error whose impact generally is controlled through double running or, as was intended in this case, by ensuring that the direction of running of the individual surveys is regularly alternated. In fact, however, the setup imbalance between the LSL forward and the SSL forward surveys, particularly toward the eastern (Palmdale) end of the line (fig. 3*C*), is in itself prima facie evidence that settlement may have amplified the LSL-SSL divergence (Castle and others, 1983, p. 1083). Stein and others (1986, p. 9042), on the other hand, suggested that the contribution of staff settlement to the LSL-SSL divergence probably was trivial. Through coupling of the average settlement figure of 0.014 mm/setup (Craymer and Vaníček, 1986, p. 9053) with still further analysis, M.R. Craymer (1992, written commun.) concluded, more or less in agreement with Stein and others, that only about 4.7 mm of the LSL-SSL divergence is due to settlement. Thus, as it relates to this study and excepting refraction, it is likely that the contribution of systematic error to the cumulative divergence (fig. 3*B*) is so small that it can be disregarded.

THE CANADIAN FIELD TEST

The Canadian field test for the systematic effect of refraction error incorporated design considerations found in both the Gaithersburg-Tucson and Saugus-Palmdale experiments. This experiment was carried out in the Eastern Townships of Quebec during the summer of 1984 and, hence, within a climatic regime similar to that which obtained during the Gaithersburg observations. The Canadian refraction test was based on repeated levelings over a six-section, 5.7-km line summing to a cumulative, simulated line length of 34.1 km and a simulated height difference of 283.4 m (Heroux and others, 1985, as modified by W. Gale, 1992, written commun.). The same equipment was used throughout the experiment, and the three-rod observing procedure was so organized as to minimize or even eliminate any bias associated with residual rod error (Heroux and others, 1985, p. 313–316). The Canadian experiment was explicitly similar to the Saugus-Palmdale experiment in that it compared long-sight-length leveling over each section against short-sight-length leveling of almost exactly half the sight length, where the two values averaged about 45 m and 22.5 m, respectively (Heroux and others, 1985, p. 311, 314). However, unlike the Saugus-Palmdale experiment and more closely akin to the Gaithersburg-Tucson tests, any likely instability (such as staff settlement) of the turning points was eliminated through the installation of permanent turning points, thereby fixing the sight lengths in advance of the observations (Heroux and others, 1985, p. 313–314). Thus, and analogous to the Gaithersburg-Tucson tests, the only restrictions on the observations were the fixed sight lengths and the levelman's ability to read the graduations on the rods.

The results of the Canadian field test are again consistent with those predicted from theory—equation (1). The accumulated LSL-SSL divergence over the 34.1-km simulated line through a simulated height difference of 283.4 m summed to −14.94 mm (Heroux and others, 1985 p. 316, as modified by W. Gale, 1992, written commun.). This divergence, even though small, probably but questionably is best explained by the dependency of the URE on the square of the respective sight lengths.

COMPARATIVE TEMPERATURE MEASUREMENTS

The thermal structure within the first three meters of the atmospheric boundary layer ultimately determines any proclivity toward the generation and accumulation of URE. This structure is controlled in turn by the sensible heat flux, which itself controls both $\Delta t'$ (equation 1) and the curvature in the thermal gradient (fig. 1). It is these considerations that prompted a detailed examination of five sets of three-point temperature measurements made at three locales: Gaithersburg, Corbin (Virginia), and Saugus-Palmdale. Three of these five sets of measurements were made in conjunction with two of the described refraction experiments; all of the indicated measurements were made by members of the National Geodetic Survey.

MEASURED AND DERIVATIVE TEMPERATURE VALUES

For each of the five data sets temperatures were measured at 0.5, 1.5, and 2.5 m above the ground surface; readings were made to the nearest 0.1°C at Gaithersburg and Corbin and to the nearest 0.1°F along the Saugus-Palmdale line. Aspirated temperature sensors were used in all of the field measurements (Whalen, 1980, p. 777; 1981, p. 3; Stein and others, 1986, p. 9035; E.I. Balazs, 1993, oral commun.). The

Gaithersburg, Saugus-Palmdale, and Corbin temperature sensors were identical in their general design (Whalen, 1981, p. 3–5; Whalen and Strange, 1983, p. 3–5; Charles Glover, 1994, oral commun.). Each sensor consisted of a thermistor mounted within a tubular shield which was itself placed within a second, polished tubular shield. Aspiration was achieved by means of a small fan which drew outside air past the thermistor and then expelled it. The fans were themselves powered by the same battery pack, thereby providing the same voltage to all three fans. The averaging time—or time required in operation for equilibration between interior and exterior temperatures (C.T. Whalen, 1994, oral commun.)—is given as approximately one minute (Whalen and Strange, 1983, p. 3). Both the Gaithersburg observations and the Saugus-Palmdale temperature measurements utilized analogue systems, whereas the much more recent Corbin data were based on digitally recorded measurements (Charles Glover, 1994, oral commun.). The Gaithersburg temperatures were recorded manually from a bank of three readouts, each of which was connected to its own sensor; the Saugus-Palmdale and Corbin measurements were recorded manually from a single readout, a procedure that required switching to successively lower sensors (Whalen, 1981, p. 4; Whalen and Strange, 1983, p. 3–5; E.I. Balazs, 1994, oral commun.). No information survives regarding the time required to complete each set of three readings at Gaithersburg; temperature measurements for the Saugus-Palmdale experiment were read from the top down and took about one-half minute to complete (G.W. Adams, 1994, oral commun.).

Three-point temperature measurements toward the base of the boundary layer permit the determination of several temperature-related functions, four of which are included here:

$t_{1.5}$ = temperature at 1.5 m above the ground surface (the approximate height of the instrument at which temperatures are generally read during any routine leveling);

$t = (t_{0.5} + t_{1.5} + t_{2.5})/3$;

$\Delta t' = t_{0.5} - t_{2.5}$ (equates with $\Delta t'$ given in equation 1);

$CURV = t_{0.5} + t_{2.5} - 2t_{1.5}$ (an index of curvature in the thermal gradient; numerically identical to a finite difference approximation to the second derivative of t with respect to height since the thermistors were spaced one meter apart).

The relations among these several parameters are fundamental to an assessment of the sense and magnitude of the URE. Under the daytime and normally unstable atmospheric conditions at a given site, where the heat flux is positive, increasing values of t (or $t_{1.5}$) are generally accompanied by increasing values of $\Delta t'$. Similarly, as $\Delta t'$ increases, *CURV* tends to increase as well. Should the heat flux drop to very low or even negative values, the preceding generalizations tend to break down. *CURV* is so defined that positive values are consistent with a normal, concave upward thermal gradient (fig. 1) and negative URE values; negative values for *CURV* indicate a reversal in the normal concavity and an expectation for positive URE values.

Means and medians for $t_{1.5}$, t, $\Delta t'$, and *CURV* have been calculated here from the temperature measurements obtained from each of the five data sets. Three sets of comparisons and derivative linear regressions have also been calculated for each of these data sets. These include: (1) $\Delta t'$ versus $t_{1.5}$, (2) *CURV* versus $t_{1.5}$, and (3) *CURV* versus $\Delta t'$. Among the five sets of temperature data considered here, only one (that obtained from measurements at the Gaithersburg instrument station) produced results that are both ambiguous and of doubtful statistical significance.

MEASUREMENT ACCURACY

Measuring temperatures to an accuracy suitable to reliable determinations of $\Delta t'$ and, especially, the curvature in the thermal gradient, can be quite challenging. Although no temperature sensor calibration records survive for those periods in which the Gaithersburg field measurements were made, these sensors were tested against standardized thermometers in advance of the experiment by the NGS Instrument and Equipment Branch and shown to agree with the standardized thermometers to within the reading precision (±0.1°C) of the meters (Whalen, 1981, p. 8). Subsequent comparative temperature measurements of those sensors of Gaithersburg (1979) vintage against those used as recently as 1992 disclosed no distinguishable differences in measurement accuracy; each tracked the other with sufficient fidelity that no bias could be detected (Charles Glover, 1994, oral commun.). The Corbin sensors were calibrated prior to any experimental observations; they were rejected for field use if they failed to meet a ±0.2°C tolerance with respect to the National Bureau of Standards temperature standard (E.I. Balazs, 1993, oral commun.). The Instrumentation and Equipment Section of the National Geodetic Survey checked the Saugus-Palmdale sensors in advance of the experiment and found that they agreed with each other to within ±0.1°C; a check against standard thermometers following the test (September 25, 1981) indicated that the 0.5-m sensor consistently read 0.1°C too low (Whalen and Strange, 1983, p. 4; Stein and others, 1986, p. 9035). Accordingly, and even though limited systematic error undoubtedly infected each of the five sets of field measurements, the Saugus-Palmdale set is the only one in which there is any clearly indicated evidence of systematic bias. Specifically, the Saugus-Palmdale field measurements could underestimate both $\Delta t'$ and *CURV* by as much as 0.1°C. Although we have no direct means of assessing the relative accuracy of these three sets of measurements, all were made in a manner that accorded with the standards and practices of the National Geodetic Survey; thus, it is unlikely that measurement accuracy figures significantly in the detection of the similarities and differences that emerge from these comparisons.

GAITHERSBURG

Temperature measurements were made at the Gaithersburg refraction test site in September 1979. The measurement periods that are specifically considered here coincided with those in which the presumably refraction-contaminated leveling measurements were made (see section on analysis of Gaithersburg-Tucson test results). The sky was generally clear during the course of the measurements; partly cloudy to mostly overcast conditions accompanied perhaps one-third of the observations. The temperature measurements were made directly over a thick, green turf; scattered trees around the fringes of the site are the only naturally occurring and possibly perturbing obstacles within the test range environment (Whalen, 1981, p. 3–4). People and artifacts that could have perturbed the thermal regime were concentrated near the instrument station; these artifacts included such things as the instruments themselves, tables, umbrellas and vehicles (Whalen, 1981, p. 3). The three-point measurements were made at two points within the test range—at the instrument (observing) station and at the nominal 1.5-m bench-mark station 60 m distant from the instrument station (fig. 2).

The temperature measurements at the two Gaithersburg stations disclosed more or less consistent, but, in part, sharply disparate results in both the means and the regressions. The means of $t_{1.5}$ and t agree quite well at the 60-m rod station, but the $t_{1.5}$ mean is about 0.1°C *below* that of t at the instrument station (table 3). Although the means of both $t_{1.5}$ and t are about 0.3°C higher at the instrument station than the rod station, this measurably significant difference is marginally within the expected difference between two stations 50 m or more apart. The mean $\Delta t'$ values at the two stations differ by about 0.03°C—well within the reasonably expected noise level of these observations. The most pronounced difference among the means is shown by the curvature in the thermal gradient: *CURV* at the instrument station is nearly 0.4°C, whereas that at the rod station is close to zero, a generalization that holds fairly well in comparisons involving the trimmed means and the medians as well. While other factors clearly intrude, the mean value for *CURV* at the instrument station—as well as the difference between the mean $t_{1.5}$ and mean t—probably is attributable in large measure to a singularly inaccurate temperature sensor at the instrument station.

Table 3. Data developed from three-point temperature measurements at the National Geodetic Survey test range in Gaithersburg.

	Number of observations	Mean (°C)	Standard deviation (°C)	Trimmed mean (°C)	Median (°C)
		INSTRUMENT STATION			
$t_{1.5}$	185	21.22	3.47	21.27	21.80
t	184	21.35	3.39	21.40	21.85
$\Delta t'$	184	.45	.39	.47	.50
CURV	184	.39	.62	.33	.30
		60-METER ROD STATION			
$t_{1.5}$	186	20.91	3.31	20.95	21.80
t	186	20.90	3.30	20.95	21.75
$\Delta t'$	186	.42	.42	.43	.40
CURV	186	−.04	.38	−.02	.00

The comparisons and corresponding regressions (figs. 5, 6) display more or less consistent differences. Regressions of $\Delta t'$ on $t_{1.5}$ produce modest and statistically significant coefficients at both the instrument and rod station (figs. 5*A*, 6*A*); the smaller and less statistically significant coefficient obtained from the instrument station measurements is attributable to the larger influence of the outliers on the regression at this station (compare figures 5*A* and 6*A*). Although the regression of *CURV* on $t_{1.5}$ produces a very statistically significant result at the instrument station, the relatively large negative coefficient is again strongly influenced by the outliers. The results obtained from the 60-m rod station, on the other hand, indicate that *CURV* varies around zero with respect to $t_{1.5}$, a determination consistent with the very low R^2 value. The plot of *CURV* against $t_{1.5}$ at the instrument station may identify the source of (and, hence, the error in) the large mean value for *CURV* at this station (table 3), even though this suspected error would have no impact on the apparently large negative coefficient. A straight-line, zero-slope fit to the data (fig. 5*B*) would project through a value for *CURV* at about 0.4°C, as contrasted with a comparable near zero value for *CURV* at the rod station (fig. 6*B*). The most likely explanation for this difference, and one that is consistent with both the close correspondence in the $\Delta t'$ means obtained from the two stations and the notably smaller value for the mean of $t_{1.5}$ compared with that for t at the instrument station (table 4), is that the 1.5-m sensor at the instrument station read ~0.2°C too low, whether on a relative or absolute scale. Regression of *CURV* on $\Delta t'$ at the rod

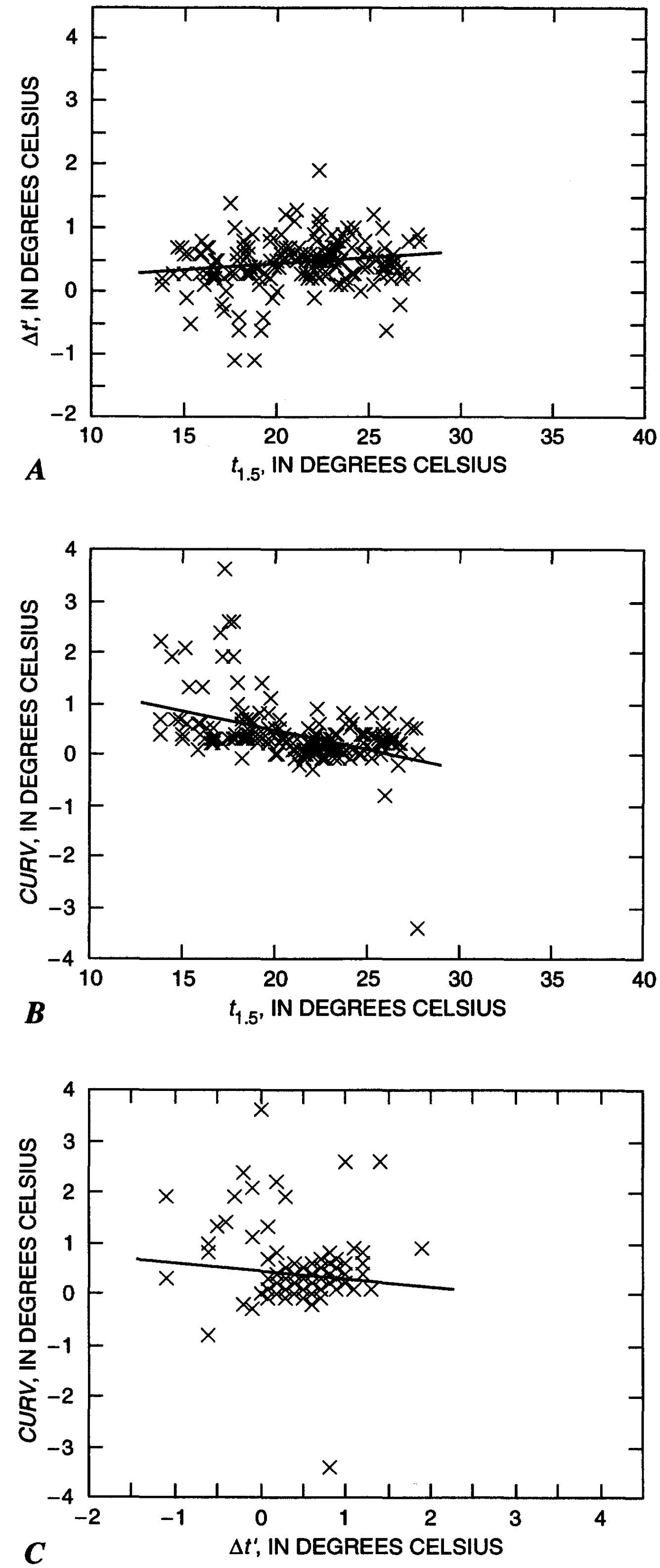


Figure 5. Comparisons among various temperature functions obtained from observations at Gaithersburg instrument station. Regression equations shown by straight lines. They are, in order: (*A*) $\Delta t' = (0.052 \pm 0.183) + (0.0189 \pm 0.0085)t_{1.5}$; R^2 = 2.7%; (*B*) $CURV = (1.86 \pm 0.270) - (0.0690 \pm 0.0125)\ t_{1.5}$; R^2 = 14.4%; (*C*) $CURV = (0.456 \pm 0.070) - (0.151 \pm 0.116)\ \Delta t'$; R^2 = 0.9%.

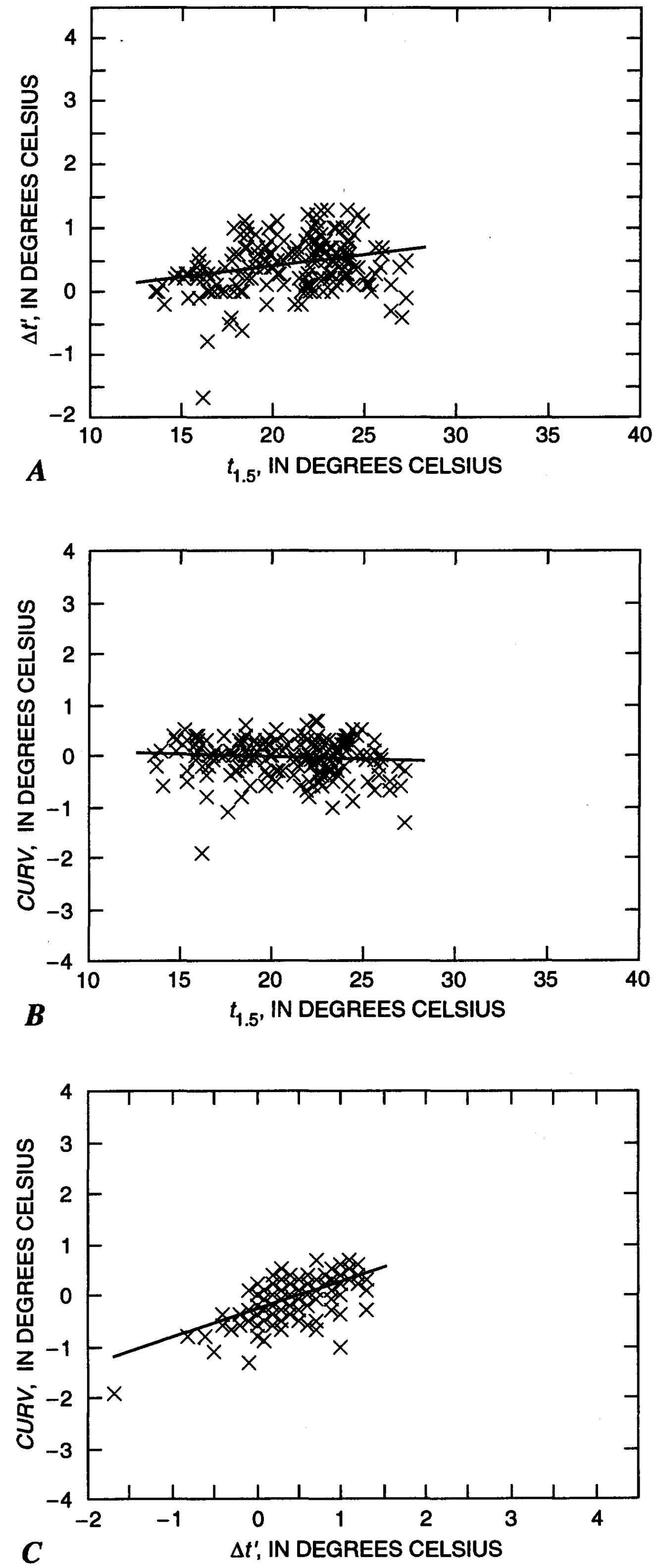


Figure 6. Comparisons among various temperature functions obtained from observations at Gaithersburg 60-m rod station. Regression equations shown by straight lines. They are, in order: (*A*) $\Delta t' = (-0.317 \pm 0.189) + (0.0353 \pm 0.0089)\ t_{1.5}$; R^2 = 7.8%; (*B*) $CURV = (0.137 \pm 0.180) - (0.0083 \pm 0.0085)\ t_{1.5}$; R^2 = 0.5%; (*C*) $CURV = (-0.265 \pm 0.032) + (0.543 \pm 0.054)\ \Delta t'$; R^2 = 35.1%.

Table 4. Data developed from three-point temperature measurements at the National Geodetic Survey test range in Corbin.

	Number of observations	Mean (°C)	Standard deviation (°C)	Trimmed mean (°C)	Median (°C)
		INSTRUMENT STATION			
$t_{1.5}$	524	27.85	2.74	27.92	28.05
t	524	27.82	2.76	27.89	28.05
$\Delta t'$	524	.33	.46	.28	.20
CURV	524	−.09	.53	−.10	−.10
		30-, 45-, AND 60-METER ROD STATIONS			
$t_{1.5}$	524	28.16	2.76	28.23	28.50
t	524	28.14	2.73	28.21	28.37
$\Delta t'$	524	.27	.31	.27	.20
CURV	524	−.05	.34	−.05	−.10

station produces a very statistically significant result that indicates a strong dependence of *CURV* on $\Delta t'$ (fig. 6*C*), a dependence that is certainly expected in view of our knowledge of turbulent exchange at the Earth's surface (see, for example, Shaw and Smietana, 1983). The absence of a comparable result obtained from the same regression at the instrument station is almost certainly attributable to the destruction of any likely correlation due to the scatter in the outliers (fig. 5*C*). Moreover, if we have correctly identified the source of the apparently discrepant mean value for *CURV* at the instrument station, a reduction in all of the values for *CURV* shown in figure 5*C* by 0.4°C would bring this plot into much closer agreement with that of the rod station (fig. 6*C*). In fact, simply raising all of the values for the 1.5-m sensor by 0.2°C brings the means for $t_{1.5}$ and t into agreement—at 21.42°C—and reduces the anomalous value for *CURV* to −0.01°C (compare with table 3). However, this same operation had no apparent impact on the regressions.

CORBIN

The Corbin temperature measurements were made at the National Geodetic Survey test range in Corbin, Virginia, in August 1992. The sky was generally cloudy or, less commonly, partly overcast; none of the observations was associated with cloudless conditions. The total environmental regime at Corbin is very similar to that at Gaithersburg. Specifically, the measurements were made over a gently sloping, thick green turf, where scattered trees around the margin of the site are the only naturally occurring obstacles that might have affected the temperature measurements (E.I. Balazs, 1993, oral commun.). People and artifacts were again concentrated around the instrument station, although their distribution probably varied from day to day. The three-point temperature measurements were made at four places within the test range; the instrument (observing) station and rod stations 30, 45, and 60 m from the instrument station. Each set of three-point rod-station measurements was accompanied by a corresponding set of observations at the instrument station, such that the total number of observations at the rod stations precisely matched that at the instrument station.

The two sets of temperature measurements obtained from the Corbin test range produced results that—with one obvious exception—agree fairly well. The several means of the two sets are internally consistent, and differences between the two are disclosed chiefly in the means for $t_{1.5}$ and t (table 4). The means for both the $t_{1.5}$ and t measurements were about 0.3°C lower at the instrument station than they were at the rod stations and, hence, close to the maximum likely difference that might be expected. The means of the $\Delta t'$ values held to within 0.06°C of each other, and those for *CURV* differed by only about 0.04°C and were small and negative in both cases (table 4). The preceding observations are equally applicable to the trimmed mean and median values. What may be the most revealing difference between the instrument and rod-station observations is the much larger standard deviation associated with the $\Delta t'$ measurements at the instrument station, as contrasted with the rod stations (table 4).

The comparisons and regressions derived from the Corbin measurements are quite similar, yet at least modestly different in comparing the two data sets (figs. 7, 8). Nonetheless, the similarities between the two data sets are much more impressive than their differences. The most obvious difference, and one which explains most of the other differences as well, is the notably larger number of outliers in the data obtained from the instrument station. For example, if one disregards the outliers that seem to peak at about 28°C at the instrument station (fig. 7*A*), the plots of $\Delta t'$ and *CURV* against $t_{1.5}$ at the two stations are certainly similar in appearance, even though the regressions produce statistically significant results at the rod stations and nothing of consequence at the instrument station (compare figs. 7*A* and 8*A* and figs. 7*B* and 8*B*). The pattern of outliers explains as well the higher standard deviations associated with the $\Delta t'$ and *CURV* means at the instrument station (table 4). The statistically significant regression of *CURV* on $\Delta t'$ at the instrument station (fig. 7*C*), in spite of the outliers, may be an expression of nothing more than chance or some unassessed atmospheric effect.

SAUGUS-PALMDALE

The Saugus-Palmdale measurements were made in conjunction with the LSL survey during May and June

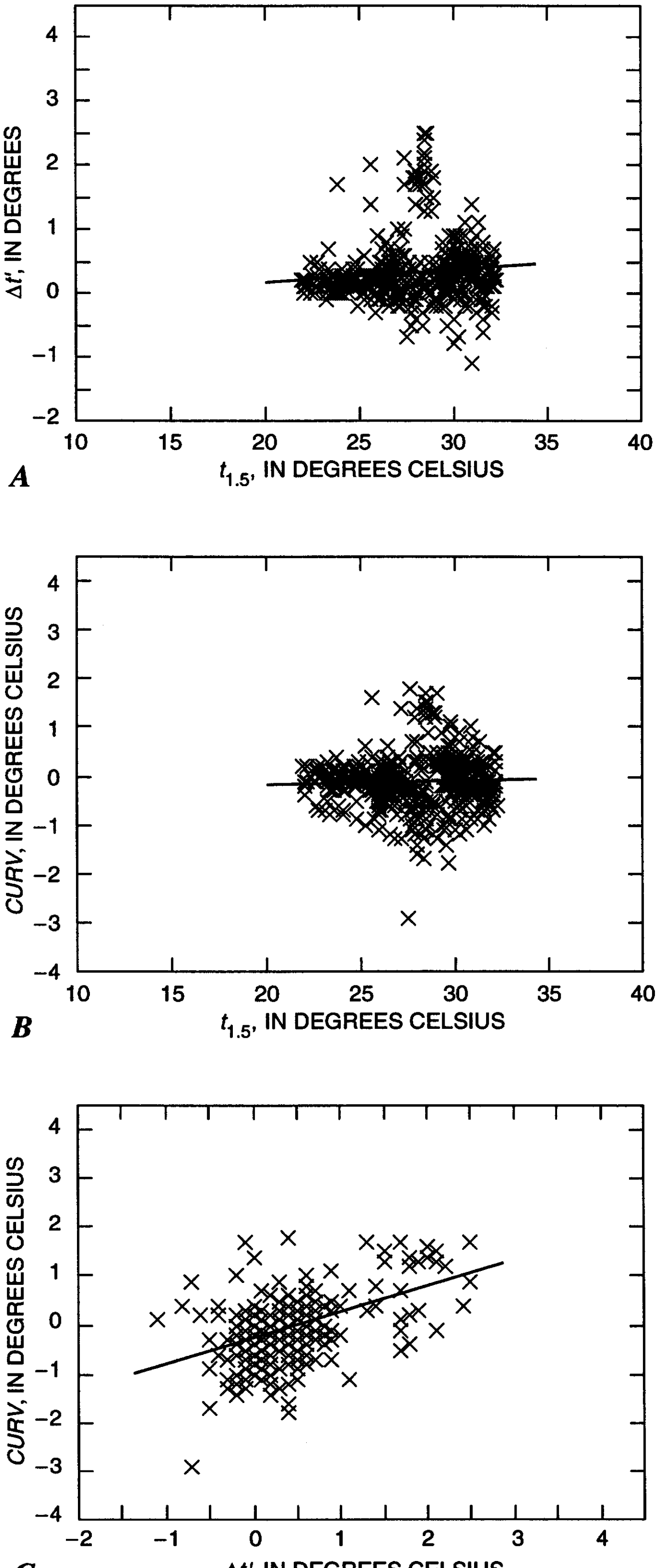


Figure 7. Comparisons among various temperature functions obtained from observations at Corbin instrument station. Regression equations shown by straight lines. They are, in order: (*A*) $\Delta t' = (-0.166 \pm 0.206) + (0.0177 \pm 0.0074)\ t_{1.5}$; $R^2 = 1.1\%$; (*B*) $CURV = (-0.333 \pm 0.236) + (0.0088 \pm 0.0084)\ t_{1.5}$; $R^2 = 0.2\%$; (*C*) $CURV = (-0.261 \pm 0.025) + (0.526 \pm 0.044)\ \Delta t'$; $R^2 = 21.3\%$.

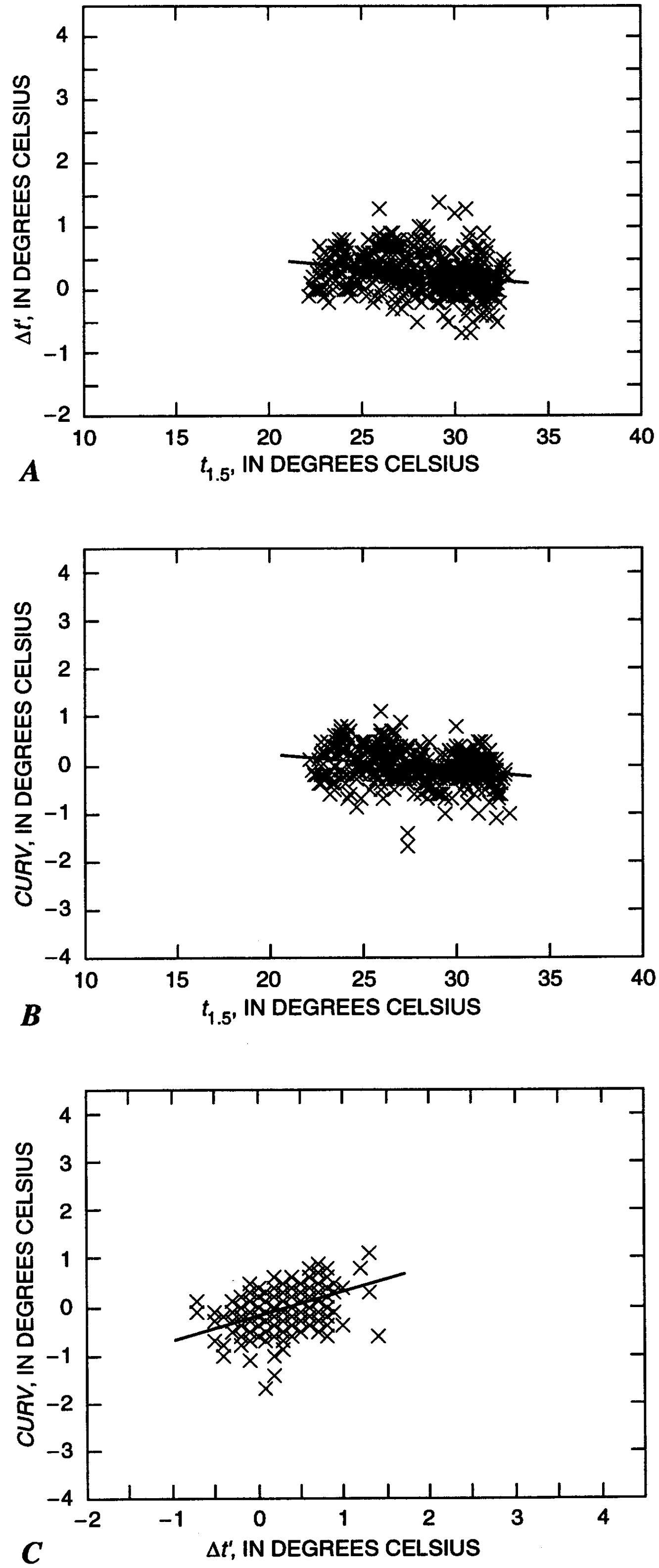


Figure 8. Comparisons among various temperature functions obtained from observations at Corbin 30-, 45-, and 60-m rod stations. Regression equations shown by straight lines. They are, in order: (*A*) $\Delta t' = (1.04 \pm 0.134) - (0.0274 \pm 0.0047)\ t_{1.5}$; $R^2 = 6.0\%$; (*B*) $CURV = (0.940 \pm 0.145) - (0.0353 \pm 0.0051)\ t_{1.5}$; $R^2 = 8.3\%$; (*C*) $CURV = (-0.193 \pm 0.017) + (0.517 \pm 0.042)\ \Delta t'$; $R^2 = 22.3\%$.

Table 5. Data developed from three-point temperature measurements made in conjunction with the Saugus-Palmdale LSL leveling.

	Number of observations	Mean (°C)	Standard deviation (°C)	Trimmed mean (°C)	Median (°C)
$t_{1.5}$	590	23.46	5.87	23.42	23.81
t	590	23.58	5.92	23.54	23.85
$\Delta t'$	590	1.25	.75	1.21	1.17
CURV	590	.34	.50	.33	.28

of 1981. The sky was generally clear, especially toward the Palmdale end of the line; overcast conditions were associated with no more than about one-third of the temperature observations. The general environment of this locale is that of a semi-arid climate within a canyon of moderate relief. There is, moreover, a good deal of variation in both roughness and the surface materials over which the temperature measurements were made. Because each leveling setup was accompanied by a three-point set of temperature measurements, this procedure is thought to have produced a good, representative sample of late spring temperatures along this line.

The relatively and predictably noisy character of the Saugus-Palmdale measurements is shown by both the means and the several comparisons (table 5; fig. 9). This attribute is especially evident in the relatively large standard deviations shown for both the $t_{1.5}$ mean and that for t (table 5). However, while the standard deviations for both $\Delta t'$ and *CURV* are relatively large in absolute terms, they are small in proportion to the means with which they are associated (table 5). Moreover, and in spite of the noisy data, all of the comparisons among $t_{1.5}$, $\Delta t'$, and *CURV* are characterized by very statistically significant regressions (fig. 9).

SIMILARITIES AND DIFFERENCES AMONG THE RESULTS OF THE COMPARATIVE TEMPERATURE MEASUREMENTS

Although some very obvious differences exist among the five sets of three-point temperature measurements examined here, the similarities may be more significant.

The rod-station observations at Gaithersburg and Corbin differ less from each other than they do from the corresponding instrument-station observations. That is, even though the mean temperature and $\Delta t'$ were quite different at Gaithersburg and Corbin, the instrument-station results obtained at these two sites are both characterized by a disproportionately larger number of outliers and less significant regressions than those obtained from the rod-station measurements. We speculate that these seeming differences and similarities may be due to the impingement of undefined artifacts on the

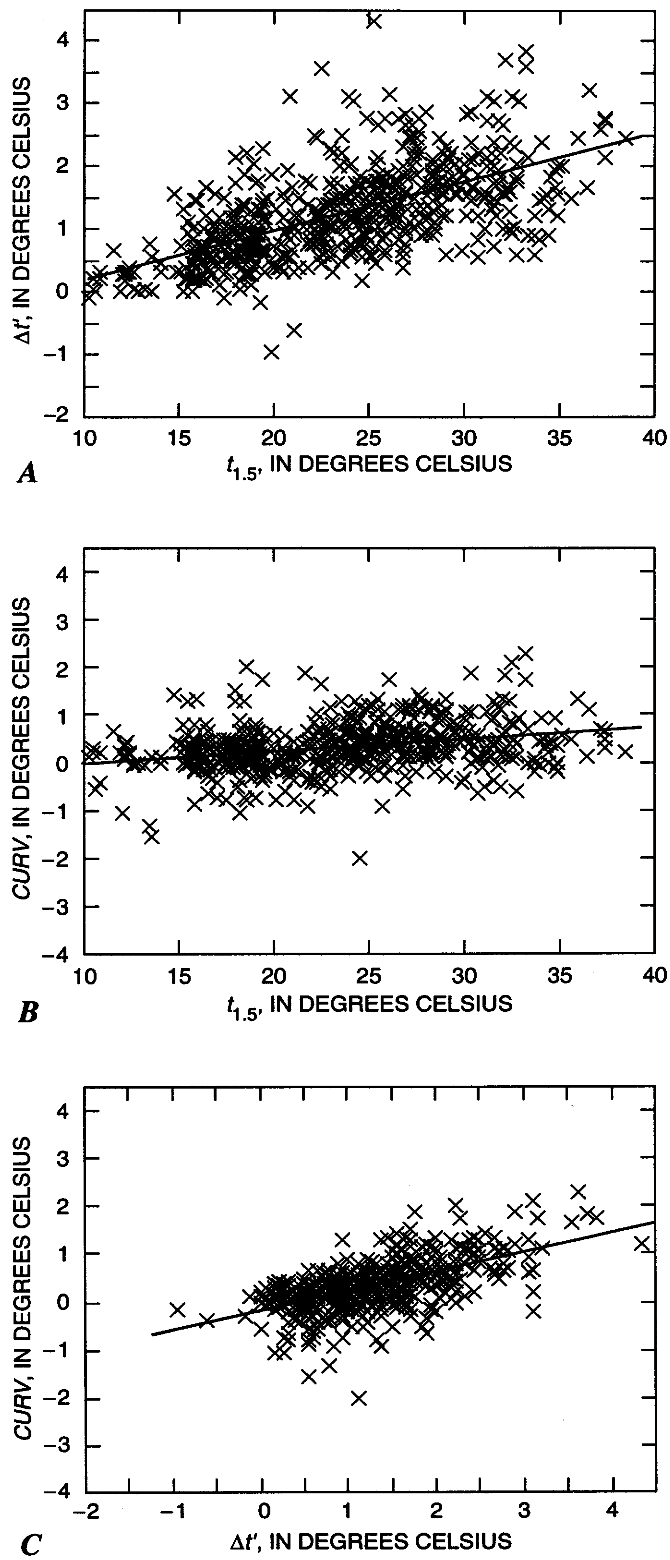


Figure 9. Comparisons among various temperature functions obtained from observations made in conjunction with LSL leveling along Saugus-Palmdale line. Regression equations shown by straight lines. They are, in order: (*A*) $\Delta t' = (-0.595 \pm 0.099) + (0.0786 \pm 0.0041)t_{1.5}$; $R^2 = 38.3\%$; (*B*) $CURV = (-0.261 \pm 0.082) + (0.0254 \pm 0.0034)t_{1.5}$; $R^2 = 8.7\%$; (*C*) $CURV = (-0.170 \pm 0.033) + (0.404 \pm 0.022)\ \Delta t'$; $R^2 = 35.6\%$.

thermal structure of the atmosphere at the instrument stations. This generalization, coupled with the possible or probable failure of the 1.5-m temperature sensor at the Gaithersburg instrument station, argue that the rod-station measurements are the more representative at both sites. Thus, they are the ones that should govern any assessment of the impact of these temperature measurements on the URE.

Accepting the preceding argument, the similarities in the results of the temperature measurements obtained from the Gaithersburg and Corbin sites far outweigh the differences (tables 3, 4; figs. 6, 8). Nevertheless, and confining ourselves exclusively to comparisons between the two sets of rod-station means and regressions, some significant differences emerge. Although the mean $t_{1.5}$ is significantly lower at Gaithersburg, the $\Delta t'$ mean is clearly greater (tables 3, 4). Moreover, while the regression of $\Delta t'$ on $t_{1.5}$ produces a shallow but predictably positive slope at Gaithersburg, this slope is similarly shallow but negative at Corbin (figs. 6*A*, 8*A*). *CURV* not only roughly randomizes around zero with respect to $t_{1.5}$ at both sites, but seems to be decreasing with increasing $t_{1.5}$ at Corbin (figs. 6*B*, 8*B*). These observations are by themselves consistent with very low values for the sensible heat flux at the two sites. Probably the most striking results obtained from the three-point temperature measurements at Gaithersburg and Corbin are shown by the regressions of *CURV* on $\Delta t'$ (figs. 6*C*, 8*C*). Recalling that we are dealing not only with different sites, but temperature measurements made 13 years apart, the correspondence between these two plots and the accompanying regressions is quite remarkable. The split between positive and negative values for *CURV* at both sites is roughly fifty-fifty, and, except for the higher density of observations at Corbin, the two plots are virtually indistinguishable (figs. 6*C*, 8*C*). While both of these data sets are based on derivative determinations obtained from the temperature measurements, it is likely that the heat flux under these presumably unstable conditions was no more than a small fraction of that along the Saugus-Palmdale line.

If, in fact, the heat flux at the Gaithersburg and Corbin sites was near negligible, some rationale should exist as to not only why this should be the case, but why it should be true even though the averaged temperatures at Corbin were about 7°C above those at Gaithersburg (tables 3, 4). An answer to the second part of this conundrum can be found, at least in part, in the difference in cloud cover. The cloud cover over the Gaithersburg site was generally much less than that at Corbin, with a corresponding and relatively large reduction in solar radiation at Corbin. While differences in cloud cover may account for the presumed similarities in heat flux, in spite of the difference in averaged air temperatures, this does not explain the apparently low heat flux at both sites. As it relates to the problem at hand, what these two sites have in common is a thick, green turf base. Moreover, the surface and near-surface moisture at both sites probably ranged from very moist to moderately moist; it is unlikely that either site was ever moisture free. Thus, that part of the thermal input taken up by the heat of vaporization of water associated with evaporation and transpiration was at the same time subtracted from that which might otherwise have contributed to the upward sensible heat. Because transpiration generally increases with increasing temperature (directly analogous to pan evaporation), the near zero to possibly decreasing values of *CURV* with increasing temperatures is clearly consistent with the notion that this heat flux was damped to trivial values owing to the impact of both direct evaporation and, especially, the actively transpiring turf canopies.

The chief difference in the temperature characteristics between the Gaithersburg and Corbin sites and the LSL Saugus-Palmdale line is expressed in the much greater (and statistically very significant) dependence of $\Delta t'$ on $t_{1.5}$ along the Saugus-Palmdale line (fig. 9*A*). This difference cannot be attributed simply to relatively elevated temperatures along the Saugus-Palmdale line, since the average temperature was well below that at Corbin and only modestly above that at Gaithersburg (tables 3, 4, 5). What obviously separates those two groups is the environment. The arid to semiarid regime that characterizes the Saugus-Palmdale line is associated with a sparse vegetative ground cover and reduced availability of surface or near-surface moisture. Damping of the heat flux attributable to evapotranspiration was much less along the Saugus-Palmdale line than it was at Gaithersburg and Corbin. The result is not only a statistically significant and positive relation between *CURV* and increasing $t_{1.5}$, but an even more impressive dependence of *CURV* on $\Delta t'$ (fig. 9*B*, *C*). Illustrative of the latter dependence is that 80 percent or more of the values for *CURV* are positive—quite unlike the distribution at Gaithersburg and Corbin (figs. 6*C*, 8*C*).

Features that are common to all the regressions of *CURV* on $\Delta t'$—except that obtained from the Gaithersburg instrument station, where it is indeterminate—are the relatively uniform slopes and the fact that the *CURV* intercept is consistently negative, generally statistically significantly so. While inconsistent with theory, in the sense that the regressions do not pass through the origin, this feature is not an artifact of the regression. Values for $\Delta t'$ identified with zero *CURV* range from 0.373°C to 0.496°C. This apparently consistent relation suggests a limiting value for $\Delta t'$, regardless of environment, that must be reached before any significant curvature moves into the positive range. One possible explanation for the existence of this seeming threshold for $\Delta t'$ is that the heat flux associated with these modest $\Delta t'$ values is so low that it fails to impart significant curvature to the thermal gradient above 0.5 m. This view is consistent with the results of at least one day of observations based on an increasingly dense concentration of temperature sensors toward the ground (Heer and Niemeier, 1985, p. 334). This example, which was identified with an ambient air temperature of only about 9.5°C and a $\Delta t'$ of about 0.2°C, indicates that nearly all the quite significant curvature in the thermal gradient occurred below 0.5 m.

Finally, we can compare the directly determined values for *CURV* at 1.5 m (tables 3–5) with the theoretically predicted values based on the measured and averaged values of $\Delta t'$ and the expression:

$$1.0 \leq -\left[\frac{zd^2t/dz^2}{dt/dz}\right] \leq 1.5.$$

This expression is obtained from an evaluation of the rate of change with height of the temperature gradient when expressed by an empirically derived formulation for thermally unstable conditions, such as that employed by Shaw and Smietana (1983, p. 10,487—equation 22). The smaller value (1.0) in the expression corresponds to near-neutral stratification, whereas the larger value (1.5) is representative of highly unstable conditions. With the further theoretical constraint that *CURV* and $\Delta t'$ should reach zero simultaneously, we can compare this expression with the profile features shown in figs. 6*C* to 9*C* (disregarding data from the Gaithersburg instrument station for the reasons stated earlier). Linear regressions of *CURV* on $\Delta t'$ disclose intercepts on the *CURV* axis ranging from −0.265 to −0.170 and slopes ranging upward from 0.404 to 0.543 (figs. 6*C* to 9*C*). The intercepts on *CURV* are small and only slightly larger than the expected accuracy of the individual thermistors. Regardless, and even though inconsistent with theory, we hesitate to dismiss completely these consistently negative intercepts as the products of measurement error. Based on the definition of *CURV* and $\Delta t'$ used here, the slopes should lie between 0.333 and 0.500 (one-third of the values in the theoretical expression, since $\Delta t'$ was measured over a height interval of 2 m and the middle instrument was 1.5 m above the surface). The regression slopes are somewhat larger than anticipated but, overall, they are in reasonable agreement with the theoretical values.

ANALYSES OF THE INDIVIDUAL EXPERIMENTAL RESULTS

The use of double-scale rods (fig. 4) in each of the described tests is at once (1) a vehicle for comparing the measured refraction values against the procedural constraints identified with the several orders and classes of 20th-century leveling, constraints intended largely to inhibit the accumulation of URE, and (2) the unifying feature that ultimately permits comparisons among the results of these experiments. Moreover, because earlier investigation (Castle and others, 1983, p. 1082–1083) suggests the existence of a probably locale-dependent threshold temperature, below which URE does not accumulate, we have examined the cumulative refraction or measured divergence as a function of not only $t_{1.5}$ (or mean t), but $\Delta t'$ and, where possible, *CURV*, in order to test for the sensitivity of the URE to each of these parameters. The existence of a locale- (and survey-) dependent threshold temperature should not be misconstrued; that is, there is almost certainly no such thing as a universal threshold temperature. By way of illustration, the movement of a subzero arctic air mass over the midwinter (but unfrozen) reaches of Lake Superior would produce a thermal gradient of precisely the configuration consistent with the generation of the URE—were the surface other than level.

THE GAITHERSBURG-TUCSON TEST RESULTS

Because the results of the Gaithersburg-Tucson tests are especially important in assessing refraction error as a function of survey order, we reiterate here the essential elements of an earlier analysis of these tests (Castle and others, 1985). The Gaithersburg-Tucson tests produced 12 sets of data. Six of these sets are identified with what were originally thought to be fairly accurate characterizations of the cumulative URE over each of the simulated sloping lines (table 1). The six sets of data obtained from the simulated lines of zero slope produced observed minus standard (O–S) values ranging between 9.2 mm and −10.8 mm (table 1), which we interpret as error associated with scintillation (simply the magnitude of the fluctuation in the rapidly changing atmospheric refraction along the line of sight), wind, advection, and so on, but clearly devoid of any systematic refraction error.

REFRACTION VERSUS STANDARD DEVIATION ABOUT THE MEAN DDH

The URE is theoretically proportional to the height difference between turns (Δh—equation 1); it is this property that permits the grouping of the Gaithersburg-Tucson refraction-contaminated measurements as if the only variables in the observations were sight length, temperature and (or) $\Delta t'$. In order to exploit this property, however, the cumulative refraction errors must be normalized to a common height difference—arbitrarily chosen as the roughly median 491.4-m height difference associated with the simulated 30-m sight-length line at Gaithersburg (table 1). Moreover, because both rod scales were read at each setup, a standard deviation about the mean of the DDH—the high-scale low-scale difference (fig. 4)—can be computed for each of the 12 data sets developed from the Gaithersburg-Tucson tests (table 1). Regression of the normalized refraction measurements (NR) on the corresponding σ_{DDH} (the standard deviation about the mean DDH) values defines a statistically significant trend that clearly shows the correlation between NR and σ_{DDH} (fig. 10). Thus:

$$NR = 4.5 - 651\sigma_{DDH}^2 \qquad (2)$$

Whalen (1981, p. 11) noted, appropriately and importantly, that the refraction errors that emerged from the Gaithersburg-Tucson tests are "worst-case" values, simply because the setup elevation differences in any routine leveling would generally average much less than the nominal 2-m difference imposed on the experimental design. To state the same conclusion in slightly different terms, because currently stipulated leveling procedures require that "the line of sight should never be nearer than 0.5 m to the ground at any point" (Federal Geodetic Control Committee, 1980, p. 32) and because the curvature in the thermal gradient increases downward (Shaw and Smietana, 1983, p. 10,488) (fig. 1), nominal 0.5-m readings on the upslope rods should maximize the URE. With modest qualification, the same generalization applies to early-20th-century (pre-

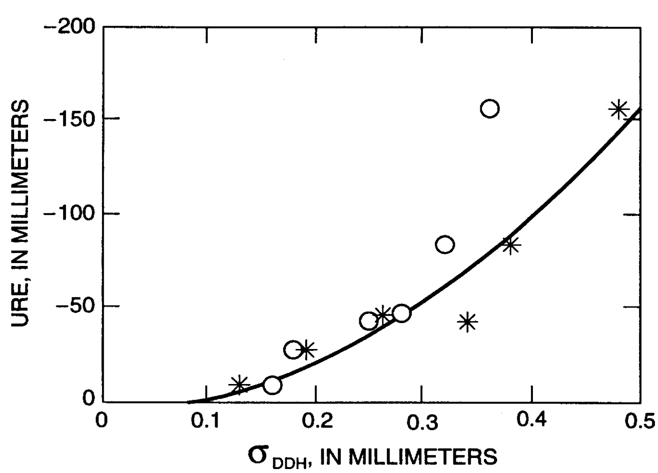


Figure 10. Standard deviation about the mean DDH (σ_{DDH}) versus summed URE measurements (asterisks) normalized to a common height difference of 491.4 m for the six simulated sloping lines obtained from the Gaithersburg-Tucson experiments (fig. 2; table 1, columns 9 and 11). Regression equation of NR (normalized refraction) on σ_{DDH} is NR = 4.5 − $651\sigma_{DDH}^2$, where all values are in millimeters; R^2 = 92.4%. Open circles based on values of σ_{DDH} computed from foresight-backsight nominal rod readings of 1.5 m for each of the six zero-slope simulated lines (fig. 2; table 1, column 10), all aligned with corresponding sloping-line (refraction-contaminated) values of identical sight length. Modified from Castle and others (1985).

1975) levelings (see Federal Geodetic Control Committee, 1975, p. 25), where the observer was admonished that the lowest of the so-called three-wire readings obtained with the Fischer level should always be at least 0.3 m above the ground (Bowie and Avers, 1914, p. 22; Rappleye, 1948, p. 40)—the equivalent of a middle-wire reading of 0.35 m with 30-m sight lengths or 0.39 m with 60-m sight lengths. In other words, because the curvature in the thermal gradient increases toward the ground surface (fig. 1), there is some expectation that were we to reformulate the Gaithersburg-Tucson tests in order to accommodate the less rigorous early-20th-century requirements, they would produce somewhat greater worst-case refraction measurements. However, these presumably increased refraction values could also be expected to be associated with somewhat greater σ_{DDH} values, so that the impact of these larger errors on equation (2) probably would be minimal. This point can be examined from quite a different perspective that still compels the same conclusion.

Implicit in the preceding discussion is the view that σ_{DDH} is a reasonable measure of atmospheric scintillation. While this is a clearly defensible position, phenomena other than scintillation may contribute to σ_{DDH}, notably wind. Under modest to intense wind conditions, the instrument or, less likely, the rods may be set in vibration and the readings will tend to "dance" or blur. With respect to the problem at hand, the response to this condition in any procedurally constrained leveling would be the same as if σ_{DDH} were due entirely to scintillation—the observer would shorten sight lengths in order to meet the rejection criteria. With this caveat in mind, we can treat σ_{DDH} as if it were an explicit measure of scintillation.

Accepting the conclusion that the refraction errors obtained from the Gaithersburg-Tucson tests are worst-case values, this does not in itself demonstrate that equation (2) defines a worst-case curve (fig. 10). This is obviously the case in any rigorously statistical sense, since more than one-half of the defining points lie above the curve. Moreover, this curve characterizes a slightly less than worst-case scenario, inasmuch as the averaged standard elevation difference on the nominal 0.5-m and 2.5-m marks is about 1.95 m (table 1), or roughly 98 percent of a worst-case figure. More to the point, however, it is conceivable that refraction errors as large as those identified with their matching σ_{DDH} values given in figure 10 could, under different conditions, be associated with significantly smaller σ_{DDH} values. While we cannot disprove such a possibility, this eventuality seems particularly unlikely. Specifically, the σ_{DDH} values associated with the otherwise comparable zero-slope lines devoid of any URE closely match the corresponding sloping-line values up to about 0.3 mm (fig. 10). This correspondence implies that for values of σ_{DDH} of up to about 0.3 mm there is almost no way that worst-case refraction errors might be associated with matching σ_{DDH} values significantly smaller than those defined by the regression curve given in figure 10. Departures from this equivalence at 50- and 60-m sight lengths (those identified with σ_{DDH} values greater than 0.3 mm) are readily explained by the notably higher scintillation associated with those sights that tend to more closely hug the ground surface over a disproportionately greater segment of the sight than do the shorter sights (see, also, Castle and others, 1985, p. 241). Thus, we conclude, with some modification, that the regression curve presented in figure 10 does in fact closely approximate, a worst-case representation of the relation between the URE and σ_{DDH}. It is not, however, to be equated with the actual or true maximum URE as a function of σ_{DDH} which, as we shall see, is both more complex and poorly defined above σ_{DDH} values of ~0.18 mm.

WORST-CASE REFRACTION ERRORS AS A FUNCTION OF SURVEY ORDER

The approximate limits of the maximum likely refraction errors associated with the several orders of late-20th-century geodetic leveling can be developed from the regression curve given in figure 10 and σ_{DDH} values obtained from a mix of levelings characterized by a variety of setup-rejection limits. The data on which we are forced to rely are heavily weighted toward the higher-order surveys and, as a result, the reliability of these error estimates (the error fields) deteriorates as the survey rejection criteria become progressively more permissive. This skewed distribution of the data is related directly to the progressively fewer lower-order levelings based on double-scale rods—owing presumably to the less demanding requirements and looser procedural constraints stipulated for these lower-order surveys. By way of illustration, the setup-rejection criterion for third-order leveling of 1.30 mm specified by the Federal Geodetic Control Committee (1980, p. 28) is so loose that the flutter of the image on the rod at the rejection limit would create such a blurred observation that this rejection criterion was rarely, if ever, invoked—probably the reason that we have been unable to find a single double-scale rod leveling run to this setup-rejection limit.

The procedure adopted here was simply to work backward from the controlling setup-rejection criterion (and hence the order and class of the survey) to the corresponding σ_{DDH} values computed from a representative set of routine levelings (tables 6, 7). Table 7 is included largely for purposes of comparison; these data were derived from levelings based on reversible compensator levels, automatic instruments that require no collimation correction. The setup-rejection criteria for reversible compensator instruments are much less restrictive than they are for comparable levels that do not operate in this mode—or have been since 1984 (compare Federal Geodetic Control Committee, 1980, p. 28; 1984, p. 3–7)—simply because the compensator in positions 1 and 2 may not be precisely symmetrical about the plumb (see Schomaker and Berry, 1981, p. 3-20). In other words, the instrument itself could provide a contribution to the DDH unrelated to observing conditions. Were the sights always precisely balanced, any tendency toward an instrument-dependent contribution to the DDH would be nullified and the setup-rejection criteria should equal those found in table 6. This point is clearly implied by the reduction in σ_{DDH} where the setup imbalance for line A is limited to those observations in which the sight lengths differ by no more than 0.4 m (table 7). Given the significant differences in the setup-rejection limit for the same classes of leveling listed in tables 6 and 7, the σ_{DDH} values are in remarkably good agreement, agreement attributable largely to the requirement that the observer meet the same corresponding section-rejection criteria at the same tolerance level. It is this requirement that constrains the observer to control sight-lengths as a function of observing conditions. Even though we have not included the data in table 7 in developing our characterization of the URE as a function of survey order, these same data demonstrate the importance of the section-rejection limits in controlling sight lengths. They demonstrate as well, of course, that the inclusion of the reversible-compensator readings would have no more than slightly enlarged the magnitude of the estimated worst-case URE associated with any stipulated order or class of leveling.

The approximate upper limits of the URE fields for both first- and second-order, late-20th-century levelings are based on the σ_{DDH} values developed from surveys of specified setup-rejection limits stipulated for the several orders and classes of leveling (table 6, fig. 11*A*). However, owing to the general absence of double-scale rod levelings run to third-order standards, this straightforward procedure could not be used in estimating the worst-case refraction errors identified with third-order levelings. Accordingly, the upper limit for the third-order URE field is based on the presumption of a general correspondence between setup- and section-rejection criteria through the several orders of leveling and the likelihood that the σ_{DDH} band widths for both the second- and third-order error fields (fig. 11*A*) were about the same out to the 12 mm/km$^{1/2}$ section-rejection limit for third-order leveling—which is the same as the loop-rejection criterion for third-order leveling that has been in force since the turn of the century (Castle and others, 1987, p. 5).

As Stein and others (1986, p. 9040) have observed, "no rigorous setup rejection criteria existed before 1965" and did not, in fact, exist until 1978 (E.I.

Table 6. Standard deviation about the mean DDH (σ_{DDH}) matched against corresponding setup- and section-rejection criteria for six representative levelings.

Line designation	Number of consecutive setups	σ_{DDH} (mm)	Setup-rejection criterion (mm)	Section-rejection criterion[a] (mm/km$^{1/2}$)	Survey order and class[a]
A[b]	418	0.155	>0.30	>4.00	First, II
B[c]	1022	.139	>0.30+[d]	>4.00+	Less than First, II
B[c]	900	.080	>0.30	>4.00	First, II
C[e]	590	.286	>0.75	>8.00+	Less than Second, II
C[e]	506	.271	>0.70	>8.00	Second, II
C[e]	384	.244	>0.60	>6.00	Second, I
C[e]	276	.142	>0.50	>4.00+	Less than First, II
C[e]	102	.176	>0.40	>4.00+	Less than First, II
C[e]	22	.130	>0.30	>4.00	First, II
D[f]	375	.161	>0.30	>4.00	First, II
E[g]	328	.165	Unspecified	>3.00	First, I
F[h]	488	.206	Unspecified	>8.40	Second, "O"

[a]Matching section-rejection criteria and survey order and class from Federal Geodetic Control Committee (1980, p. 28) and Rappleye (1948, p. 2).
[b]Wilson (1983).
[c]Short-sight-length survey (Teledyne Geotronics, 1981).
[d]Includes sections (6 out of 60) in which the DDH/setup exceeded 0.30 mm and ranged as high as 0.45 mm.
[e]Long-sight-length survey (Teledyne Geotronics, 1981).
[f]R.C. Wilson (1984, written commun.).
[g]National Geodetic Survey line L-22449; one setup reading dropped owing to illegible copy.
[h]National Geodetic Survey line L-22643; five setup readings dropped owing to illegible copy.

Table 7. Standard deviation about the mean DDH (σ_{DDH}) matched against corresponding setup- and section-rejection criteria for two representative reversible compensator levelings uncorrected for collimation error.

Line designation	Number of setups	σ_{DDH} (mm)	Setup imbalance (m)	Setup-rejection criterion[a] (mm)	Section-rejection criterion[a] (mm/km$^{1/2}$)	Survey order and class
A[b]	693	0.172	≤5.0[c]	>1.00[d]	>4.00	First, II
A[b]	137	.160	≤ .4	>1.00	>4.00	First, II
B[e]	18	.226	≤10.0[c]	>2.00[f]	>8.00	Second, II
B[e]	17	.230	≤ .8	>2.00	>8.00	Second, II

[a]Matching setup- and section-rejection criteria from Federal Geodetic Control Committee (1984, p. 3–7).
[b]National Geodetic Survey line L-25365.
[c]Maximum permitted foresight-backsight imbalance from Federal Geodetic Control Committee (1984, p. 3–7).
[d]The DDH exceeded 0.30 mm in 30 setups out of total of 693; none exceeded 0.70 mm.
[e]National Geodetic survey Line L-25271.1.
[f]The maximum DDH recorded for any setup was 0.40 mm.

Balazs, 1992, oral commun.) (compare also Federal Geodetic Control Committee, 1974, p. 9, and 1980, p. 28). Accordingly, we have resorted to an indirect procedure in order to establish the URE fields for the several pre-1978 (or, equally likely, pre-1975) orders of leveling. Specifically, the setup-rejection criteria for modern leveling are so stipulated that the section-rejection limits should meet a 5 percent tolerance level (Federal Geodetic Control Committee, 1975, p. 23). The pre-1975 section-rejection tolerance is given by both Bowie and Avers (1914, p. 11–12) and Rappleye (1948, p. 7) as 5 to 15 percent; thus, we take 10 percent as a reasonable mean (compare with Federal Geodetic Control Committee, 1975, p. 23). Accepting this value, a statistical projection based on the two tails of an assumed normal distribution curve indicates that the hypothetical pre-1978 (or pre-1975) double-scale rod setup-rejection criteria (DDH) should have been about 1.12 times that given for modern leveling. Because the σ_{DDH} values obviously are controlled by the setup-rejection limits, we have expanded the post-1977 URE fields by a factor of 1.12 to obtain their pre-1978 equivalents (fig. 11*B*). Although the upper limit on the timeframe given in figure 11*B* might be pushed back to as early as 1965, we have recovered no documentation suggesting that the pre-1975 section-rejection tolerance level was ever any more restrictive than about 10 percent. We should note here that even were we to have adopted an expansion factor of 1.25 or 1.30, the worst-case refraction errors identified with any specified

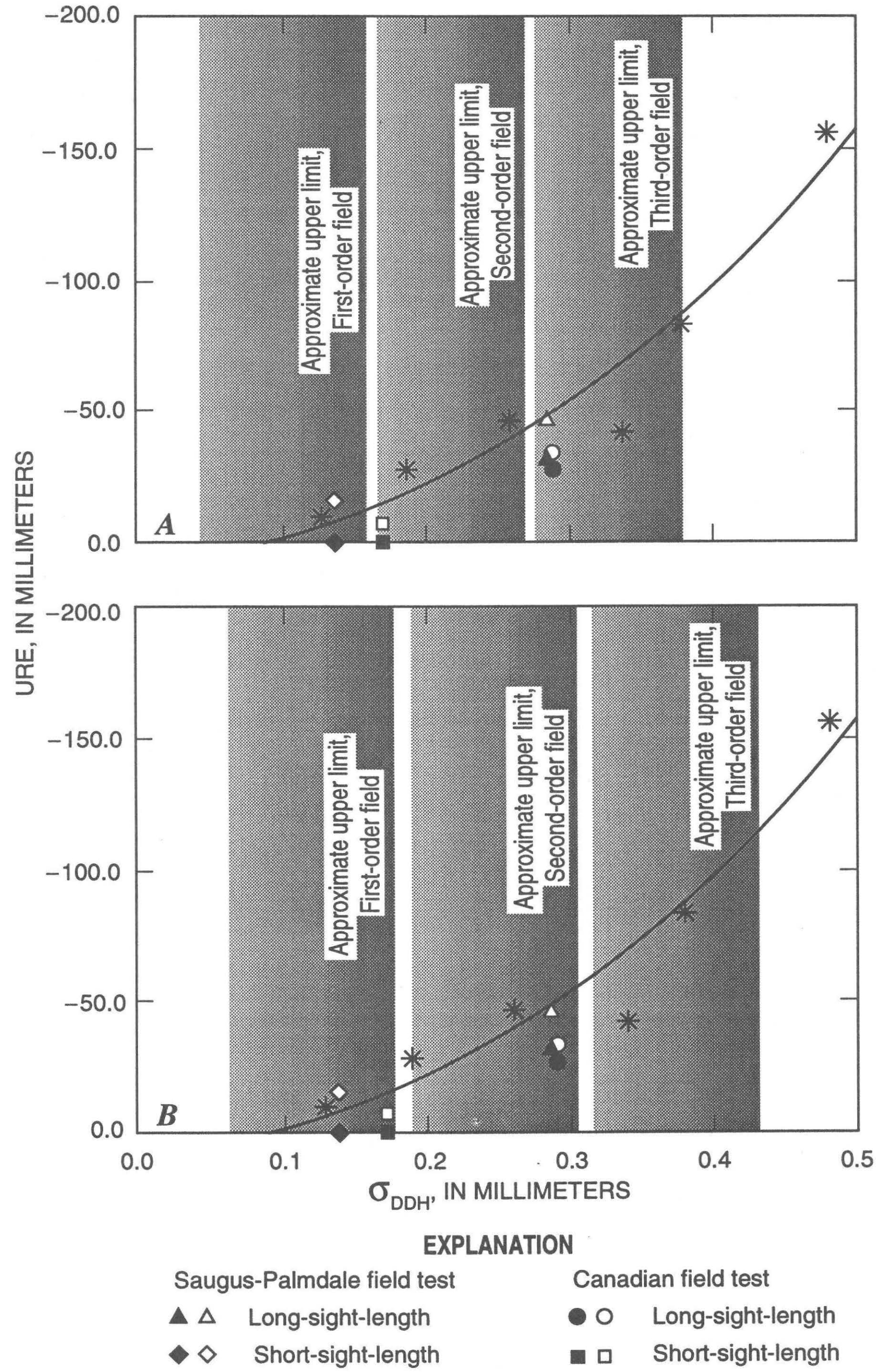


Figure 11. Standard deviation about the mean DDH (σ_{DDH}) versus summed and normalized refraction measurements (asterisks) obtained from the Gaithersburg-Tucson refraction tests (see fig. 10) and approximate maximum values of refraction errors associated with the several orders of leveling based on σ_{DDH} values drawn chiefly from routinely produced double-scale rod levelings of specified order (table 6). *A*, 1975 and later error fields as a function of survey order. *B*, 1974 and earlier error fields as a function of survey order. Various symbols show σ_{DDH} obtained from LSL and SSL surveys for the Saugus-Palmdale and Canadian field tests plotted against cumulative refraction errors normalized to a common height difference of 491.4 m. Computed field-test refraction errors based on assumption that LSL-SSL divergence is due exclusively to atmospheric refraction error (URE) which is: (1) confined entirely to the LSL surveys (solid symbols) or, alternatively, (2) directly proportional to average sight length squared (open symbols). Saugus-Palmdale values based on field data and absolute height (704.5 m) and maximized divergence (–44.2 mm) concept of Stein and others (1986, p. 9034); use of corrected data enlarges long-sight-length solid-symbol by –4.1 mm and the long-sight-length and short-sight-length open symbol values by –6.0 and –1.5 mm, respectively (short-sight-length solid symbol remains unchanged). Modified from Castle and others (1985) and data developed by Heroux and others (1985), as modified by W. Gale (1992, written commun.). See text for details.

order of leveling would have increased by no more than 15 to 20 percent. That the 1.12 expansion factor is reasonable or even conservative is suggested by at least two observations: (1) As we have already indicated, reformulation of the Gaithersburg-Tucson tests to accommodate the more permissive pre-1975 minimum rod reading would have only slightly increased the worst-case errors for any specified order of leveling. (2) Even with the especially permissive setup-rejection criteria of 1.00 mm for reversible compensator first-order leveling, the section-rejection criterion of 4.00 mm/km$^{1/2}$—identical to that in force since the turn of the century—imposed on this leveling constrained σ_{DDH} to 0.172 mm (table 7), a value marginally within the pre-1975 first-order URE field (fig. 11*B*).

URE VERSUS $t_{1.5}$, $\Delta t'$, AND *CURV*

A comparison of the cumulative URE obtained from the simulated Gaithersburg 60-m line against setup temperature, where the setups are ordered according to temperature (fig. 12), is at once revealing and inconsistent with what we would have inferred from Whalen's (1980; 1981) earlier reports. We restricted our examination to the 60-m line simply because it produced the largest of the three measured URE (O–S) values and was characterized by the lowest noise/signal ratio (table 1). Our corrections to the observed data are based on the same, relatively simple procedures used originally by Whalen (1980), which produced a summed refraction error of –65 mm, as contrasted with his subsequent, more rigorously corrected value of –77.5 mm (table 1) (Whalen, 1981, p. 12). The form of the second of his computed cumulative curves, however, is virtually identical to that of the first (Whalen, 1980, p. 779; 1981, p. 10). The 60-m standards (that is, the presumably near error-free 60-m datums) used in developing the plot given in figure 12 are the same as those given by Whalen (1981, p. 18) in his later report. Our data set differs from that used by Whalen, however, in that the September 14 record has not been recovered for his period 2 (Sept. 7–14); we have attempted to compensate for this loss by adding the September 27 record to his period 3 (Sept. 24–26). In spite of the differences between these two data sets, the cumulative refraction error (fig. 12) sums to about –54 mm—or only 11 mm off Whalen's (1980, p. 758) –65-mm value.

The most intriguing observation that emerges from an examination of figure 12 is the apparent insignificance of temperature—and, hence, $\Delta t'$ (fig. 6*A*)—as a constraining parameter in the accumulation of the URE. That is, the summed value (–54 mm) seems to depend on nothing more than the number of setups, regardless of temperature considerations and quite inconsistent

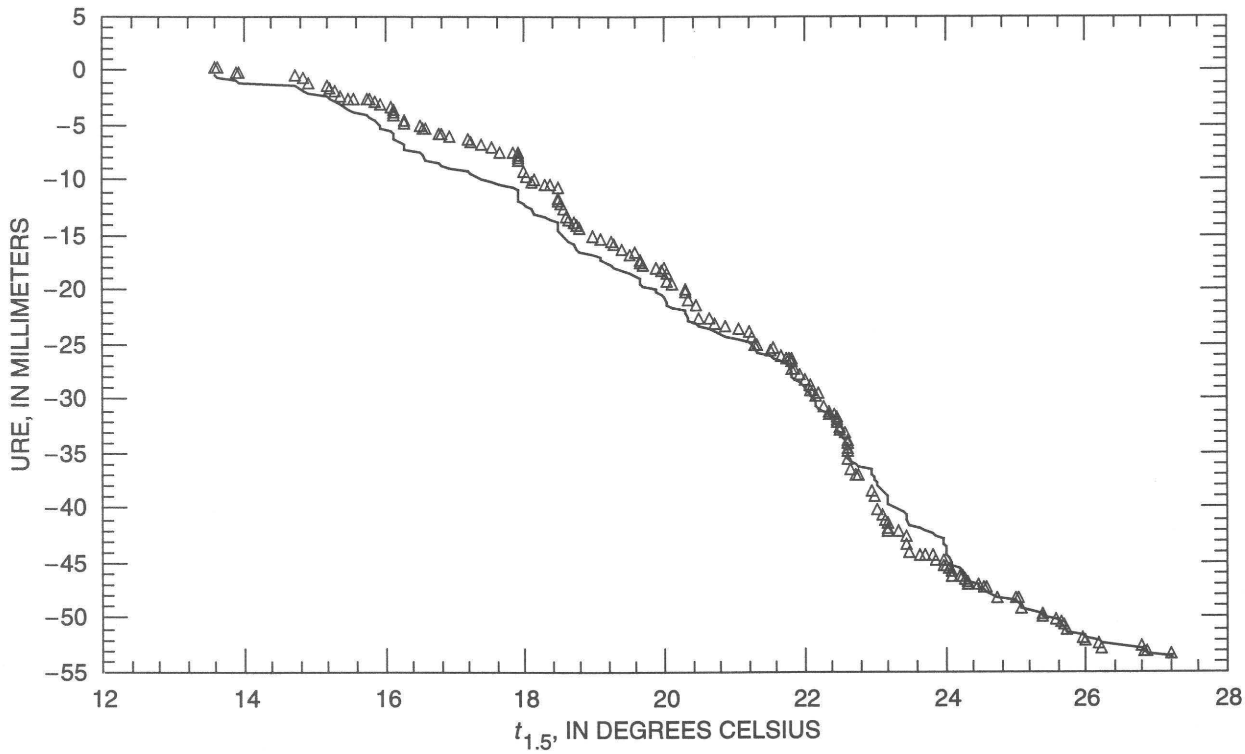


Figure 12. Cumulative URE versus setup temperature obtained from the Gaithersburg 60-m simulated sloping line measurements, where the setups are ordered according to setup temperature (triangles). Manufactured curve (solid line) based on averaged value of 0.288 mm/setup. Temperature data from rod-station measurements.

with Whalen's (1980, p. 761) expectations. In order to test this postulate we have again ordered the setups according to temperature, but have substituted the average observed minus standard difference of –0.288 mm/setup for the actually measured values. The difference between the resulting curve and the similarly constructed observed curve (fig. 12) verges on the trivial, and the differenced residuals tend toward zero. If the points that define the manufactured curve were equally spaced along the abscissa (that is, showed no temperature bias), this curve would appear as a straight line and the similarly constructed observed curve would be nearly so. While a threshold of sorts appears at about 22°C, where the observed curve steepens sharply (fig. 12), this is probably nothing more than an artifact attributable to the irregular distribution of the refraction measurements as a function of temperature. Taken together, accordingly, these observations suggest significant contamination of the refraction measurements by systematic error of uncertain origin.

Multivariate linear regression analyses, which consider measured refraction, $\Delta t'$, and curvature of the thermal gradient at each setup of the 60-m Gaithersburg line, disclose a statistically significant constant of 0.239 mm/setup for the second observation period (Sept. 7–13), but none for the third (Sept. 24–27). Subtraction of this constant from the second-period observations roughly halves the cumulative refraction associated with the simulated 60-m Gaithersburg line (fig. 12). However, because the regression equation explains very little of the variance (R^2 = 13.2 percent), it is hardly surprising that a plot of the cumulative refraction values minus this constant, where the setups are again ordered by temperature, offers little more hint of sensitivity to temperature than can be found in figure 12.

A comparison of the cumulative URE against $\Delta t'$ for the 60-m simulated line, where the setups are ordered according to $\Delta t'$ (fig. 13), indicates about 5 to 6 mm of cumulative URE with $\Delta t'$ values of 0.00°C and below. This obvious inconsistency with theory (equation 1) is difficult to dismiss as the product of inaccurate temperature measurements, since this accumulation occurs over a negative $\Delta t'$ range of about 1.5°C. Accordingly, this comparison is again clearly suggestive of a large systematic error in either the Gaithersburg standard or the subsequent leveling observations.

If we consider in turn the measured refraction as a function of *CURV*, the likelihood of a large systematic error in the Gaithersburg 60-m measurements is further enhanced. A scatter plot of the observed-minus-standard measurements by setup against corresponding values for *CURV* (fig. 14) suggests that the curvature in the thermal gradient played little role in controlling cumulative refraction. That is, we should have expected to see a sharply biased distribution of these points toward positive values for *CURV* if the URE accumulated to

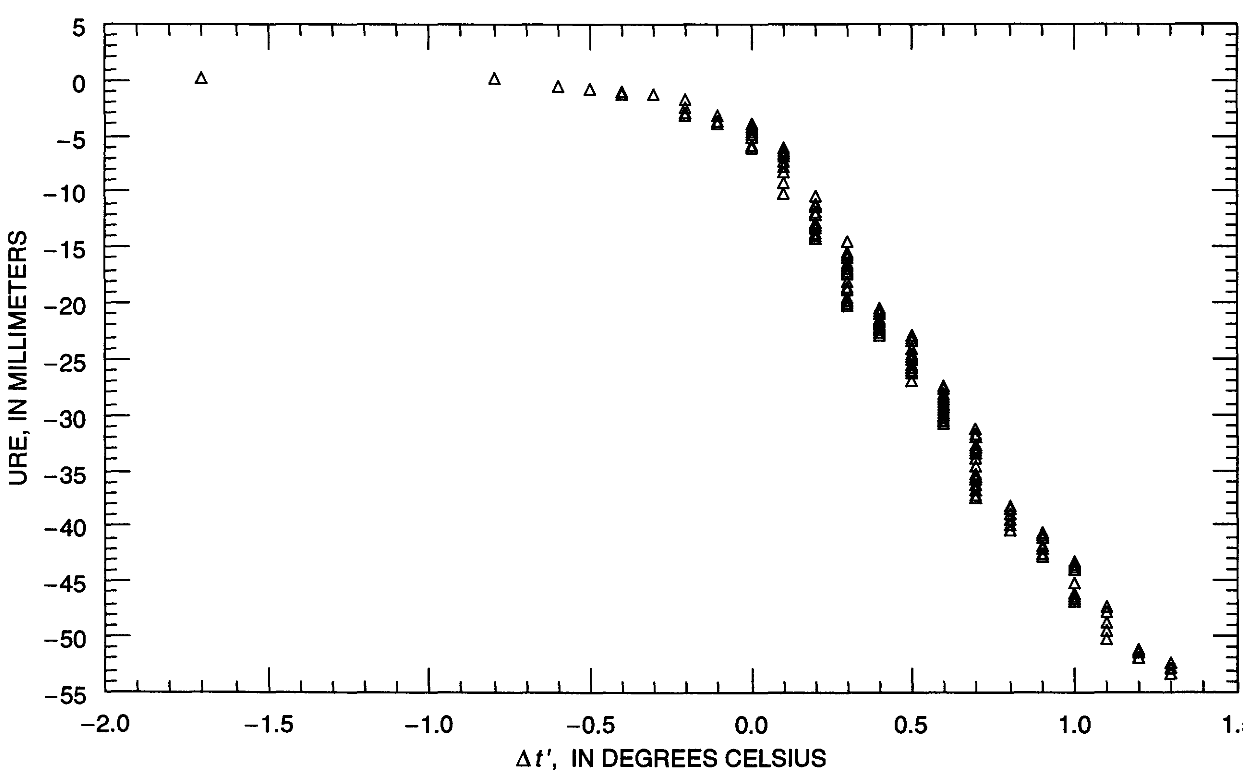


Figure 13. Cumulative URE versus setup $\Delta t'$ obtained from the Gaithersburg 60-m simulated sloping line measurements, where the setups are ordered according to setup values for $\Delta t'$. Temperature data from rod-station measurements.

any significant degree—even if the uncertainty in the temperature measurements were significantly greater than the 0.1°C to which the temperatures were read. Accordingly, it is especially likely that the Gaithersburg 60-m line was contaminated by very little URE and (or) that any temperature threshold was well above 26°C (fig. 12). This conclusion, of course, is consistent with what could be inferred directly from the comparisons of *CURV* against either $t_{1.5}$ or $\Delta t'$ (fig. 6*B*, *C*)—specifically, the random distribution about zero for the derivative values of *CURV* indicates that we should not have anticipated any significant URE in the Gaithersburg observations. Because the values for *CURV* (figs. 6, 14) were derived from the same set of temperature measurements, this predicted expectation for virtually zero URE in the Gaithersburg measurements cannot be viewed as independent evidence of the presence of systematic error in the leveling observations. Nevertheless, it is difficult to dismiss the suggestion that unmodeled environmental factors may strongly influence both the curvature in the thermal gradient and any proclivity toward the accumulation of URE.

Even though our examination of the Gaithersburg test results supports the view that locale may impose a significant limitation on the generation of any URE, this same examination also suggests probable deficiencies in the worst-case representations shown in figures 10 and 11. For example, our multivariate linear regression analyses indicate a systematic error in the summed 60-m cumulative URE which would change the normalized value of −83.5 mm (table 1) to −43.4 mm. Because the change in standards for the 30-m line was only 0.01 mm between periods 2 (Sept. 7–14) and 3 (Sept. 24–26) (table 2) and because the 30-m summed refraction value was very small in any case (−9.1 mm; table 1), we assume that it was relatively free of any systematic error—other, of course, than URE. However, the cumulative refraction error associated with the simulated 50-m Gaithersburg line is suspect because (1) the setup observations were associated with 24 unexplained errors of ~25 mm; (2) the combined index errors for the 50-m sloping-line rods summed to the unusually large value of more than 0.1 mm; and (3) two of the three cumulative curves obtained from the Gaithersburg tests flatten markedly with the change in standards between periods 2 and 3, whereas that for the 50-m cumulative line flattens sharply before the change in standards and not at all following this change (Whalen, 1981, p. 9–10). Because of all these measurement uncertainties in the 50-m line, we have not even attempted regression analyses analogous to those used in connection with the 60-m data. Accordingly, in computing the change in the cumulative normalized value of −43.2 mm for the 50-m line (table 1), we have simply taken the apparent error in the 60-m line and prorated a percentage change based on the 0.17-mm change in the standards at the 60-m marks and the 0.11-mm change at the 50-m marks between the

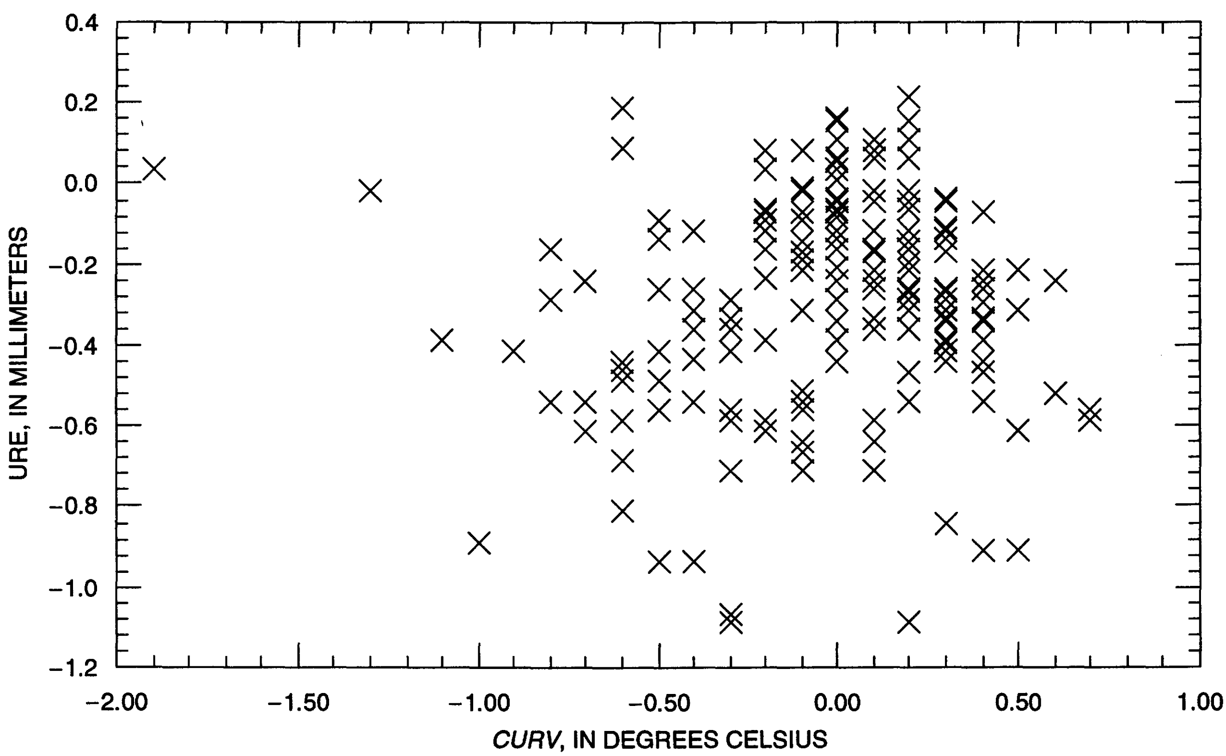


Figure 14. Scatter plot of URE versus *CURV* obtained from the Gaithersburg 60-m simulated sloping line measurements. Temperature data from rod-station measurements.

periods Sept. 7–14 and Sept. 24–26 (table 2), a procedure that raises the 50-m URE to –29.8 mm. Substituting these two newly determined normalized refraction (NR) values (–43.4 and –29.8 mm) for those given in table 1 results in two new regression equations:

$$\mathrm{NR} = 7.2 - 585\sigma_{\mathrm{DDH}}^{2}, \tag{3}$$

$$\mathrm{NR} = -6.1 - 1200\sigma_{\mathrm{DDH}}^{3}. \tag{4}$$

Although the first of these two equations is of the same form as equation (2), it gives a poorer fit to the data than equation (4)—R^2 values of 75.6% and 83.7% for equations (3) and (4), respectively. The first of these curves falls below that defined by equation (2) and tends to diminish the magnitude of the URE, whereas the second equation more closely matches that defined by equation (2) (fig. 15). Thus the upper limit estimates described by the curves given in figures 10 and 11 probably are about as accurate as can be obtained from the existing experimental data.

The distribution of the refraction error as a function of the curvature in the thermal gradient associated with the Gaithersburg 60-m line (fig. 14), which suggests that the Gaithersburg observations were devoid of any cumulative URE, invites yet another approach for obtaining a clearly locale-dependent worst-case URE curve. Regression through the three Tucson points (table 1) produces a curve that slightly increases the worst-case assessment associated with leveling of any specified order:

$$\mathrm{NR} = -3.19 - 675\sigma_{\mathrm{DDH}}^{2}. \tag{5}$$

This curve, while limited in its applicability to specifically arid regimes, may represent the worst-case curve that could be reasonably anticipated anywhere (fig. 15). Regrettably, the loss of the Tucson data precludes any attempt at determining a threshold temperature in this environment. Based on the Tucson temperatures for those days during which the standards were established (tables 2, 8), it is likely that the test temperatures were generally above any postulated temperature threshold, such that its presence would have been difficult to detect—the inverse of the Gaithersburg problem.

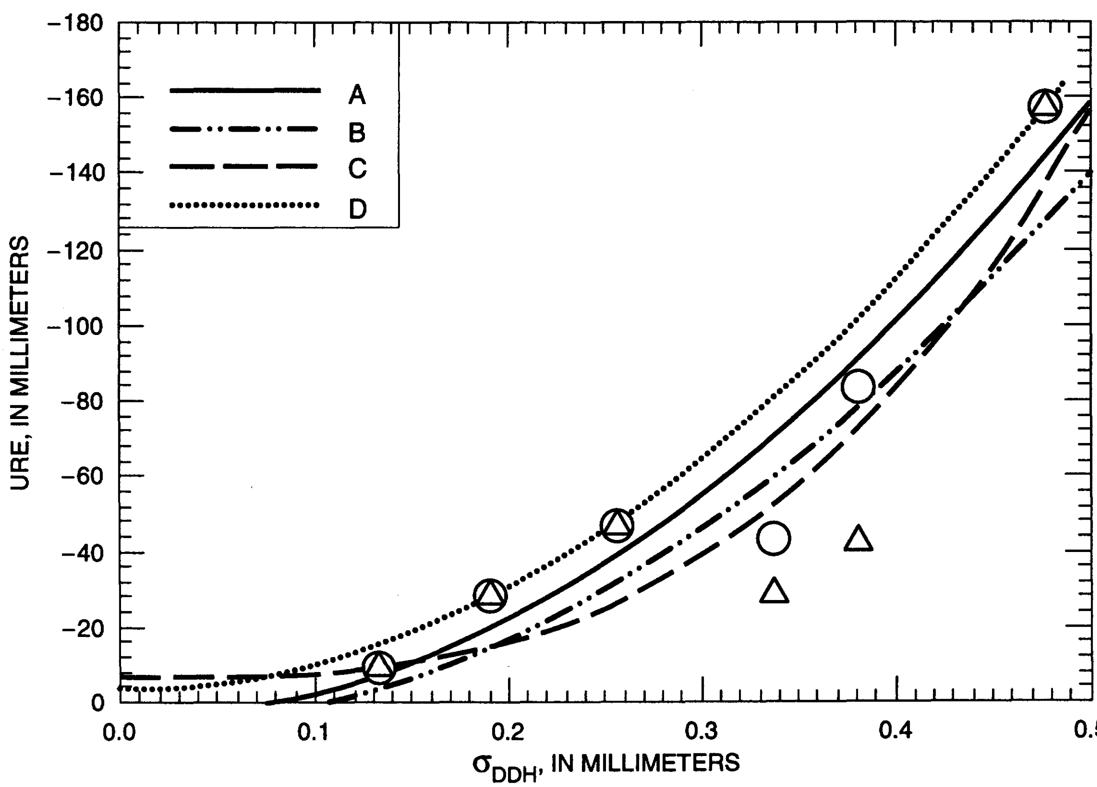


Figure 15. Worst-case curves obtained from normalized refraction on σ_{DDH}. (A) Original curve (equation 2). (B) Revised curve (equation 3) of the same order as equation (2); based on corrected, normalized refraction values for 50- and 60-m Gaithersburg lines. (C) Revised curve (equation 4) based on corrected normalized refraction values for 50- and 60-m lines with better fit to data than equation (3) (D) Curve based on fit to three Tucson points only (equation 5). Circles represent points used in calculation of original regression curve (equation 2); triangles represent points used in calculation of revised and Tucson-only regression curves (equations 3, 4, 5).

Table 8. Temperatures at Tucson for April 1980 days associated with the establishment of the refraction test standards.[a]

Date	Temperatures (°F)[b]		
	Maximum	Minimum	0800 hours
8	83	47	57
9	86	49	57
10	85	52	60
11	80	53	63
14	78	41	53
15	86	48	58
17	90	54	66
18	93	66	74
21	88	65	75
23	62	44	51

[a]Standards dates from Whalen (1981, p. 20).
[b]Temperatures from National Climatic Center, Ashville, North Carolina.

THE SAUGUS-PALMDALE TEST RESULTS

The Saugus-Palmdale experiment produced results generally compatible with those obtained in connection with the Gaithersburg-Tucson tests. However, unlike the Gaithersburg-Tucson tests, the Saugus-Palmdale experiment compared the divergence between two presumably refraction-contaminated levelings, as opposed to measuring the absolute magnitude of the URE. Accordingly, the value of this experiment rests in large measure on the extent to which we are able to assess the URE content in each of the two surveys that is independent of theoretical modeling. Several comparative procedures indicate that this approach is not only feasible, but clearly preferable to any model-dependent technique.

In considering the results of the Saugus-Palmdale field test, we have in all cases invoked the "absolute topography" concept of Stein and others (1986, p. 9034). This procedure reverses in sense the relatively few down-to-the-east (toward Palmdale) sections disclosed in the actual topography (fig. 3), thereby achieving not only the maximum possible elevation difference (704.5 m) but the maximum possible LSL-SSL divergence (−44.2 mm based on the field data and −50.1 mm based on the corrected data) that can be obtained from this experiment.

DIVERGENCE AS A FUNCTION OF DDH

If the Saugus-Palmdale experimental sections are ordered according to the maximum DDH per section for the LSL survey, the cumulative LSL-SSL divergence hardly begins to accumulate until the maximum DDH per section reaches about 0.50 mm (fig. 16*A*)—or about 0.45 mm based on the suspect corrected data (fig. 16*B*). Mark and others (1987, p. 2785) considered two possible explanations for the apparent cutoff at or near 0.50 mm, both of which implicitly accept the view that the divergence is attributable entirely or nearly entirely to the URE—a position clearly consistent with that of Stein and others (1986, p. 9042): (1) Measurably significant URE in the SSL survey matched that in the LSL survey, thereby challenging the generally accepted theoretical relation between refraction and the squared dependence on sight length. (2) Alternatively, and far more likely, the systematic component of the atmospheric refraction error was largely eliminated in the survey subset composed of those sections in which the DDH for the LSL survey was kept at or below 0.50 mm. Thus, under those circumstances in which the setup-rejection limit was held to ±0.50 mm (roughly comparable with σ_{DDH} values of ~0.17–0.19 mm, marginally within the second-order field—table 6, fig. 11), one might conclude that the URE could not even begin to accumulate. That this may not necessarily be the case is shown, for example, by a worst-case, 16-mm estimated URE for first-order leveling over the 491.4-m normalized height difference used in the construction of figure 11. Accordingly, while the σ_{DDH} values may provide useful limits on worst-case errors identified with stipulated orders of leveling, they do not by themselves provide estimates of the *probable* magnitude of the URE associated with a specified leveling—as could be reasonably inferred from the close correspondence between the smaller (less than 0.3 mm) σ_{DDH} values associated with both the refraction-contaminated, sloping-line surveys and the zero-slope surveys obtained from the Gaithersburg-Tucson tests (fig. 10).

Nevertheless, and focusing specifically on the Saugus-Palmdale experiment, we can break this data set into two groups based on the maximum DDH per LSL section, where this grouping is independent of either temperature or any temperature-derivative parameter. Our choice of the limiting DDH is governed by the break at about 0.50 mm disclosed in figure 16*A*. If for each group we plot the cumulative LSL-SSL divergence against distance, we find that the subset in which the DDH nowhere exceeded 0.50 mm per setup is associated with a cumulative divergence of no more than 6 to 7 mm—or about 10 mm based on the corrected data—whereas that based on those sections in which the maximum DDH exceeded 0.50 mm in one or more setups was associated with nearly all of the divergence (about 40 mm) accumulated over the full length of the Saugus-Palmdale line (fig. 17). Significantly, the data set in which the DDH ≤ 0.50 mm for all setups is also identified with an average LSL sectional temperature of 20.0°C (fig. 17), whereas the less tightly controlled data set is identified with an average LSL sectional temperature of 28.0°C (fig. 17). Equally germane, where the DDH ≤ 0.50 mm, the ratio of the square of the average LSL sight length to that of the average SSL sight length is 3.23, whereas this same ratio in which the DDH > 0.50 mm is 3.08 (fig. 17)—nearly

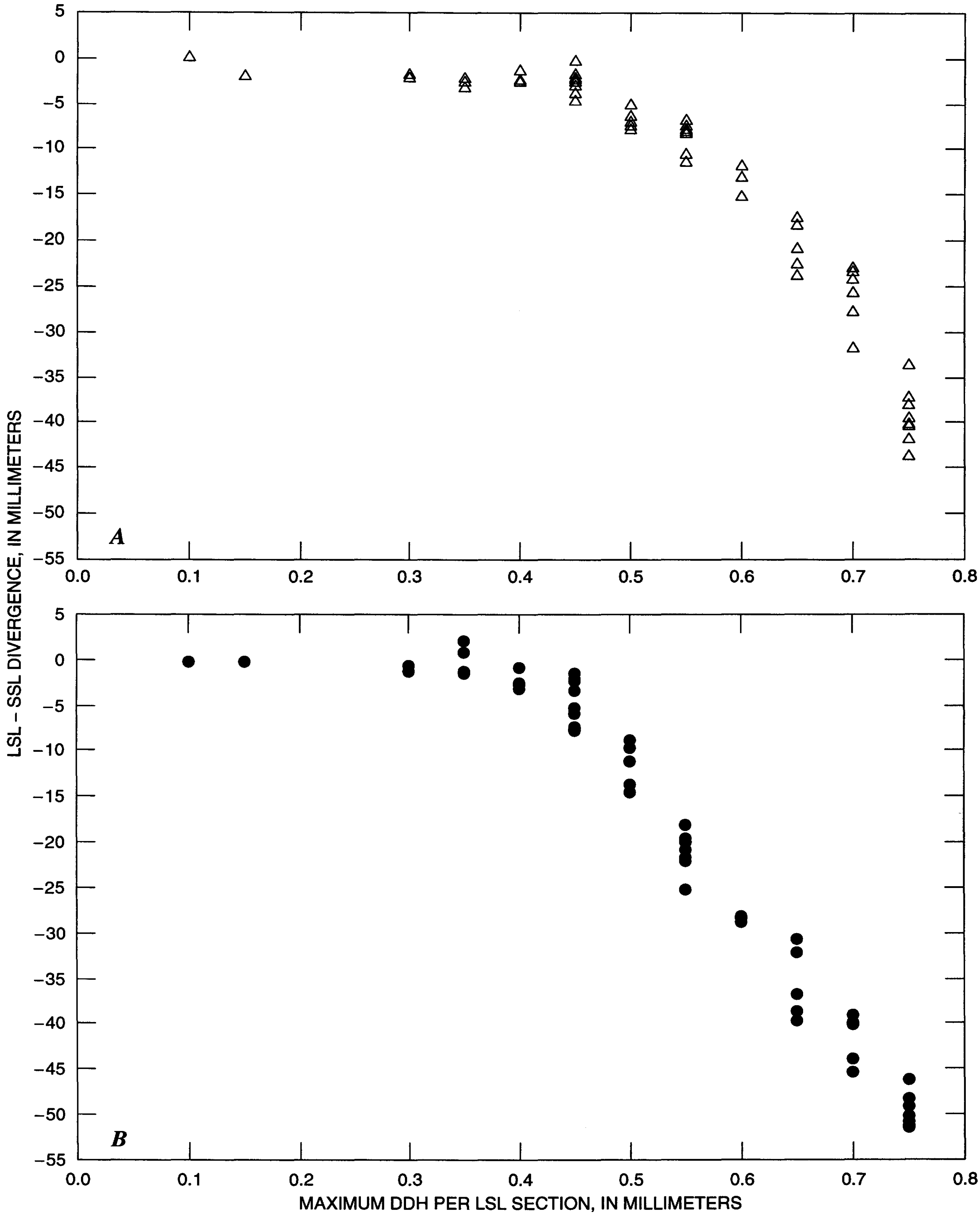


Figure 16. Cumulative LSL-SSL divergence obtained from the Saugus-Palmdale field test, where the sections are ordered according to the maximum DDH per LSL section and sectional slopes are so fixed that direction of running is in all cases uphill—the "absolute topography" concept of Stein and others (1986, p. 9034). *A*, Based on field data. *B*, Based on corrected data.

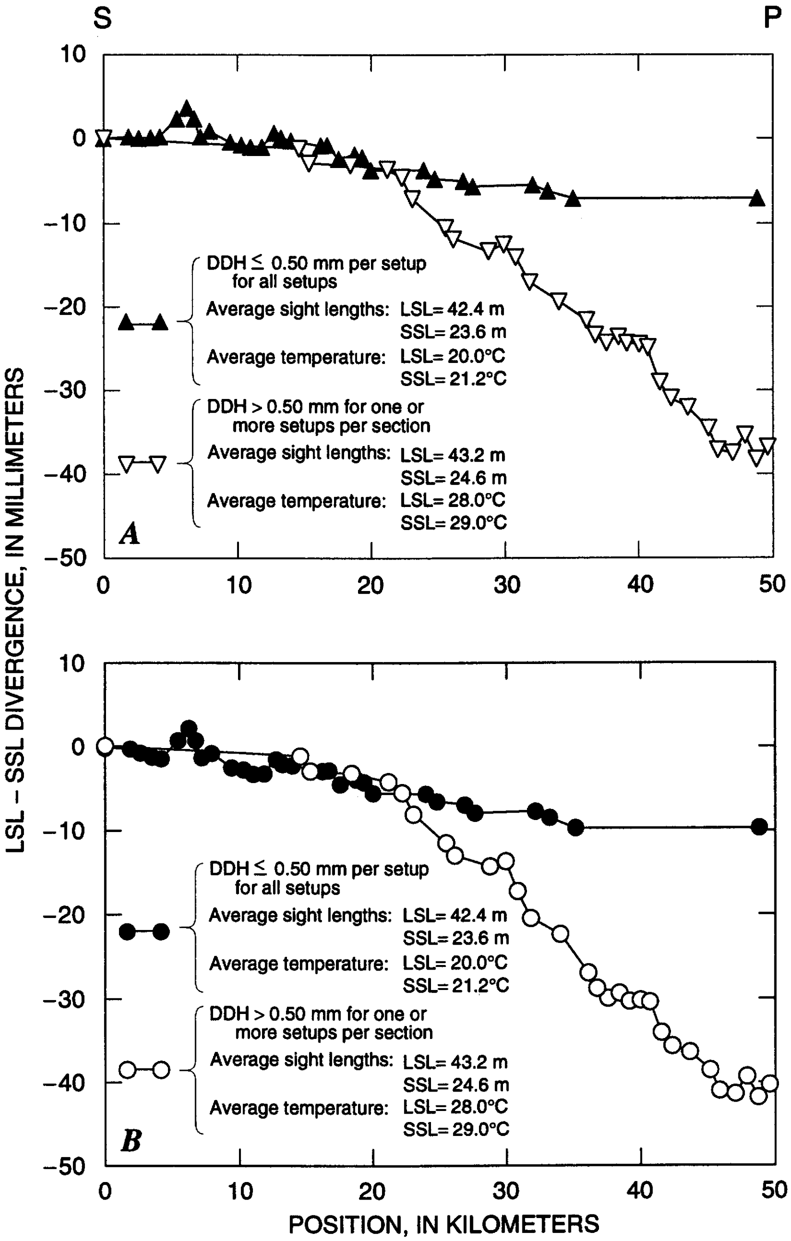


Figure 17. Cumulative divergence between LSL and SSL Saugus(S)-Palmdale(P) surveys as a function of position along line, where the experiment has been reformulated as two separate experiments: one composed of sections in which DDH ≤0.50 mm/setup for all setups, and the other composed of sections in which DDH >0.50 mm/setup for one or more setups. Sectional slopes are so fixed that the direction of running is in all cases uphill (see fig. 16). *A*, Based on field data. *B*, Based on corrected data.

equal values, but of the opposite sense to what one might expect if sight length is the overwhelming determinant.

URE CONTENT OF THE SSL SURVEY

If we accept the measured divergence between the LSL and SSL surveys obtained from these two experiments as an accurate representation of the cumulative difference in the URE, and if we can independently determine the cumulative URE in either the LSL or SSL survey, the other may be obtained by difference.

The preceding descriptions of the relations between DDH, divergence, and sight length constitute in themselves presumptive evidence that the SSL survey in the Saugus-Palmdale experiment was contaminated by a very modest cumulative URE. Because the subset consisting of those sections in which the LSL DDH was $\leq$ 0.50 mm was associated in turn with near-trivial divergence (fig. 17)—a millimeter or two above the 1σ random error estimate for this manufactured line based on the field data (Stein and others, 1986, p. 9039)—and because the stipulated setup-rejection limit of ±0.30 mm for the SSL survey was exceeded in only 6 out of 60 sections and nowhere exceeded ±0.45 mm over the entire line (table 6), the 0.50-mm (or even the 0.45-mm) cutoff (fig. 16) argues that a setup-rejection criterion as great as ±0.45 mm, but generally no greater than ±0.30 mm, constrained the URE accumulation in the Saugus-Palmdale SSL survey to no more than a few millimeters. This position is reinforced by the observation that *none* of those sections in which the SSL setup-rejection exceeded the prescribed ±0.30-mm limit fall within the subset in which the LSL DDH was >0.50 mm, such that nearly all of the LSL-SSL divergence accumulated within those sections in which the SSL setup-rejection limit was held to ±0.30 mm (fig. 17). It is this 0.30-mm limit associated with this particular survey that produced a σ_{DDH} value of 0.080 mm (table 6)—for a *worst-case* cumulative URE very close to zero (figs. 10, 11). Alternatively, the SSL URE can be deduced from the worst-case URE computed by means of equation (2) and the σ_{DDH} given for the entire line—0.139 mm (table 6)—a procedure that produces a worst-case URE of −11.6 mm based on an absolute elevation difference of 704.5 m (or −23.3 mm based on equation 5). However, because the setup-elevation difference over the SSL line averaged about 0.68 m (based on an absolute elevation difference of 704.5 m and 1033 setups—table 6), as contrasted with the 2.0-m (or 1.95-m) difference on which the worst-case estimate is based, these computed values can be expected to overestimate the URE—probably significantly—if the URE is not a linear function of Δh .

The various estimates of the cumulative URE associated with the SSL survey for the Saugus-Palmdale field test indicate that it was measurably insignificant. Accordingly, we conclude that the measured LSL-SSL divergence represents a very tight determination of the cumulative URE identified solely with the LSL survey.

The preceding conclusion conflicts with that of Stein and others (1986, p. 9033), whose modeled value for the URE assigned to the Saugus-Palmdale SSL URE is −27 mm. At least three arguments in addition to those cited above indicate that this modeled value overestimates the actual error. (1) The results of the Gaithersburg-Tucson tests indicate that the modeled corrections for the shorter sight-length observations both over- and under-correct the measured URE as a percentage of the error, commonly significantly (Whalen, 1981). (2) Stein and others (1986, p. 9033) assigned a modeled correction of −79 mm to the LSL survey and average sight lengths of 42 and 22 m to the LSL and SSL surveys, respectively, for the Saugus-Palmdale experiment.[3] Thus, owing to the theoretical relation between refraction error and the square of the sight length, the SSL refraction error should be of the order of 27 percent of the still very large −79-mm value—or roughly −21 mm—as contrasted with −27 mm. (3) Provided that the DDH is held to ±0.50 mm (or even ±0.55 mm), the NGS refraction models grossly overcorrect for any URE, whereas above this DDH cutoff these same models tend to undercorrect (Castle and others, 1983). At the very least, accordingly, the uncertainties in the modeled corrections implied by these observations indicate that the empirically estimated values for the SSL URE probably are more soundly based than the modeled values.

URE VERSUS $t_{1.5}$, $\Delta t'$, AND *CURV*

A comparison of the cumulative divergence along the Saugus-Palmdale line (effectively, the LSL URE) against the LSL sectional temperatures, where the sections are ordered according to temperature, indicates little accumulation of URE below a threshold value of about 23.8°C—or 21.9°C based on the corrected data (fig. 18). This determination is obviously at variance with the comparable comparison based on the Gaithersburg 60-m simulated line data (fig. 12) and, in this sense, suggests a conceptual contradiction. Nevertheless, and as we have already indicated, the cumulative URE in figure 12 probably is attributable largely to systematic error in the comparative measurements, such that any threshold temperature at the Gaithersburg site could easily have been above the higher observed temperatures—26°C or greater (fig. 6). We cannot, however, appeal to measurement error in attempting to explain the phenomenon disclosed in figure 18. That is, measurements of temperature to within 0.5°C are easily achieved and no more than a modest fraction of the LSL-SSL divergence can be attributed to measurement error other than that related to refraction, much as argued by Stein and others (1986, p. 9033–9034, 9038–9039). While we should expect to see increasing $\Delta t'$ values associated with increasing temperatures, there is no evident theoretical reason why the accumulation of URE should be so explicitly dependent on temperature,

[3]Although Stein and others (1986, p. 9033) report a mean LSL sight length of 46 m in their table 1, this figure apparently is in error, since it is at variance with the 42-m figure given in their abstract (Stein and others 1986, p. 9031), the 42-m value given by the former Director of the National Geodetic Survey, and our own, independent determination (fig. 3*B*).

nor, indeed, is there any indication of any threshold relation between $\Delta t'$ and $t_{1.5}$ developed from the Saugus-Palmdale measurements (fig. 9A). Accordingly, we are forced to the conclusion that temperature proxies for the upward sensible heat flux and, hence, $\Delta t'$ and *CURV* as well.

A comparison of the cumulative divergence along the Saugus-Palmdale line against the LSL sectional means for $\Delta t'$ (fig. 19) indicates little or no URE accumulation below a $\Delta t'$ value of at least 1.1°C—or 1.0°C based on the corrected data. Unlike the preceding comparison (fig. 18), however, a greater likelihood of significantly distorting error exists in these temperature measurements, such that the actual threshold value for $\Delta t'$ may be somewhat below (or above) that shown in figure 19. Thus, while we would expect to see the divergence begin to accumulate (or a reversal in the sense of this accumulation) at $\Delta t'$ values close to zero, this does not seem to be the case. The implied corollary, moreover, is that *CURV* should be close to zero for $\Delta t'$ values of up to about 1.0°C, which again contradicts our theoretical expectations.

A plot of the cumulative LSL-SSL divergence versus the averaged sectional values for *CURV* (fig. 20) discloses a threshold value for *CURV* at about 0.3°C, below which the accumulation of any URE is negligible. This determination is again inconsistent with theory, which would argue that this threshold value should be close to zero. However, this inconsistency, as well as that associated with the 1.1°C threshold value for $\Delta t'$ (fig. 19), is largely one of appearance rather than substance. Specifically, *CURV* has been calculated from measurements over the full 2.0-m range between 0.5 and 2.5 m, whereas the average LSL setup elevation difference (Δh) was less than 1.2 m (1.194 m—704.5 m/590 setups). Accordingly, because the curvature in the thermal gradient increases progressively toward the ground surface (Shaw and Smietana, 1983, p. 10,488), at or below measured values for *CURV* of 0.3°C the curvature over that part of the thermal gradient actually sampled by the leveling observations probably was indistinguishable from linear. Had the average LSL sighting embraced the entire 2.0-m permissible setup-elevation difference, the threshold

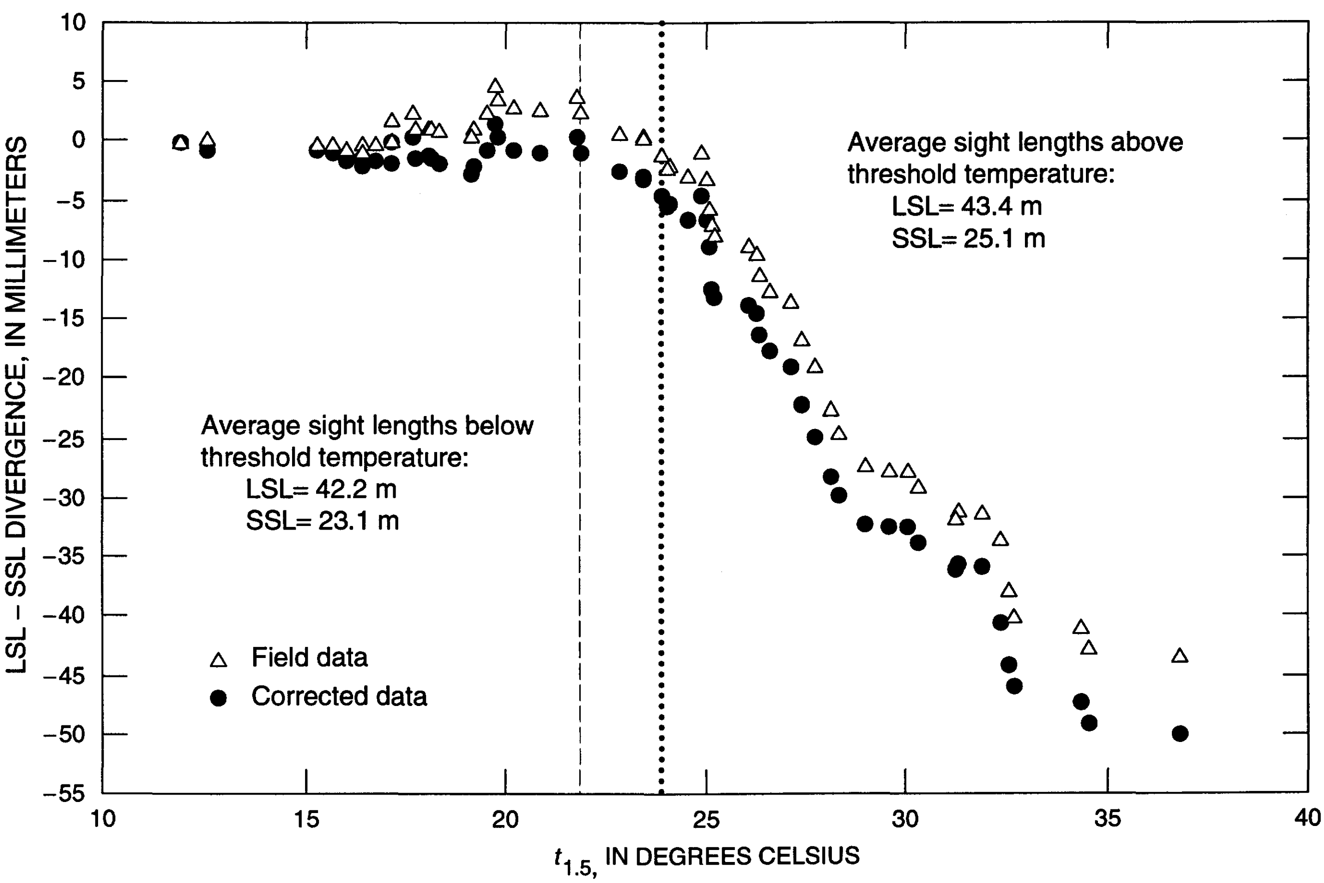


Figure 18. Cumulative LSL-SSL divergence versus LSL sectional temperatures obtained from the Saugus-Palmdale field test, where sections are ordered according to LSL sectional temperature. The threshold temperature, below which divergence does not accumulate, is 23.8°C based on field data (dotted line); that based on corrected data is 21.9°C (dashed line). Average sight lengths calculated from field data. Sectional slopes are so fixed that direction of running is in all cases uphill (see fig. 16).

value for *CURV* probably would have been much closer to zero—and the LSL-SSL divergence would have been much greater. In practice this situation probably would never arise, because the observer could be expected to position the instrument in such a way that he could be sure that the upslope sight would be well above the minimum permitted rod reading—thereby avoiding any likelihood that he would be forced to break and then reestablish the setup.

A joint examination of the three comparisons considered here (figs. 18–20) invites several observations. Were we to "correct" the values for $\Delta t'$ and *CURV* in order to accommodate the 0.1°C error in the 0.5-m sensor—based on the post-survey calibration—all of the values for both $\Delta t'$ and *CURV* shown in figures 19 and 20 would be increased by 0.1°C. In other words, the threshold value for $\Delta t'$ would increase to about 1.2°C and that for *CURV* would jump to a bit over 0.4°C. Thus, one of the important messages derivative from the results of the Saugus-Palmdale experiment seems to be this: In any routinely produced leveling (Δh <<2.0 m), it is likely that both $\Delta t'$ and *CURV* must be well above zero before URE can begin to accumulate. A second and probably more significant observation emerges from the identification of those sections over which the LSL-SSL divergence actually accumulated. Specifically, if we consider just the field data, this divergence accumulated over 31 sections above the threshold temperature (fig. 18), over 31 sections above the threshold value for $\Delta t'$ (fig. 19), and over 29 sections above the threshold value for *CURV* (fig. 20). If we arbitrarily select the divergence above the threshold value for temperature as a basis for comparison, we find that 25 of those sections above the threshold value for $\Delta t'$ and 24 of those above the threshold value for *CURV* are common to the 31 sections above the threshold temperature.[4] Two fairly important conclusions can be drawn from this correspondence: (1) Temperature does a less than perfect, but still quite a good job of proxying for both $\Delta t'$ and *CURV*, and

[4]The same set of comparisons based on the corrected data indicate that the divergence accumulated over 35 sections above the threshold temperature, over 36 sections above the threshold value for $\Delta t'$, and over 30 sections above the threshold value for *CURV*. Again, selecting the threshold value for temperature as a basis for comparison, 25 sections above the threshold value for $\Delta t'$ and 24 sections above the threshold value for *CURV* are common to the 35 sections above the threshold temperature.

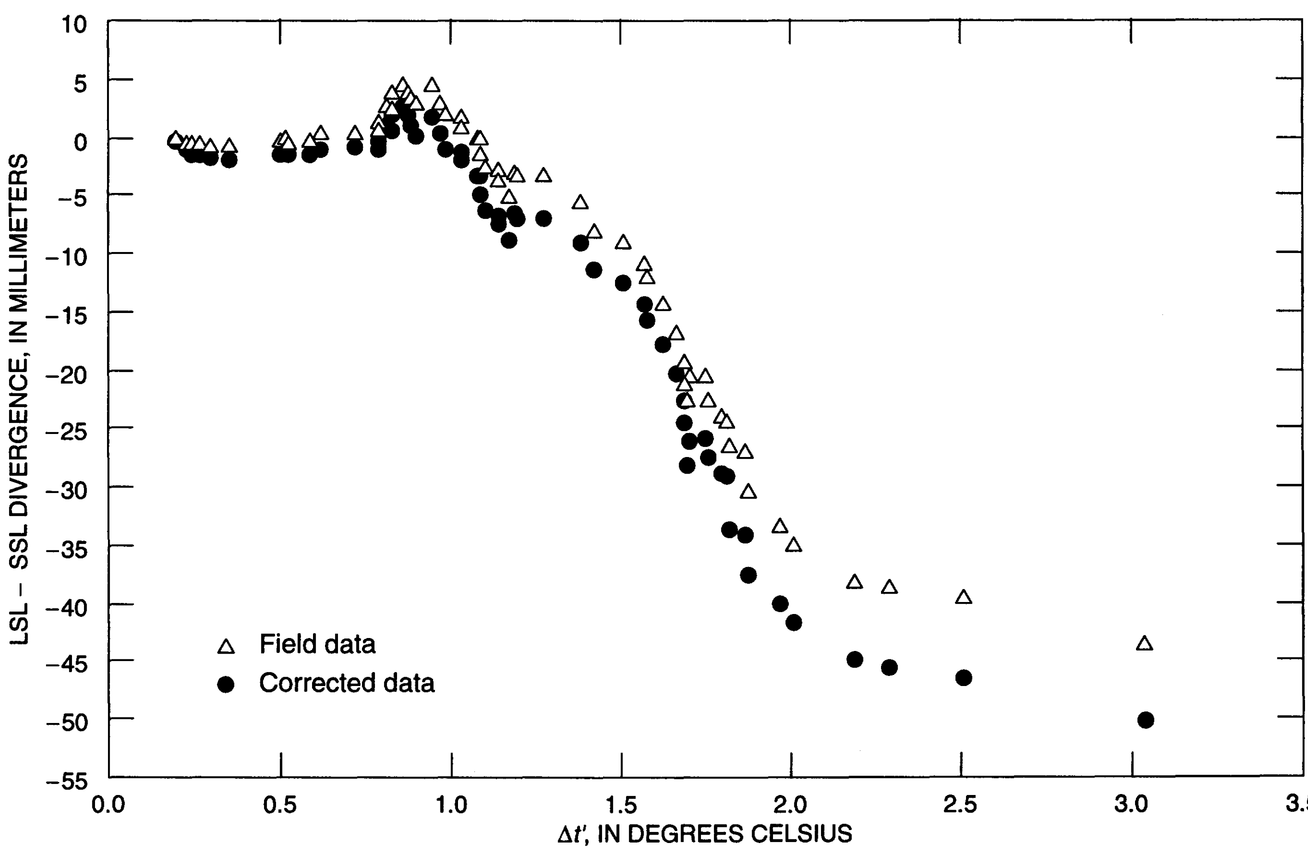


Figure 19. Cumulative LSL-SSL divergence versus LSL sectional $\Delta t'$ values obtained from the Saugus-Palmdale field test, where sections are ordered according to LSL sectional $\Delta t'$ values. Sectional slopes are so fixed that direction of running is in all cases uphill (see fig. 16).

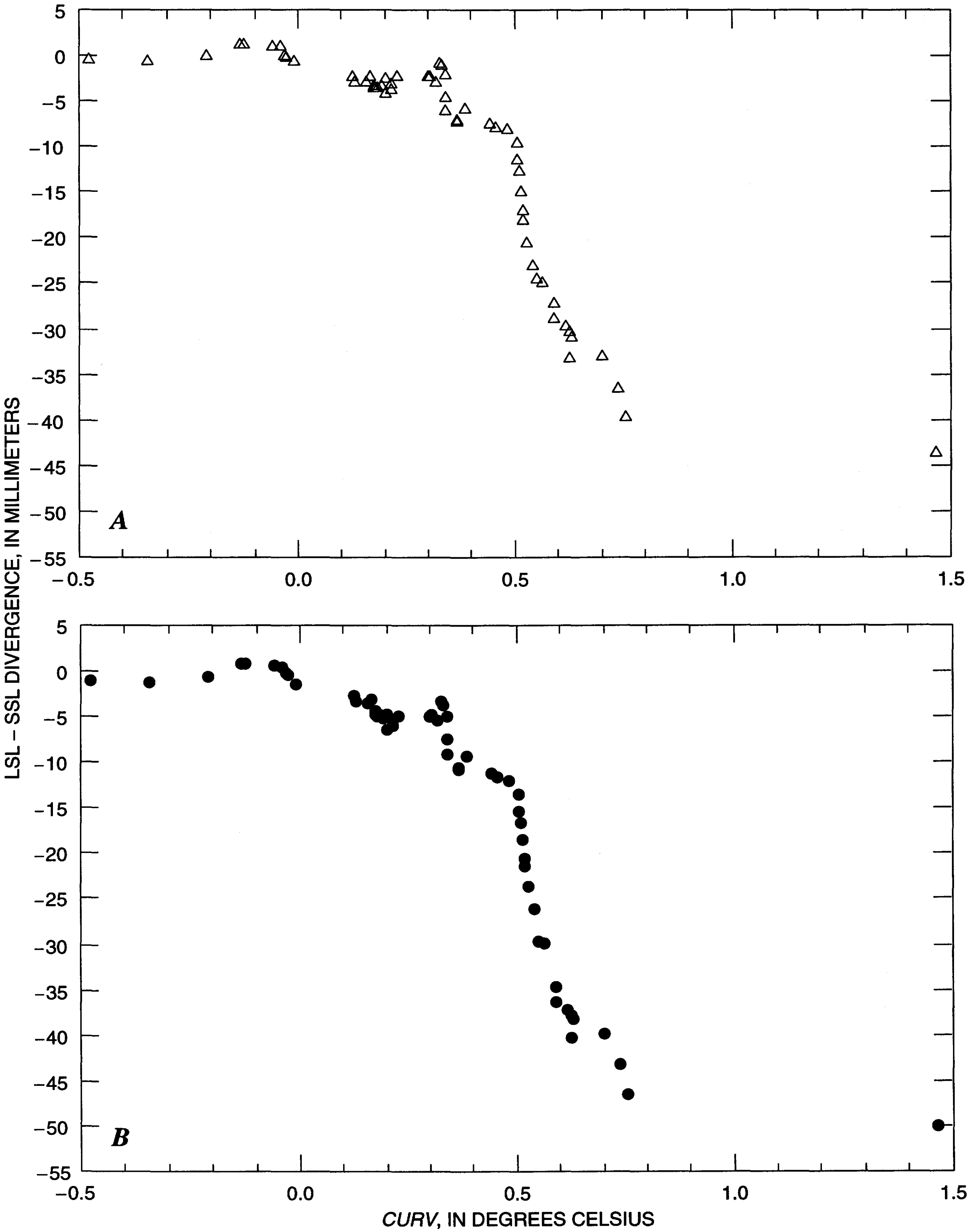


Figure 20. Cumulative LSL-SSL divergence versus LSL sectional values for *CURV* obtained from the Saugus-Palmdale field test, where sections are ordered according to LSL sectional values for *CURV*. Sectional slopes are so fixed that direction of running is in all cases uphill (see fig. 16). *A*, Based on field data. *B*, Based on corrected data.

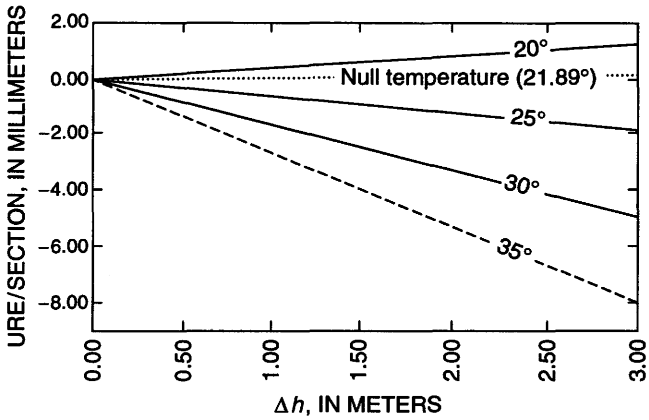
(2) the temperature measurements made in connection with the Saugus-Palmdale field test were acceptably accurate.

REGRESSION ANALYSIS OF URE ON t AND Δh

While placing upper limits on the URE as a function of survey order is clearly feasible (fig. 11), identification of those criteria that could contribute to improved estimates of the *likely* magnitude of the URE for any specific leveling is a good deal more challenging if only because neither of the two defining parameters (σ_{DDH} and URE—see fig. 10) is normally measured. Nevertheless, owing to the nature of both the URE and σ_{DDH}, several procedures should, in theory, permit better estimates of the URE than can be obtained directly from figure 11. When leveling over sloping terrain, the URE varies with both L^2 and Δh (equation 1)—as, indeed, L^2 may vary inversely with Δh owing to the limitation that increasing slope puts on sight length. Accordingly, if one or the other of these two parameters can be compared with measured estimates of the URE, which may in turn be functionally related to some routinely measured parameter—for example, temperature—it may be possible to develop additional, empirically based determinations for the URE. With these points in mind, we can again consider the results of the Saugus-Palmdale field test in order to establish the relation, if any, between the average temperature and (or) averaged setup-elevation difference (Δh) per section, on the one hand, and measured values of URE on the other. This requires, of course, that we accept the provisional conclusion that the URE was confined entirely to the LSL survey. Although we could have selected L (or L^2) as the variable to be compared against the URE, L probably varied within a narrow range in the LSL survey.

Multivariate linear regression analysis of refraction (URE) on t and Δh obtained from the LSL survey produced statistically significant results (fig. 21) that are subject in part to independent verification. The regression equation obtained from our analysis of the Saugus-Palmdale experiment (fig. 21) suggests alternative procedures for improving the estimated magnitude of the URE in any procedurally constrained leveling. However, in using this explicitly empirical approach we can extrapolate beyond these results over no more than a very limited range. We can go beyond this point only by invoking the generally accepted theoretical relation between the URE and the square of the sight length—which, as we show above, may be of limited utility, if not actually misleading, when dealing with higher order, procedurally constrained levelings.

The most striking result disclosed by the analysis of the Saugus-Palmdale data is the sensitivity of the URE to t (fig. 21). Owing to the statistical insignificance of the intercept, the family of curves defined by successively higher temperatures has been constrained to pass through the origin. Thus, the null temperature (21.89°C) for the stipulated average sight length of 42 m may slightly underweight a real proclivity toward the accumulation of any URE. Nevertheless, this computed value agrees remarkably well with the threshold temperature (or temperatures) disclosed through a comparison of cumulative refraction against the LSL sections ordered according to temperature (fig. 18).



REGRESSION EQUATION IS:

URE/section = –0.738 + 4.514Δh – 0.206$t\Delta h$, where Δh and t are in meters and degrees Celsius, respectively

STATISTICS

Predictor	Coefficient	Std. Dev.	t-statistic
Constant	–0.738	0.601	–1.23
Δh	4.514	1.173	3.85
$t\Delta h$	–0.206	0.035	–5.28

R^2=38.1%

Figure 21. URE per averaged 0.83-km section versus average setup-elevation difference (Δh) as a function of temperature. Based on regression analysis of Saugus-Palmdale LSL survey, where average sight length (~42 m) is treated as a constant. URE per section is based on normalization of all sections to 10 setups/section. Because constant term is not statistically significant, all regressions are shown as passing through origin. Because only a single sectional temperature value exceeded 35°C, that curve is dashed.

CANADIAN TEST RESULTS

The results of the Canadian experiment are again more or less consistent with both the Gaithersburg-Tucson and Saugus-Palmdale field tests. They are perhaps less instructive than the other two experiments in that they were associated with only two-point temperature measurements (at 0.5 and 2.5 m), thereby precluding comparisons against *CURV*. Moreover, our examination of the Canadian data suggests an even higher level of uncertainty than that which emerged in considering the results of the other two experiments.

URE CONTENT OF THE SSL SURVEY

The cumulative URE associated with the Canadian SSL survey can be estimated by means of the same worst-case procedure adopted in estimating the probably maximum Saugus-Palmdale value. Specifically, the σ_{DDH} for the SSL survey is given by Heroux and others (1985, p. 318) as 0.17 mm, which leads to a maximum URE value over the simulated 283.4 m height difference of −8.3 mm (or −13.7 mm based on equation 5). However, because the setup-elevation difference over the SSL line averaged only about 0.39 m, these computed values probably grossly overestimate the actual URE simply because the curvature in the thermal gradient can be expected to increase sharply downward. Moreover, since this modest 0.39-m value was coupled with an average $\Delta t'$ of only 0.55°C, there is a reasonable physical basis for concluding that the curvature over the actually sampled part of the thermal gradient was insignificant and that any resulting URE was near negligible. Finally, were we to accept the computed values of −8.3 mm and −13.7 mm as realistic estimates of the SSL URE, it would require that they comprise 56 and 92 percent, respectively, of the measured, cumulative LSL-SSL divergence—certainly not impossible, but *extremely* unlikely. In any case, because the average sight lengths were about the same (22.5 m vs. 24.3 m), the average sectional temperatures were nearly identical (23.8°C), and the mean $\Delta t'$ was only about 40 percent of that measured along the Saugus-Palmdale line (Stein and others, 1986, p. 9036–9037), the cumulative SSL-URE probably was no more than a small fraction of whatever URE may have accumulated in connection with the Saugus-Palmdale SSL survey and, as such, was measurably insignificant.

CUMULATIVE URE VERSUS t AND $\Delta t'$

A plot of the cumulative divergence against the average sectional temperature derived from the Canadian field test, where the sections are again ordered by temperature (fig. 22), produces a less definitive characterization than that obtained from the Saugus-Palmdale experiment—at least in the sense that only about one-fifth of the leveled sections plot below the apparent threshold temperature of ~20°C. While not directly demonstrable, it is possible that this significantly different distribution of the data (compare figures 18 and 22)

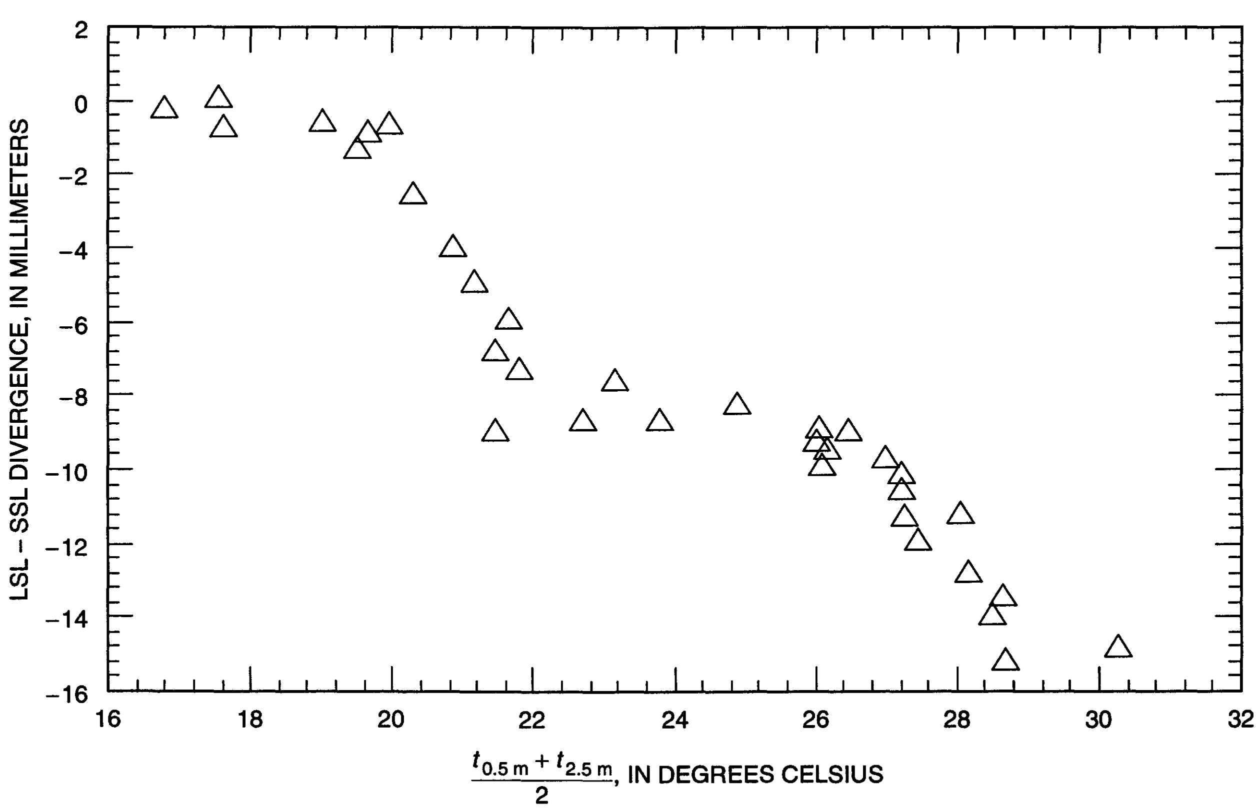


Figure 22. Cumulative LSL-SSL divergence versus average sectional temperature obtained from Canadian field test, where sections are ordered according to the average sectional temperature (Heroux and others, 1985, as modified by W. Gale, 1992, written commun.). Average sectional temperature equates with mean of 0.5-m and 2.5-m temperature-probe measurements.

is somehow related to unassessed physical differences among the Canadian leveled sections. A bias of sorts is suggested especially by the accumulation of over half (−7.78 mm) of the cumulative LSL-SSL divergence (−14.94 mm) within only one of the six measured sections—one associated with an intermediate LSL Δh of 1.1083 m, an average t of 22.47°C—or less than the line average (23.77°C)—and an average $\Delta t'$ approximately equal to the line average value (0.55°C). This particular section is unique only in its maximum LSL sight length of 48.2 m, as contrasted with a minimum of 42.2 m in the shortest of the six sections, and the largest end-to-end elevation difference, differences which ranged between 4.5 m and 13.3 m (Heroux and others, 1985, p. 315–316). It is equally likely that the indicated difference in the distribution of the data is due to the temporal distribution of the observations that comprise the Canadian field test, observations scattered over the period June 22–September 18, 1984 (Heroux and others, 1985, p. 315). Specifically, because the intent of this experiment was the production of refraction-contaminated measurements, a deliberate effort was made to select for leveling days in which the URE could be expected to accumulate (W. Gale, 1993, oral commun.), a determination that may have led to the relatively poorly defined threshold temperature. Moreover, because the computed standard error of 7.4 mm (Heroux and others, 1985, p. 317) roughly matches the divergence summed through temperatures of up to about 25°C, it is likely that the effective threshold temperature may be somewhat greater than 20°C. That is, if all of the experimental levelings had been deliberately restricted to sections in which the temperature was no less than 22°C—hardly an abnormally elevated value—we could easily have concluded that the still poorly defined threshold temperature was about 25°C (fig. 22).

Comparison of the cumulative LSL-SSL divergence against $\Delta t'$, where the sections are ordered according to increasing values of $\Delta t'$ suggests a threshold value of about 0.5°C to 0.6°C (fig. 23). While this comparison is somewhat analogous to that shown in figure 22 in terms of the nebulous threshold, the threshold value for $\Delta t'$ is even more imprecisely defined than that for temperature. Equally perplexing is the absence of any accumulation of URE over the $\Delta t'$ range 0.8°C to 1.5°C, even allowing for the small number of included sections.

THE UNCERTAINTY FACTOR

The results of the Canadian experiment are particularly enigmatic largely because they almost completely defy prediction. For example, a multivariate linear regression analysis, identical to that employed in connection with the Saugus-Palmdale data, yielded no correlations among URE, t and Δh. The absence of any

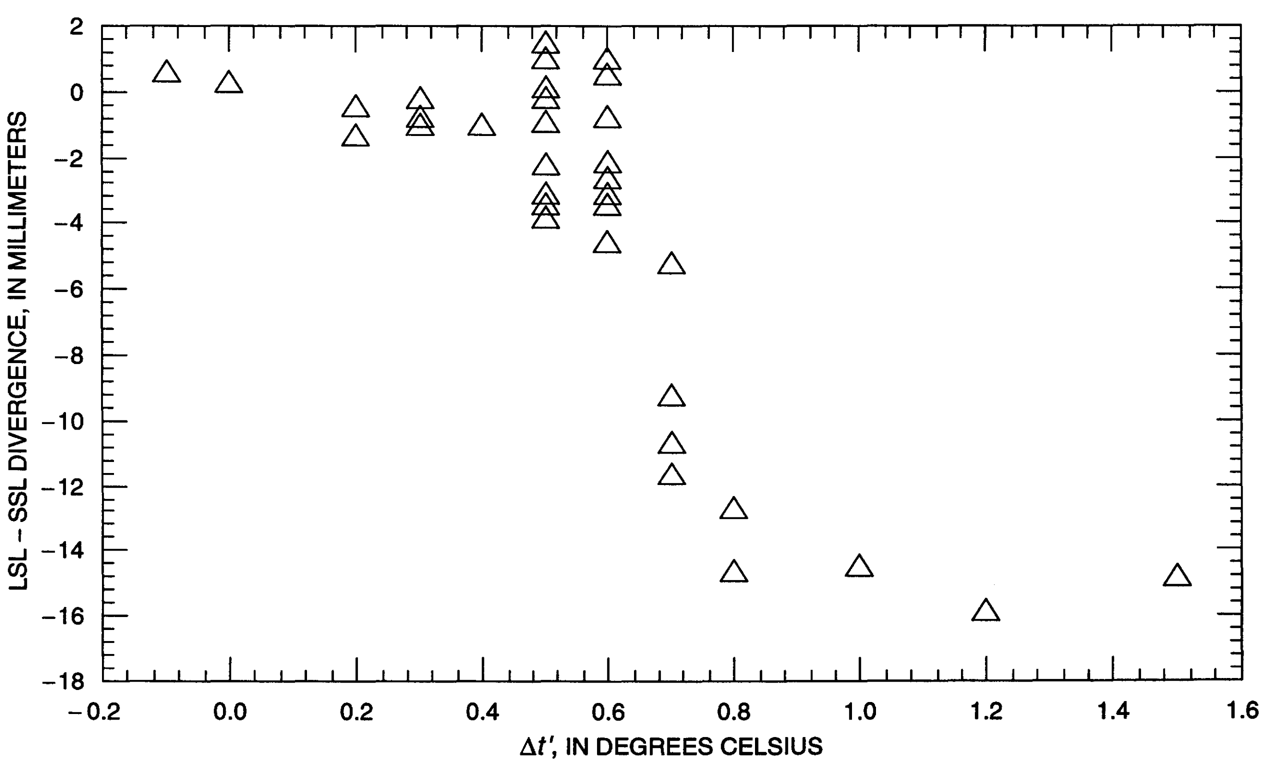


Figure 23. Cumulative LSL-SSL divergence versus sectional $\Delta t'$ values obtained from Canadian field test, where sections are ordered according to $\Delta t'$ values (Heroux and others, 1985, as modified by W. Gale, 1992, written commun.).

correlation in the Canadian data may be a function of the sectional bias described in the preceding section. This particular bias is compounded by the observation that a single running over the suspect section accounted for 27 percent (–4.05 mm) of the cumulative LSL-SSL divergence, a running associated with neither an especially elevated t (21.5°C) nor an unusually large $\Delta t'$ (0.7°C). In fact, because this –4.05 mm value is about 3.15 standard deviations (σ = 1.133 mm) off the mean (–0.427 mm), the probability of its occurrence is of the order of 2 to 3 × 10^{-3}. Because there is nothing that is otherwise unusual about this specific running, it could be argued that it is so unrepresentative that it should be expunged from the 35 runnings that comprise the results of this experiment. Were this suggestion to be adopted, the divergence would drop to –10.89 mm—or only about 3.5 mm off the computed standard error of 7.4 mm for the entire line (Heroux and others, 1985, p. 317). This operation, perhaps unwisely, would cast doubt on any conclusions developed from this experiment.

The relatively low and poorly defined threshold values for both temperature and $\Delta t'$ (figs. 22, 23) are seemingly inconsistent with those obtained from the Saugus-Palmdale results. For example, the average setup-elevation difference (Δh) for the LSL Canadian leveling was 0.788 m (Heroux and others, 1985, as modified by W. Gale, 1992, written commun.), as contrasted with 1.194 m for the Saugus-Palmdale LSL leveling. Accordingly, because the Canadian experiment should have sampled a smaller difference in the thermal gradient in foresight and backsight observations, higher threshold values could reasonably have been predicted for the Canadian experiment, rather than vice-versa. These apparent contradictions probably are related to the contrasting nature of those factors that comprise the local environmental regime. The survey route chosen for the Canadian experiment is a gravel shoulder adjacent to an asphalt-paved highway (W. Gale, 1993, oral commun.), which itself traverses a generally moist, heavily vegetated landscape. It is, in other words, an environment analogous to one that might juxtapose a Saugus-Palmdale (or Tucson) regime against an almost limitless Gaithersburg (or Corbin) site. It is these same environmental complexities that may at the same time account for an average measured $\Delta t'$ of only 0.55°C (Heroux and others, 1985, p. 319).

COMPARISONS AMONG THE EXPERIMENTAL RESULTS

It should be clear at this point that no one of these three individual field tests for refraction error should be considered definitive vehicles for the assessment of this error in any *geodetic* (procedurally constrained) leveling. Nevertheless, inclusive examination of the experimental results not only enhances the value of the individual experiments, but highlights as well their clearly defined limitations. An improved understanding of both the benefits and deficiencies of these experiments provides the basis for a proposed experimental investigation of the more pertinent factors that enhance or inhibit the accumulation of URE.

COMPARISONS OF THE SAUGUS-PALMDALE AND CANADIAN TEST RESULTS AGAINST THE GAITHERSBURG-TUCSON RESULTS

The only two levelings currently available in which the URE was actually measured, exclusive of the Gaithersburg-Tucson simulated lines, are the Saugus-Palmdale and Canadian LSL surveys. The results of these surveys provide reasonable—if obviously limited—bases for testing both the utility and the reliability of the URE fields defined in figure 11. Normalization of the LSL refraction errors to the 491.4-m height difference used in the development of the worst-case curve in figure 11 leads to cumulative URE values of –30.8 mm (or –34.9 mm based on the corrected data) and –25.9 mm for the Saugus-Palmdale and Canadian LSL surveys, respectively; the corresponding σ_{DDH} values obtained from table 6 and from Heroux and others (1985, p. 318) are 0.286 mm and 0.29 mm. While we have also computed the normalized refraction errors for the LSL and SSL surveys where the errors are assumed to be directly proportional to the average sight length squared (fig. 11), all the evidence developed in the preceding paragraphs indicates that the URE was confined nearly entirely to the LSL levelings. Given this determination, several useful tests are suggested in addressing the questions of the reliability and utility of the empirically derived fields identified in figure 11.

URE VERSUS Δh

An obvious test that emerges from the results of the Saugus-Palmdale and Canadian experiments can be developed from the measured difference in the LSL URE values. Both the average sight lengths (42 m vs. 45 m) and the σ_{DDH} values were about the same for the two surveys. While the average $\Delta t'$ value (equation 1) for the Canadian experiment was less than half that associated with the Saugus-Palmdale experiment (Stein and others, 1986, p. 9033, 9036–9037), the averaged sectional temperatures were identical (23.77°C). Moreover, even though $\Delta t'$ is considered the theoretically more significant parameter, evidence developed in earlier sections of this report suggests that the ambient air temperature, in combination with locale, may be nearly as good an index of near-surface atmospheric refraction as $\Delta t'$. Thus, based on the assumption that the URE is a linear (or near linear) function of Δh (equation 1)—and disregarding any impact of the differences in sight lengths and $\Delta t'$ values—normalization of the LSL URE values to the same height difference should, in theory, produce a discrepancy of zero between the LSL URE results obtained from these two experiments. In fact, however, the Saugus-Palmdale value exceeds that obtained from the Canadian experiment by 4.9 mm (or 9.0 mm were the comparison to be based on the corrected Saugus-Palmdale data). This discrepancy is obviously consistent with the two-fold difference in the average $\Delta t'$ values, but it is inconsistent with the modest difference in sight lengths. Thus, to the extent that the differences in these two variables tend to nullify each other as contributing factors to the resulting URE values, it implies that equation (1) is nonlinear in Δh. However, while this likelihood is supported by the sense of the difference in the averaged LSL values for Δh obtained from the experiments, it is much less convincingly supported by the magnitude of this difference. That is, an increase in Δh of about 50 percent (from 0.788 m to 1.194 m) accounts for no more than a 19 percent increase in the URE (or about 35 percent based on the corrected Saugus-Palmdale data).

We can explore this issue still further by considering the relations between either the worst-case error associated with a σ_{DDH} value of 0.29 mm (about −51 mm—figs. 10, 11) or the cumulative and normalized −46.9 mm error associated with the simulated 45-m Tucson line (table 1) and the results of the Canadian and Saugus-Palmdale experiments. Both of the suggested comparisons are encumbered by drawbacks. Nevertheless, because the sight length matches the averaged Canadian LSL sight length, and because the corresponding Tucson 45-m σ_{DDH} value is only about 0.034 mm less than that identified with the LSL Canadian survey (table 1), we view the second approach as the more defensible—even while recognizing our ignorance of the averaged $\Delta t'$ values associated with the Tucson 45-m line. Because Δh for the simulated 45-m Tucson line is given as 1.922 m (table 2), it constitutes a 144 percent increase in Δh over that for the Canadian LSL survey. Accordingly, because this 144 percent increase is associated in turn with an 81 percent increase in the URE (or corresponding increases in Δh and URE of about 61 percent and 52 percent, respectively, based on a comparison against the Saugus-Palmdale field data), this particular comparison strongly reinforces the tentative conclusion that equation (1) is significantly nonlinear in Δh.

MEASURED URE AS A FUNCTION OF THE WORST-CASE CHARACTERIZATIONS

A second test of the approach developed in figure 11 can be obtained from comparisons between the results of the two field tests and the represented error fields (fig. 11). The Saugus-Palmdale experiment was the only one in which the order of the leveling was actually specified for both the SSL and the LSL surveys. Nevertheless, because the average sight lengths and average sectional temperatures were about the same, it is likely that the *unimposed* procedural constraints on the Canadian test levelings were about the same as those specified for the Saugus-Palmdale levelings. Thus, and reading directly from figure 11*A*, *B*, the Saugus-Palmdale SSL survey met both 1974 and earlier and 1975 and later first-order standards, whereas the Canadian SSL survey met only pre-1975 first-order standards. Similarly, both of the LSL surveys met no better than 1975 or later third-order standards or 1974 and earlier, upper-limit, second-order standards (fig. 11*A*, *B*). Because (and even though improperly implemented) the setup-rejection criterion imposed on the Saugus-Palmdale SSL survey was ±0.30 mm (Stein and others, 1986, p. 9038–9039), the determination that both of the SSL surveys met first-order (or near first-order) standards could be reasonably anticipated (Federal Geodetic Control Committee, 1980, p. 28). Again, because the setup-rejection criterion imposed on the LSL Saugus-Palmdale survey was 0.75 mm, we could anticipate that the LSL surveys were run to no better than post-1974 third-order standards (Federal Geodetic Control Committee, 1980, p. 28) or pre-1975 upper-limit, second-order standards (fig. 11*B*). This position, however, conflicts with that of Stein and others (1986, p. 9038) who state that "the long-sight leveling conducted during the experiment meets all the standards and specifications for first-order leveling performed during 1955–1965, which it was designed to duplicate."

The inconsistency between our determination and the preceding statement of Stein and others is rooted in the fact that the LSL survey was *not* run to first-order standards. A 1974 and earlier procedural requirement for first-order leveling states that "***each section should be run forward and backward, the two runnings of a section not to differ by more than 4 mm. √K *** where K is the length of the section in kilometers***" (Rappleye, 1948, p. 2). This particular requirement, moreover, dates back to at least the turn of the century (Hayford, 1904, p. 213).

In order to justify the adoption of the 0.75-mm setup-rejection criterion for the LSL survey as a means of achieving a single-run equivalent of a pre-1966 first-order leveling, Stein and others (1986, p. 9040) quote Rappleye (1948, p. 7) as follows:

> Observers have found that a convenient rule, in fixing the length of sight, is to shorten the sights whenever the upper and lower intervals subtended on the rod [as viewed through the Fischer three-wire level] are found to differ frequently by more than a selected limit. Each observer should fix this limit from his own experience, by noting the relation between a provisional limit and the amount of rerunning found to be necessary while using it.

Stein and others (1986, p. 9040–9041) continue with these derivative statements and additional observations:

> Because the subtended intervals (the height difference between center and distal reticles) were only estimated to the nearest millimeter, the most restrictive standard possible was δ = 1 mm (i.e., rejecting all sights for which the thread intervals do not agree exactly). In 1955 the observer used a 3-mm tolerance (associated with a 60-m average sight length), and in 1961 a 2-mm tolerance was adhered to (with L = 51 m) [both along the Saugus-Palmdale line]. Thus the sighting standards of the 1981 long-sight survey (δ = 0.75 mm) were tighter than those in 1955–1965 (δ = 2–3 mm).

Each of the preceding points is considered here, more or less in order. (1) "Rappleye's rule," as suggested by his phrase "differ frequently," was no more than a general guideline; it clearly was *not* a rejection criterion. For example, a National Geodetic Survey observer whose experience with the Fischer level extended over nearly a decade states: "The spider-web threads could sag by as much as to give a 3-mm difference on the subtended interval figure. The reading commonly would be rejected if the difference went to 3 mm, and it was (in any case) only the section-rejection limit that could be codified. The clarity of the image on the rod was the most important control in ensuring that sight lengths were not overextended" (Rodney Jack Lee, 1991, written commun.). It is the last part of this statement that focuses on the correlation between scintillation and cumulative refraction error recognized by any experienced observer. (2) The "most restrictive standard possible" obviously could have been less than 1 mm. Moreover, even were the observer to choose 0 mm as his standard, there was certainly no requirement that he reject *all* sights for which "the thread intervals do not agree exactly." (3) In equating the three-wire "rejection criterion" based on the thread interval difference subtended on the rod with that based on the high-scale low-scale difference (DDH) obtained from optical micrometer readings of the precisely delineated marks on double-scale rods, Stein and others (1986, p. 9040) improperly equate conceptually different measurements. The design thread interval between upper and lower threads for the Fischer three-wire reticule is 1.23 mm—or 0.615 mm between middle and upper or lower threads (E.I. Balazs, 1991, written commun.). The stadia constant of this instrument is such that at 100 m the image of upper and lower threads would subtend an interval on the rod of 300 mm. Had the reticule threads been set with micron accuracy and remained absolutely invariant with respect to each other, a more restrictive setup-rejection criterion than "Rappleye's rule" might have been possible. However, neither of these requirements was met in conjunction with each other, and there is some doubt that they could be met today. (4) Finally, simple parallax introduces yet another uncertainty in viewing the reticule lines as subtended on the rod. By its very nature, accordingly, a three-wire, totally instrument (reticule)-dependent, setup-rejection criterion can be neither as accurate nor as restrictive as one which can be achieved through the use of double-scale rods in conjunction with the optical micrometer. All of this by way of saying: (1) that the Stein and others (1986, p. 9041) conclusion that "the sighting standards of the 1981 long-sight survey * * * were tighter than those in 1955–1965" is based on an invalid comparison; and (2) that the pre-1975 first-order sight lengths were controlled not by "Rappleye's rule," but rather and necessarily by the section-rejection criterion of 4 mm/km$^{1/2}$.

In support of their postulate that the LSL survey matched the standards and specifications for 1955–1965 first-order leveling, Stein and others (1986, p. 9038) note that the misclosure in only one of the experimental sections exceeded the 1955–1965 section-rejection limit of 4.0 mm (L)$^{1/2}$, where L is given as the section length in kilometers. However in any routine, double-run leveling there is no a priori basis for assuming that one or the other of the runnings was more accurate. The described misclosures of the LSL survey, on the other hand, were against the SSL running, where the much more restrictive setup tolerance offers presumptive evidence that the SSL survey was much more accurate than the LSL survey. Because both surveys utilized the same rod pair, we assume as a first approximation that rod error made no contribution to the LSL-SSL sectional divergence. Refraction and staff and (or) instrument settlement remain as the most likely sources of systematic error in the SSL survey. As we have already shown, however, the SSL survey was contaminated by little if any cumulative refraction error over the full length of the line; thus, the SSL URE must have approached zero within an individual section. Moreover, if we accept as representative the average settlement value of 0.014 mm/setup developed by Craymer and Vaníček (1986, p. 9053), it is unlikely that settlement could have contributed more than a few tenths of a millimeter to the sectional divergence in even the longest sections. Accordingly, these several considerations argue that the SSL survey was sufficiently free of systematic error that it would not have been surprising had all of the LSL closures on the SSL sections met the 4.00 mm (L)$^{1/2}$ rejection limit.

An alternative, more defensible assessment of the extent to which the LSL leveling met pre-1965 first-order standards is based on a root-mean-square determination of the first-order rejection limit, where the SSL survey is assumed to have been free of significant systematic error only. Thus,

$$R_C = \sqrt{\varepsilon_S^2/2 + \varepsilon_L^2/2}, \tag{6}$$

where

- R_C = calculated first-order rejection-criterion equivalent;
- ε_S = sectional error limit for the SSL survey, where the estimated 1σ random error of 1.00 mm $(L)^{1/2}$ (Stein and others, 1986, p. 9039) is assigned entirely to the SSL survey; and
- ε_L = sectional error limit for the LSL survey, where it has been run to pre-1965 first-order standards—4.00 mm $(L)^{1/2}$.

Accepting this characterization of the pre-1965 first-order rejection-criterion equivalent for the sectional closures of the LSL survey on the SSL survey, R_C = 2.92 mm $(L)^{1/2}$. Accordingly, *six* of the finally accepted sections in the Saugus-Palmdale experiment failed to close to pre-1965 first-order standards—*following*, that is, any required rerunnings in either direction.

Because the three-wire "Rappleye's rule" comparison with the double-scale rod setup-rejection criterion leads to an invalid conclusion and because their misclosure argument lends little if any support to their basic postulate, the arguments presented by Stein and others (1986) fail to contribute to their contention that the single-run LSL survey met 1955–1965 first-order standards. Thus, we find the combined experimental-observational evidence to be far more compelling: specifically, the Saugus-Palmdale LSL survey barely met 1975 and earlier second-order standards (fig. 11).

URE AS A FUNCTION OF THE TOTAL ENVIRONMENT

It would be difficult to reconcile the differences among the results of the three experiments examined in this study without appealing to various environmental factors, only two of which, cloud cover and wind, are normally considered in the modeled refraction corrections. Any environmental factor that limits the incident solar radiation or the derivative upward sensible heat flux will tend to damp the URE. Nearly all of the interpretative complexities, however, are related to those environmental elements that impact the upward sensible heat flux and the resultant curvature in the thermal gradient.

The contrast in the results obtained from the Gaithersburg observations and those that emerged from the conceptually identical Tucson experiment documents the importance of these environmental factors and, at the same time, provides the basis for an instructive comparison with the Saugus-Palmdale and Canadian field tests. The Gaithersburg-Tucson comparison takes on added significance because of, rather than in spite of, the fairly compelling evidence that the O–S sloping line values obtained from the Gaithersburg experiment (table 1) grossly exaggerate the actual URE. Any comparison between the Gaithersburg and Tucson results obviously requires that we accept the Tucson geodetic results as accurate representations of the URE accumulated along the simulated lines. Moreover, because the Tucson temperature data are also missing, we can only infer that the $\Delta t'$ and *CURV* values at Tucson probably were much greater than they were at Gaithersburg and, hence, that the cumulative URE at Tucson should be several times that at Gaithersburg. Even without these data, however, there is an excellent rationale for this inference.

One of the two environmental controls generally considered in URE modeling is cloud cover—a factor whose significance may be underestimated in the modeling process. During the course of the Gaithersburg observations the sky was partly to mostly overcast about one-third of the time, whereas the Tucson experiment was accompanied by virtually cloudless skies. However, the difference in cloud cover cannot by itself account for the sharply disparate accumulation of URE at the two sites, especially since the temperature range in the two locales was about the same (compare figures 5 and 6 with table 8). The chief difference between the two sites was not in the cloud cover, but rather in the nature of the surface and near-surface regime. Owing to the absence of vegetation—other than scattered, near dormant desert shrubs—and the thoroughly dessicated surface materials, a large fraction of the incident radiation at Tucson probably was returned as sensible heat. At Gaithersburg, on the other hand, and quite unlike Tucson, an undetermined but sizable fraction of the thermal input went into latent heat associated with evaporation and transpiration off an actively transpiring turf, thereby robbing the boundary layer of thermal energy that could otherwise have contributed to the sensible heat flux. The source of the evaporated water, other than that obtained through transpiration, could have been residual surface moisture from earlier rain, dew, or even capillary rise through the underlying soil. Nevertheless, judging from the positive values for $\Delta t'$, that portion of the incident radiant energy soaked up by evapotranspiration at Gaithersburg (or Corbin) was less than that required to totally suppress the sensible heat. It may have been quite sufficient, on the other hand, to reduce *CURV* to near zero values with corresponding reductions in the URE to well below those supposedly obtained from the experimental observations (table 1).

The significance of wind, the other of the two environmental factors considered in the refraction model of the National Geodetic Survey, is ill defined by any of the experimental results examined in this report. Although a three-tiered coded value for wind speed was recorded at Gaithersburg (and presumably at Tucson as well), we have not even attempted to compare the URE against the coded value for wind speed at this site. Although comparisons between wind velocity and $\Delta t'$ or *CURV* could have been made at the Gaithersburg 60-m instrument station, because the actual wind speeds are unknown and because both $\Delta t'$ and *CURV* are characterized by very modest values coupled with significant scatter (table 3, fig. 6), any results obtained from this comparison would be so clouded with uncertainty as to be useless. This generalization is nearly as valid in considering the much more rigorously characterized measurements obtained from the Saugus-Palmdale field test.

With the Gaithersburg-Tucson controls as background, we can turn to a consideration of the threshold values for temperature, $\Delta t'$ and *CURV* detected in the Saugus-Palmdale (figs. 18–20) and, less definitively, the Canadian experiments (figs. 22, 23). This consideration is predicated on the acceptance of a functional relation between Δh and the several threshold values. By way of illustration, the mean Δh along the Saugus-Palmdale LSL line was 1.194 m and was associated with threshold values of 23.8°C for $t_{1.5}$, of 1.08°C for $\Delta t'$, and of 0.33°C for *CURV*—all based on the field data. Relying solely on the regression equations (fig. 9), an increase in Δh to the maximum setup elevation difference of 2.0 m—since the temperature measurements were made over the full 2-m interval—and a stipulated reduction to zero in both $\Delta t'$ and *CURV* leads to significantly reduced threshold values for $t_{1.5}$ of 7.6°C and 10.3°C, respectively. Similarly, any decrease in Δh would produce an undetermined increase in all of the threshold values given in figures 18–20.

If we consider now the threshold values for $\Delta t'$ and *CURV* (figs. 19, 20) as a function of the National Geodetic Survey's sun code (fig. 24), neither of these parameters exceeded their indicated threshold values until the sun code had reached 1.0 or above. Because, as we have already seen, $t_{1.5}$ proxies fairly well for both $\Delta t'$ and *CURV*, it is hardly surprising that the threshold value for $t_{1.5}$ was not exceeded until the sun code had itself exceeded 1.0 (fig. 24). From this observation, we can reasonably conclude that the sensitivity of the URE to the ambient air temperature is attributable in large measure to the fact that $t_{1.5}$ proxies not only for the sensible heat flux but, less specifically, for the incident solar radiation as well. Moreover, while the Saugus-Palmdale line traversed a more-or-less arid terrain, it was much less so than that at Tucson. For example, moist surfaces were actually observed during

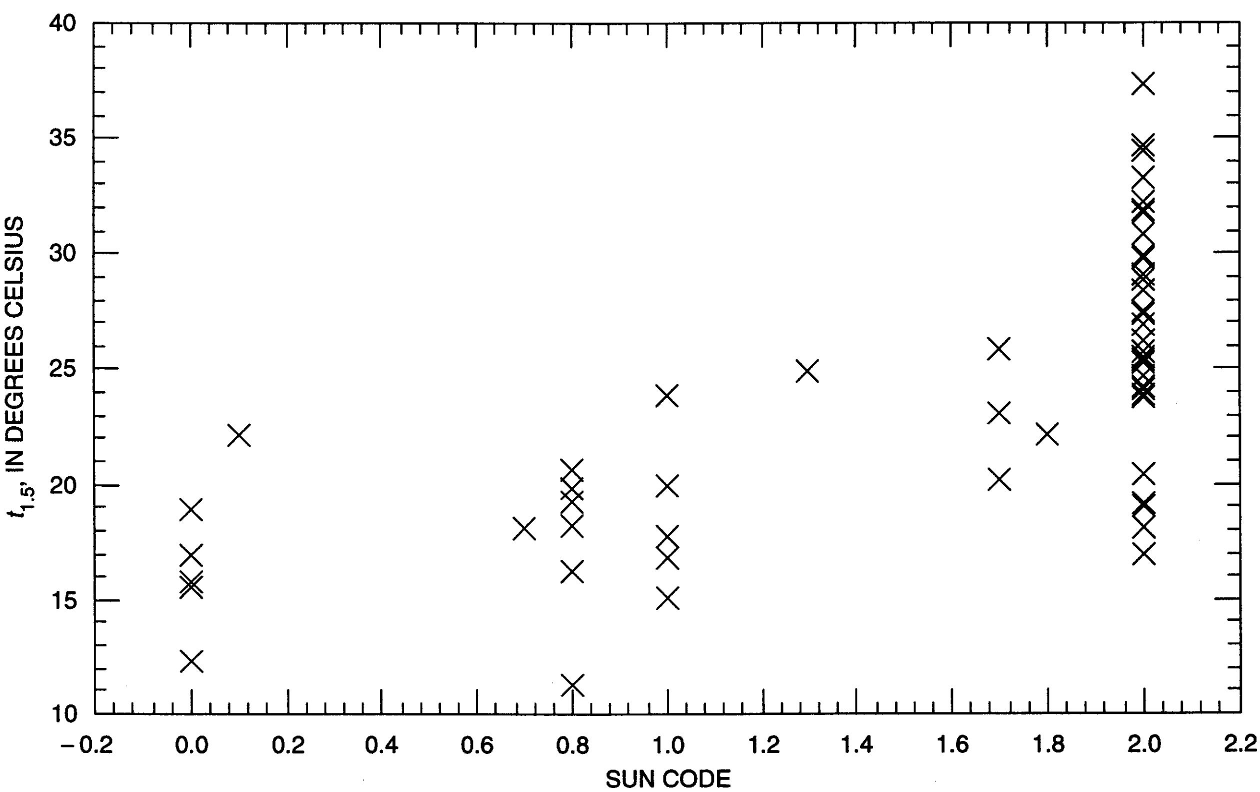


Figure 24. $t_{1.5}$ versus National Geodetic Survey sun code on a section-by-section basis along Saugus-Palmdale LSL line. 0, cloudy (less than 25 percent sunny); 1, partly cloudy (from 25 to 75 percent sunny); 2, clear (more than 75 percent sunny). Sun-code values obtained from averaging of setup observations.

the course of the experiment, and the occasional rain guaranteed that the surficial materials were far less desiccated than those at Tucson. In addition, a mix of grasses and other types of vegetation was recognized at various points along or adjacent to the survey route. Because extraction of heat from the thermal input in response to evapotranspiration diminishes the sensible heat—and hence the ambient air temperature—quite like that which occurs with increasing cloud cover, temperature becomes a combined index of the potential for the accumulation of any URE. In other words, until the upward sensible heat flux meets *its* threshold value, the associated threshold values for $t_{1.5}$, $\Delta t'$, and *CURV* for any stipulated Δh could not exceed those values at which the URE would begin to accumulate.

The impact of wind on the LSL URE (the LSL-SSL divergence) generated in connection with the Saugus-Palmdale field test is both puzzling and seemingly counter intuitive. That is, a comparison of the URE against wind velocity on a section-by-section basis discloses statistically significant positive correlations between URE and wind velocity (fig. 25). While wind velocity is also correlated with sun code (correlation coefficient = 0.486), wind may be a less significant parameter in controlling the accumulation of URE than common sense would suggest. For example, if we consider the Saugus-Palmdale data to be representative, a multivariate linear regression of URE (the LSL-SSL divergence) on wind, sun code, and temperature ($t_{1.5}$) indicates that temperature is the only one of these variables associated with a statistically significant coefficient. At least a part of this apparent violation of common sense may be explained by the theoretical modeling of Shaw and Smietana (1983) who show that the relations between the URE, heat flux and friction velocity (which is approximately proportional to wind velocity for any specified roughness length) are both unexpected and complex. For any given positive heat flux the URE *increases* with increasing friction velocity to some URE maximum, above which the URE diminishes with further increases in friction velocity. Moreover, because this URE maximum migrates to progressively lower friction velocity values with decreasing heat flux, wind may assume a more important role with increasing cloud cover and roughness lengths of the order of 0.10 m or greater. Finally, because experienced surveyors find that the effect of wind on URE (or scintillation) depends in large measure on a variety of ancillary considerations—such as the direction of wind versus the direction of sighting, the nature of the adjoining ground cover, and so forth—it would at this point be unwise to fall into the trap of overstating the significance of wind as an inhibiting factor in the accumulation of URE.

If, as we have seen, a threshold value exists for $t_{1.5}$ (as well as $\Delta t'$ and *CURV*) associated with the LSL Saugus-Palmdale leveling, we should also expect to see a corresponding threshold value for scintillation (σ_{DDH}). While not directly measured, several procedures permit fairly accurate estimates of the threshold scintillation. The first is based on the estimated value for σ_{DDH} associated with the 0.50-mm cutoff—or the 0.45-mm cutoff based on the corrected data—developed from the maximum DDH per section for the LSL survey (fig. 16), where neither the LSL nor the SSL surveys were contaminated by measurably significant URE below this cutoff. This approach suggests that the threshold scintillation (σ_{DDH}) for the LSL survey was about 0.17 to 0.19 mm (table 6). Alternatively, because the SSL survey probably was contaminated by near zero URE, we can examine the upper bound of the

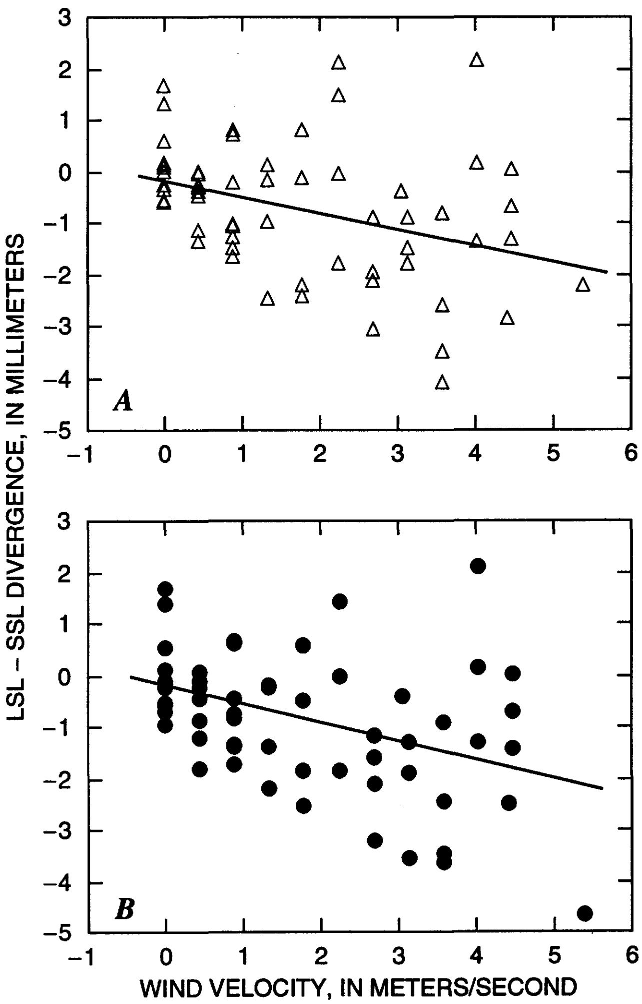


Figure 25. LSL-SSL divergence per section versus average LSL sectional wind velocity (v_w) obtained from Saugus-Palmdale field test. *A*, Based on field data. *B*, Based on corrected data. Regression equations shown by straight lines are: *A*, URE (LSL-SSL divergence) = $-(0.180 \pm 0.240) - (0.314 \pm 0.103)v_w$; R^2 = 13.8%; *B*, URE (LSL-SSL divergence) = $-(0.199 \pm 0.250) - (0.363 \pm 0.107)v_w$; R^2 = 16.4%. Sectional slopes are so fixed that direction of running is in all cases uphill (see fig. 16).

σ_{DDH} values for the SSL survey on a section-by section basis, a procedure that yields a threshold value for scintillation (σ_{DDH}) of about 0.18 mm (fig. 26). In either case, it is this value for the scintillation below which we could expect to see no significant LSL-SSL divergence—effectively, the URE in the LSL leveling. Unlike the threshold values for $t_{1.5}$, $\Delta t'$, or *CURV*, however, this threshold is explicitly associated with sighting conditions and, hence, specific rejection criteria, rather than the particulars of the survey (*L*, Δt, and so on). Thus, the corresponding setup- and section-rejection criteria could be expected to be of the order of 0.50 mm and 4.00+ mm/km$^{1/2}$, respectively, in the first case and 0.40 to 0.50 mm and 4.00+ mm/km$^{1/2}$, respectively, in the second case (table 6). It is clear, in any case, that the threshold level for scintillation is attributable to the same environmental elements that ultimately control the threshold values for $t_{1.5}$, $\Delta t'$, and *CURV*.

In the context of the preceding discussion, the results of the Canadian experiment are particularly enigmatic if one grants that these results are generally consistent with what one might expect in comparing LSL leveling against SSL leveling. For example, we recognize both threshold sectional temperatures and threshold sectional $\Delta t'$ values in the Canadian data, below which there seems to be little accumulation of URE (figs. 22, 23). Both, however, are well below those that emerged from the Saugus-Palmdale comparisons, even though the mean value for Δh (0.788 m) identified with the LSL survey is less than the corresponding Saugus-Palmdale value (1.194 m). These differences are the inverse of those expected in a comparison between a leveling through a semi-arid regime (Saugus-Palmdale) and one through a temperate regime (Eastern Townships). In addition, the average

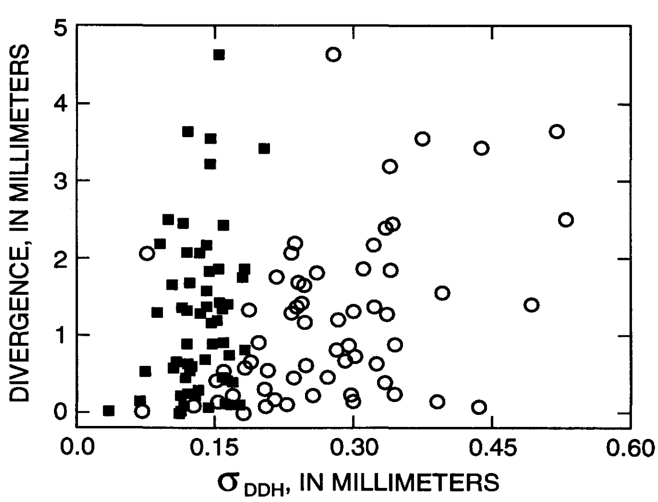


Figure 26. Divergence versus standard deviation about the mean DDH (σ_{DDH}) on a section-by-section basis for both LSL (open circles) and SSL (solid squares) Saugus-Palmdale experimental levelings (from Castle and others, 1983, p. 1082).

$\Delta t'$ given for the Canadian experiment (0.55°C) is less than half that for the Saugus-Palmdale experiment (table 5) and only slightly above that obtained from the Gaithersburg field test (table 3). Equally puzzling is the result obtained from a comparison of the sectional temperatures against the sectional values for $\Delta t'$ (fig. 27), where the absence of any correlation is inconsistent with that expected on both theoretical and observational grounds (compare with figure 9). Because a deliberate effort was made to select test days characterized by a high potential for the generation of URE, it is no surprise that 77 percent of the sections are identified with a sun code of 2 (less than 25 percent cloud cover). Accordingly, there is little variability in the sample population, such that the sun code and sectional URE are uncorrelated—even though both the sectional temperatures and $\Delta t'$ predictably increase with increasing values for sun code.

While it could be argued that the significance of the Canadian data should be largely discounted as concatenations of error in either the geodetic or temperature measurements, a variety of environmental factors could and probably must have influenced the LSL-SSL divergence—and thereby provide at least a partial explanation for some of the seeming contradictions described in the preceding paragraph. We can begin by considering the contrast in the average measured value for $\Delta t'$ (0.55°C) and that predicted for this same parameter (Heroux and others, 1985, p. 319). The two-fold increase in the predicted over the observed value is consistent with the near-maximum time-averaged sun code that could be expected in eastern Canada during the summer months. Moreover, because oil- and tar-soaked surfaces are represented as absorbing and reradiating heat especially efficiently (Stein and others, 1986, p. 9034), the predicted $\Delta t'$ should have been even higher had this factor been considered in the modeled value for $\Delta t'$. It is fairly clear, in other words, that the mitigating influence of various environmental factors damped the upward sensible heat along the line of sight to much lower values than might have obtained had the gravel/asphalt surface extended broadly beyond the survey route. Owing to the abruptly contrasting thermal regimes identified with an asphalt surface on one side and one identified with a generally or intermittently moist and actively transpiring mat of weeds, grasses, and other plants on the other side, the thermal structure within the atmospheric boundary layer along the join between these regimes (where most of the leveling observations were actually made) must have been remarkably chaotic. Evidence of this chaos is certainly suggested, for example, by the uncorrelated—exceptionally noisy—relation between the average sectional temperatures and corresponding values for $\Delta t'$ (fig. 27).

If the preceding description of the atmospheric boundary layer along the route of the Canadian experiment—relatively low heat flux marked by an extremely convoluted thermal structure—is indeed plausible, why, it might be asked did *any* URE accumulate in connection with these surveys? Several observations may contribute to a resolution of this question. For example, the LSL line-average value for scintillation (σ_{DDH}) is given as 0.29 mm (Heroux and others, 1985, p. 318); this is almost 60 percent above the probably minimum threshold value (~0.18 mm), such that there is an expectation for at least some contamination of the LSL survey by URE. This is, of course, a permissive argument, since the threshold value defines a value below which we could anticipate zero URE; it does not compel that URE must accumulate where this value is exceeded (see, for example, the σ_{DDH} values given for the zero-slope simulated lines obtained from the Gaithersburg-Tucson experiment—table 1). Secondly, the $\Delta t'$ values (fig. 27) are sectional averages that could easily mask significant thermal episodes within the timeframe required to complete a full section of geodetic measurements. The occurrence of occasional convective plumes accompaning advective surges across the hot asphalt surface could have momentarily increased $\Delta t'$ by several multiples, which would in turn have produced exceptionally large increments of URE. Finally, the large number of reversals in the accumulated divergence along the simulated 34-km line suggests that this divergence may be little more than unsystematic noise—although the near-perfect concordance between the cumulative divergence and the modeled correction based on measured values for $\Delta t'$ seems to argue otherwise (Heroux and others, 1985, p. 319).

In addition to the already enumerated environmental elements that could impact the curvature in the idealized thermal gradient, there are a number of other unmeasured parameters that could perturb this gradient in such a way as to randomize or even reverse it. For example, partial shading along the line of sight would chop up the boundary layer into a series of discrete cells in which the contrasting and interfering flow of the upward sensible heat flux would set up air-current eddies that would almost certainly modify the idealized curvature—and the URE. Similarly, drainage of air through areas of irregular topography could lead to structurally unpredictable effects at the confluence between drainage channels. Alternation of surface materials of contrasting thermal properties (specifically, heat capacity and conductivity) would also tend to set up mutually interfering eddies that could or should minimize any proclivity toward the accumulation of the normally expected URE. Finally, while the albedo of the surface materials tends to control the sensible heat and, hence, $\Delta t'$, it would not in itself lead to destructive interference within the normal thermal structure. It could, however, seriously limit $\Delta t'$ and reduce *CURV* to insignificant values.

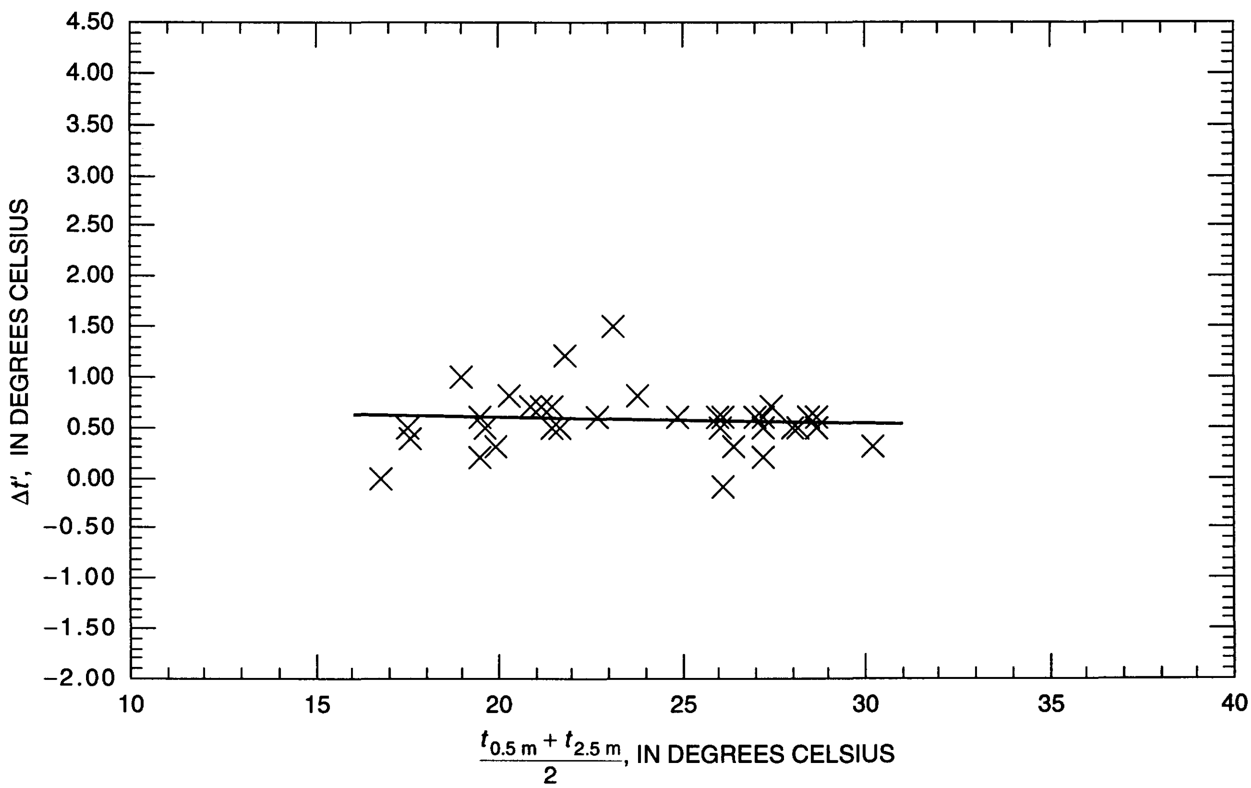


Figure 27. Averaged sectional $\Delta t'$ as a function of averaged sectional temperatures obtained from Canadian refraction experiment. Regression equation is $\Delta t' = (-0.689 \pm 0.321) + (0.0054 \pm 0.0133)t$; $R^2 = 0.5\%$.

URE ASSOCIATED WITH FIRST-ORDER LEVELING

The inclusive results of two of the three refraction tests examined here permit fairly accurate estimates of the maximum and extremely unlikely URE that could be expected to have been associated with any 20th-century first-order leveling.

Provided that the DDH was constrained to a limiting value of 0.50 mm—or 0.45 mm based on the suspect corrected data—the URE apparently summed to no more than 6 to 7 mm over a height difference of ~350 m—the Saugus-Palmdale line-height difference less that associated with those sections in which the DDH >0.50 mm (fig. 17). The 0.50-mm—or 0.45-mm—setup-rejection criterion is identified with (σ_{DDH}) values that fall within a range of about 0.17 to 0.19 mm (table 6), and thus marginally above the limiting boundary of the pre-1975 first-order field (fig. 11*B*). Accordingly, had a 0.45 to 0.50-mm setup-rejection limit been in force throughout the 20th century, a criterion that met no better than second-order standards, we could anticipate the accumulation of URE of no more than about 15 to 20 mm over a height difference of 1000 m. In fact, however, the corresponding *section*-rejection equivalent of the 0.50-mm (or 0.45-mm) setup-rejection limit is above the 4.00 mm/km$^{1/2}$ section-rejection criterion for first-order leveling at the 10 percent tolerance level—great enough, that is, to put it well within the σ_{DDH} range of 0.17–0.19 mm (fig. 11*B*). Because the 4.00 mm/km$^{1/2}$ section-rejection criterion has served as the chief procedural constraint on first-order leveling throughout the 20th-century, we could (or should) reasonably conclude that any 20th-century first-order leveling was contaminated by near-negligible URE that, at worst, was comparable to random error.

The preceding conclusion might be challenged on several grounds. For example, the modest URE associated with a setup-rejection criterion of 0.50 mm (figs. 16, 17) was derived from the Saugus-Palmdale LSL leveling identified with a mean Δh of <1.2 m. Assuming identical sighting conditions within the cumulative sections in which the DDH ≤0.50 mm, any increase in Δh and (or) L could be expected, if not guaranteed, to produce a measurably significant increase in URE; however, this postulated increase in Δh or L would at the same time ensure increased scintillation along the line of sight and DDH values in excess of 0.50 mm. Thus, this procedural restriction would be equally effective in inhibiting the accumulation of URE, regardless of the particulars of the survey. Yet another observation that seems to suggest that near-negligible URE in any first-order leveling may be an unwarranted exaggeration is the correlation of the Gaithersburg −9.1-mm cumulative refraction error with a σ_{DDH} value of only 0.133 mm (table 1)—well within the first-order error field (fig. 11*B*). Nevertheless, and without reiterating our bases for suspecting that none of the Gaithersburg observations were significantly contaminated by URE, the −9.1-mm value is an aberration in its own right. Specifically, the noise/signal ratio of 1.54 (table 1) is not only the highest of any of the six sloping-line cumulative observations obtained from the Gaithersburg-Tucson experiment, but demonstrates as well that the −9.1-mm value is statistically indistinguishable from zero. Moreover, the computed σ_{DDH} for the simulated 30-m sloping line (0.133 mm) is anomalously low; the corresponding value obtained from the zero-slope line is 0.163 mm (table 1).

The regression curve based exclusively on the Tucson test observations (fig. 15) carries with it the strongest objection to any conclusion that 20th-century first-order levelings must have been contaminated by trivial values of URE. If we adopt 0.17 mm as the upper limit for first-order values of σ_{DDH} (fig. 11*B*), then figure 15, curve D, suggests that a worst-case first-order URE over a 1,000-m height difference is about −45 mm, a good deal greater than "near negligible." Nevertheless, and disregarding the uncertainty in the measurements of the three points that define this curve (table 1), the −45-mm estimate is a worst-case value not only in the sense that it is based on the rarely—if ever—realized maximum permissible 2-m Δh constraint, but also in terms of the site charateristics and meteorologic conditions. The surface materials at the Tucson site, probably an abandoned stream bed, consist of a mix of coarse sand and pebbly gravel interspersed with the occasional mesquite or other desert shrub. Other than Death Valley, perhaps, it would be difficult to identify a more arid regime in North America. Moreover, based on weather records obtained from the National Climate Center, the percent of possible sunshine during the April 1980 observation period was at a near maximum (97.6 percent). Thus, the combination of the surface environment and clear skies indicate that the heat flux as a function of ambient air temperature (which averaged 27.1°C at 1400 hrs, MST) must have been at near peak values at the Tucson site. The resulting thermal structure of the boundary layer at this site could be expected to produce near maximum URE values at the indicated temperature levels (table 8). Accordingly, while −45 mm is significantly greater than near negligible, the likelihood of ever accumulating URE of this magnitude in any routinely produced first-order leveling is vanishingly small.

Finally, the threshold scintillation (σ_{DDH}) of about 0.18 mm, below which the URE probably does not accumulate (see preceding section), indicates that the URE should rarely exceed zero in any first-order leveling (fig. 11*B*). Thus it seems that what should command our attention is not the differences among these three semi-independently obtained estimates, but rather the surprisingly good correspondence between the doubly worst-case value of −45 mm and the zero to near-trivial estimates based on analyses of the Saugus-Palmdale experiment—all of which are clearly at odds with representations of URE in routinely produced first-order leveling of as much as 2×10^{-4} times the measured elevation difference (Strange, 1981, p. 2814).

The experimentally based determination of near-negligible accumulations of URE in any first-order leveling should not be construed as an extreme position. It is, for example, clearly supported by the analyses of Mark and others (1987), who concluded from an examination of thousands of kilometers of generally first-order leveling that the survey procedures designed to suppress the URE have worked remarkably well during the full span of the 20th century. It is fully consistent as well with the experience-based view of countless observers, all of whom quickly learned that they could not meet the several rejection criteria if the scintillation along the line of sight—and hence a proclivity toward the accumulation of both the URE and atmospheric noise—rose to such levels that the image on the rod tended to blur. In meeting the rejection criteria under scintillating conditions, the levelman would automatically shorten his sight lengths, which, consistent with theory (equation 1), would diminish the accumulation of any URE. Moreover, at least three arguments suggest that this action would so diminish this error as to eliminate its accumulation. (1) The smaller the anticipated URE, the more likely that the impact of various unmodeled effects—such as advection, partial shading, contrasting conductivities and heat capacities in adjacent surface materials, and so forth—would lead to its randomization. (2) In shortening his sights, the observer would be less apt to trap (include) one of the convective plumes that characterize the unstable atmospheric conditions that produce optical refraction (see, for example, Castle and others, 1985, p. 241). (3) With shortened sight lengths, the dt/dz values in the contributing parts of the thermal gradient—and hence the refraction errors—in foresight and backsight observations are much more apt to equalize simply because the foresight-backsight readings necessarily sample a much narrower band of the atmosphere (fig. 1).

Surveying logistics in the United States encourage control leveling in the Southern States during the winter, migrating northward with the advancing seasons. Adherence to this approach tends to ensure that much of the leveling was carried out under temperate conditions, and thus commonly below the locale and probably seasonably dependent threshold temperatures below which the URE does not seem to accumulate. These procedures have not, of course, been rigorously implemented. For example, the first-order line L-22449 (table 6) was run in the high desert of southern California during June, a consideration that probably accounts for the relatively large σ_{DDH} value (0.165 mm) associated with this leveling. Similarly, the second-order line L-22643 was run through southern Minnesota in December and was identified with the relatively, but not necessarily unusually, small second-order σ_{DDH} value of 0.206 mm (table 6).

ESTIMATES OF URE ASSOCIATED WITH LOWER-ORDER SURVEYS

We can attempt to deal with the more general problem of the probable accumulation of URE in any given second- or third-order leveling by drawing on regressions of the sort developed in figure 21. A significantly limiting feature in this approach is that the URE is compared against Δh, which in reality represents the combined effects of both Δh and L^2, where the average value for L (42 m) is treated as a constant. Accordingly, if a determination of the URE were to be made for some leveling in which the average value for L departed from 42 m, we would be forced to base our computation of the URE on the theoretical relation between the URE and the square of the sight length. Accepting these limitations, we can consider two hypothetical levelings run to pre-1975 second-order standards over a 491.4-m elevation difference at a line-averaged temperature of 23.77°C. The two levelings are stipulated as differing significantly in their average Δh values—1.194 m for the first and 0.788 m for the second—and, less significantly, in their average sight lengths—42 m for the first and 45 m for the second. These hypothetical levelings are, of course, simply the LSL surveys carried out in connection with the Saugus-Palmdale and Canadian field tests; our obvious goal is to test the extent to which the regression regenerates the measured and normalized values shown in figure 11.

The results of this test suggest that the regression equation given in figure 21 is an inadequate vehicle for estimating the probable URE. At the same time, however, these results provide us with some insight as to how the experimentally based estimates of the URE might be significantly strengthened. Calculation of the cumulative URE values based on our regression (fig. 21) produces normalized values of −19.1 mm for the Saugus-Palmdale LSL survey and −21.1 mm for the Canadian LSL survey—roughly 62 and 81 percent, respectively, of the URE values shown in figure 11. The contradictory nature of these values, whereby the larger predicted value is associated with the smaller of the two measured values (fig. 11), clearly demonstrates the limitations of the regression as a predictive tool. We suggest above that the URE *may* be improperly characterized as a linear function of Δh in equation (1). In fact, the regression results suggest that the relation between the URE and Δh departs significantly from linearity—a conclusion at variance with that of Stein and others (1986, p. 9034–9036) who conclude that the equation (1) approximation accounts for nearly all the LSL-SSL divergence.

It could be argued that we are expecting too much from a single regression characterized by an R^2 value of 38.1 percent. Nevertheless, the indicated inconsistency in the calculated values of the URE suggests that it is equally likely that the approximately linear relation between Δh and the URE breaks down dramatically as the sampled band of the atmospheric boundary layer broadens: the URE will tend to be proportionally underestimated with increasing values in the average setup-elevation difference (Δh). This view is clearly consistent with the theoretical formulation of the thermal gradient under unstable conditions, which indicates both a significant height dependence for the curvature in the gradient and an increasing curvature toward the ground surface (Shaw and Smietana, 1983, p. 10,488), much as shown in figure 1. Regardless, as we argue here and has been demonstrated elsewhere (Castle and others, 1983; Mark and others, 1987), currently invoked theoretical models of the URE are equally or more inadequate predictive tools when dealing with procedurally constrained leveling. The only question that remains is whether more rigorous theoretical modeling or improved experimentation will contribute to more reliable estimates of the probable URE associated with any specific second- or third-order leveling.

EFFECTIVENESS OF NORMALLY APPLIED PROCEDURAL CONSTRAINTS

Although both theory and fairly compelling experimental observations argue for the accumulation of potentially large URE over gently sloping terrain, evidence of the presence of significant URE in procedurally constrained (*geodetic*) leveling is generally nonexistent. This is explicitly true in the case of first-order leveling (Mark and others, 1987); it may be less explicitly true, but commonly so as it applies to lower-order leveling as well (Gilmore, 1986; Castle and others, 1987). The reason that these experience-based procedural constraints have proved so effective is somewhat uncertain, but we have for the first time a soundly developed empirical basis for speculating as to why these constraints generally work.

Consider the observing conditions along a gently sloping line where the clarity of the image on the rod is so obscure—that is, the scintillation is so great—that the observer cannot accept the readings without at the same time ensuring that he probably will be unable to meet one or more of the several rejection criteria. Both theory and experience indicate that the clarity of the image will improve and the likely accumulation of any URE will diminish with reduced sight lengths. That is, as sight lengths are reduced the scintillation will diminish owing to two considerations: (1) shorter sight lengths are intrinsically compatible with reduced scintillation, and (2) retention of longer sight lengths should have required that the ray paths travel through a proportionately greater section of the boundary layer close to the ground—through just that region, in other words, where both the scintillation and the curvature in the thermal gradient increase sharply, such that the potential for any URE could be expected to increase equally sharply (see section on "The Gaithersburg-Tucson Test Results"). To state the same conclusion in somewhat different terms, any reduction in sight length will automatically reduce Δh as well, thereby compelling the sampling of a less *differentially* refractive atmosphere in foresight and backsight observations and a corresponding reduction in any likely URE.

The results of the Saugus-Palmdale experiment indicate that the accumulation of any URE will be damped to zero provided that the scintillation (σ_{DDH}) does not exceed 0.18 mm, a value that emerges from a survey run—even if inadvertently—to something less than first-order, class II standards (see section on "The Saugus-Palmdale Test Results"). This zero accumulation of URE probably is attributable chiefly to the impact of unassessed variables that tend to diminish or randomize any likely URE. The 0.18-mm scintillation value is not a magic number; it is, however, derived from more-or-less rigorously imposed procedural constraints and is apparently above the σ_{DDH} values commonly associated with first-order leveling (table 6). Moreover, if the scintillation exceeds 0.18 mm—or whatever level the observer associates with the maximum acceptable value for first-order leveling—it is likely that he will either reject the readings or shorten sight lengths (or even cease operations), regardless of the actual order of the survey. To do otherwise would require that he be able to discriminate among acceptable levels of scintillation as a function of survey order, a highly subjective if not almost impossible task. While the Saugus-Palmdale LSL test results seem to challenge this generalization (fig. 11), it is especially likely that the number of setup rejections—a number that remains unknown, since no record was kept—was a good deal greater than the 5-percent figure stipulated by J.D. Bossler. Finally, if the survey were being conducted at temperatures well below the threshold temperature—as defined for any given value for Δh—the observer would have considerable latitude in his choice of sight lengths with little fear of accumulating significant URE. This might be considered an unsupportable view were it not for the fact that most observers are very sensitive to the effects of temperature in meeting the rejection criteria.

EXPANDED EXPERIMENTAL INVESTIGATION

Mark and others (1987, p. 2788) observed that a number of variables which may impinge on the URE have not been considered in any theoretical model, chiefly because they are so difficult to quantify. By way of illustration, because the theoretical sensitivity of the URE to instrument height is such that it is reduced by about 12 percent simply through raising the instrument from 1.5 to 1.6 m (Shaw and Smietana, 1983, p. 10,489), one can easily imagine any number of unassessed variables that may play a role in limiting the URE in procedurally constrained surveys. Accordingly, we are inclined to believe that improved estimates of URE are much more apt to emerge from more intensive and carefully implemented experimental studies than through attempts to develop more sophisticated theoretical models.

A PROPOSED EXPERIMENT

The observatory concept invoked in the Gaithersburg-Tucson refraction tests is thought to be the most reliable approach to defining the URE as a function of survey order and locale; it incorporates the explicit advantage of model independence, minimizes the effects of unassessed variables, and, as a result, provides much more tightly controlled results than could otherwise be obtained. However, neither the Gaithersburg-Tucson results nor, as we show above, those obtained from the Saugus-Palmdale and Canadian field tests provide the definitive constraints needed to accurately estimate the URE identified with lower-order levelings in particular. The regression results developed from the Saugus-Palmdale experiment are diminished by the requirement that we assume a linear relation between the average setup-elevation difference (Δh) and the URE. Supplementary experimental data that would tend to mitigate this requirement should lead to improved estimates of the probable URE associated with any specific survey. In other words, additional experimental observations that could define the extent to which such routinely measured parameters as L, Δh, and t together control the accumulation of the URE are certainly desirable, if not imperative. With these considerations in mind, we believe that a broadened set of experiments identical in concept to those conducted at Gaithersburg and Tucson offers the best hope in attempting to more accurately define the magnitude of the URE associated with procedurally constrained 20th-century levelings.

The two sets of measurements that clearly demand refined definition are the URE as functions of temperature (t) and intermediate (as opposed to worst-case) values of the average setup-elevation difference (Δh). Probably less necessary, but still potentially valuable, are measurements based on greater sight lengths than those used in connection with the original Gaithersburg-Tucson experiments. Referring again to figure 2, we suggest the establishment of additional bench marks at nominal rod-reading heights of 0.83, 1.16, 1.83 and 2.16 m, in addition to those at 0.50, 1.50 and 2.50 m, at each of the designated sight lengths. We suggest as well the addition of marks 75 m distant from the instrument station, all intended to match the seven recommended nominal rod-reading marks associated with the shorter sight lengths. The finally adopted procedures for the creation of the near error-free datums at both Gaithersburg and Tucson (Whalen, 1981) should be adhered to in concept, but improved in their implementation. Specifically, while any systematic error in the Tucson test results probably was modest, the Gaithersburg results were almost certainly contaminated by relatively large systematic error(s)—conceivably large enough that the Gaithersburg URE determinations may be entirely artifact. In order to ensure that the standards are accurately determined, two procedural changes are encouraged. (1) The test-range marks should be established at least months in advance of any observations. A good deal of frictional heat and probably a degree of elastic strain is generated in driving the bench-mark rods to refusal. By waiting several months both this heat and any elastic strain should be largely dissipated, whereby any continuing contraction or expansion of the rods and resulting changes in the elevation differences would be minimized or eliminated. (2) To further minimize systematic error associated with changes in the standards, the standards should be established both immediately preceding and immediately following each set of observations. Should the successively determined-elevation differences among each of the sets of bench marks consistently hold to within 0.02 to 0.03 mm over any given period of observation, the mean of the before and after values should provide reference elevation differences no more than negligibly contaminated by error.

Acceptance of the modifications proposed above would require, among other things, that the number of observations jump from 12 (or 24 if each double-scale reading is counted as two measurements) to 32 with each sweep through the experimental array. Thus, simply on this basis alone, the costs associated with the implementation of the expanded experiment could be expected to escalate sharply—although certainly by less than a factor directly proportional to the additional rod readings.

Improved and enhanced temperature measurements, well removed from the instrument station, and direct measurements of heat flux should also be considered were a new observatory-type experiment actually implemented. If only because the ambient air temperature seems to operate as an index of heat flux—and, thus, both $\Delta t'$ and the total boundary layer environment—it should be automatically monitored at 1.5 m at two widely separated points within the test site. Moreover, because it is the incremental differences in temperatures at successive levels above the ground that explicitly control the magnitude of any URE, improved accuracy and independent datalog monitoring of temperature differences over the 0.5- to 2.5-m range should be attempted, perhaps through use of a differential thermocouple system. Even if this system were to drift off temperature, if the accuracy in the measurement of the temperature differences were maintained to within a few hundredths of a degree, it would produce dramatically improved estimates of both $\Delta t'$ and, especially, the curvature in the thermal gradient. In this same connection, at least one sensor should be added below a point midway between the 0.5- and 1.5-m sensors used at Gaithersburg and Tucson. Finally, even with more accurate temperature measurements, direct measurements of heat flux would provide a useful element of redundancy and, at the same time, remove much of the guesswork associated with our assessment of the impact of environmental factors on the accumulation of URE.

STRATEGY

Measurements obtained from only two sites probably form an inadequate basis for extrapolating to real-world situations. Accordingly, the addition of at least one site typical of eastern or northwestern North American leveling routes is especially desirable. For example, an eastern site characterized by grasses and weeds mixed with bare ground or gravel would be more typical of those conditions found along gravel shoulders or service roads adjacent to railways or highways. Nevertheless, and even though control levelings are rarely propagated through golf courses, we would strongly endorse the retention of a Gaithersburg- or Corbin-like site. Sites such as these represent near-limiting cases of the degree to which environmental factors may mitigate the accumulation of URE. A determined effort should be made to ensure that both varied and representative periods of occupation be selected for each of the several arrays. For example, it would be fatuous even to attempt to make measurements at Tucson in August simply because the image on the rod could be expected to be fluctuating so rapidly that the observer would have trouble in reading the graduations at even the shortest sight lengths. For quite the opposite reason, measurements at Gaithersburg in January or February probably should be dismissed from consideration. That is, owing both to low solar heating (low sun angle, cloud cover) and the loss of large fractions of any radiant energy to latent heat associated with surface evaporation and (probably) melting, we should reasonably expect small values of sensible heat flux and minimal refraction effects during the winter months. Thus, separated periods of observation in December, February and March or April could be expected to produce a range of representative readings at Tucson. At Gaithersburg, Corbin, or some nearby substitute site, one might reasonably select similarly short, successive periods of occupation in March, May, and June or August, September, and October. The additional logistical costs imposed by the deliberate separation of the periods of occupation would be more than compensated by the much greater applicability of the experimental results.

Both the directly emergent measurements and computed values obtained from the expanded experiment (fig. 28) suggest a variety of regressions explicitly applicable to estimates of the URE associated with any specified leveling. For example, simple regression of the URE on σ_{DDH} would produce a family of curves—of a form comparable to that shown in figure 10—where each curve is identified with a specified Δh. Results drawn from regression of σ_{DDH} on both t and L might be especially valuable, since they could provide a link relating the URE directly to these routinely tabulated parameters. Additional and unspecified regression analyses should also be attempted; it is only through experimentation with this sort of analysis that one might discover those relations that constitute the most reliable vehicles for estimating the URE. At worst, the expanded experiment would produce 48 individual data sets, whereby one could estimate the URE associated with any particular leveling through simple interpolation among routinely determined values of L, Δh, and t. Even though our present capabilities permit limited estimates of the likely contamination of any given leveling by URE, a refined experimental investigation would put much tighter limits on the estimated URE than the modeled values which consider so few of the contributing variables that tend to either enhance or damp this error.

CONCLUSION

The results of the three experiments described in this report place fairly well defined limits on the maximum accumulation of the URE as a function of survey order and, less certainly, the ambient air temperature and locale. These limits narrow progressively with progressively higher-order levelings. The procedures in force for first-order leveling throughout the 20th century are such that these surveys must have rarely

URE IN MILLIMETERS
L, IN METERS →

Δh, IN METERS ↓	30	45	60	75
0.00	URE_1	URE_2	URE_3	URE_4
0.67	URE_5	URE_6	URE_7	URE_8
1.33	URE_9	URE_{10}	URE_{11}	URE_{12}
2.00	URE_{13}	URE_{14}	URE_{15}	URE_{16}

t, IN DEGREES CELSIUS
L, IN METERS →

Δh, IN METERS ↓	30	45	60	75
0.00	t_1	t_2	t_3	t_4
0.67	t_5	t_6	t_7	t_8
1.33	t_9	t_{10}	t_{11}	t_{12}
2.00	t_{13}	t_{14}	t_{15}	t_{16}

σ_{DDH}, IN MILLIMETERS
L, IN METERS →

Δh, IN METERS ↓	30	45	60	75
0.00	σ_{DDH} (1)	σ_{DDH} (2)	σ_{DDH} (3)	σ_{DDH} (4)
0.67	σ_{DDH} (5)	σ_{DDH} (6)	σ_{DDH} (7)	σ_{DDH} (8)
1.33	σ_{DDH} (9)	σ_{DDH} (10)	σ_{DDH} (11)	σ_{DDH} (12)
2.00	σ_{DDH} (13)	σ_{DDH} (14)	σ_{DDH} (15)	σ_{DDH} (16)

Figure 28. Suggested format for tabulation of results obtained from expanded experiment at one site for each of 16 simulated test lines over a simulated height difference of ~500 m for each of 16 lines.

admitted URE in excess of the predicted random error, regardless of slope. Similarly, where the setup-elevation difference was ≤1.2 m and the average air temperature associated with any leveling (of whatever order) in a semi-arid to temperate environment remained at or below about 23°C, the URE probably did not even begin to compete with random error. In special cases, such as leveling directly over an asphalt highway or through a richly vegetated terrain, these threshold temperatures may drop to well below or increase to significantly above 23°C.

The greatest uncertainty surrounds the systematic accumulation of refraction error associated with second- and third-order levelings. That is, over a height difference of 500 m, accumulation of URE in any second-order survey in excess of 50 mm or in excess of 100 mm in any third-order survey must have been extremely uncommon. Nevertheless, we are at present unable to estimate with any confidence expected intermediate values of the URE for these lower-order surveys, where the temperatures rose above some threshold value that might be as low as 20°C—or even less. It is largely for this reason that we strongly encourage additional experimental observations modeled on the Gaithersburg-Tucson refraction tests of the National Geodetic Survey.

REFERENCES CITED

Bowie, W., and Avers, H.G., 1914, Fourth general adjustment of the precise level net in the United States and the resulting standard elevations: U.S. Coast and Geodetic Survey Special Publication 18, Washington, D.C., 328 p.

Castle, R.O., Brown, B.W., Jr., Gilmore, T.D., Mark, R.K., and Wilson, R.C., 1983, An examination of the southern California field test for the systematic accumulation of the optical refraction error in geodetic leveling: Geophysical Research Letters, v. 10, p. 1081–1084.

Castle, R.O., Elliott, M.R., and Gilmore, T.D., 1987, An early-20th-century uplift in southern California: U.S. Geological Survey Professional Paper 1362, 70 p.

Castle, R.O., Gilmore, T.D., Mark, R.K., and Shaw, R.H., 1985, Empirical estimates of cumulative refraction errors associated with procedurally constrained levelings based on the Gaithersburg-Tucson refraction tests of the National Geodetic Survey: Geophysical Research Letters, v. 12, p. 239–242.

Craymer, R.M., and Vanícek, P., 1986, Further analysis of the 1981 southern California field test for leveling refraction: Journal of Geophysical Research, v. 91, p. 9045–9055.

Federal Geodetic Control Committee, 1974, Classification, standards of accuracy, and general specifications of geodetic control surveys (1977 reprint): National Oceanic and Atmospheric Administration, Rockville, Maryland., 12 p.

———1975, Specifications to support classification, standards of accuracy, and general specifications of geodetic control surveys (1977 reprint): National Oceanic and Atmospheric Administration, Rockville, Maryland., 30 p.

———1980, Specifications to support classification, standards of accuracy, and general specifications of geodetic control surveys (revised): National Oceanic and Atmospheric Administration, Rockville, Maryland, 46 p.

———1984, Standards and specifications for geodetic control networks:National Geodetic Survey, Rockville, Maryland, various p.

Gilmore, T.D., 1986, Historic vertical displacements in the Salton Trough and adjacent parts of southeastern California: U.S. Geological Survey Open-File Report 86-380, 44 p.

Hayford, J.F., 1904, Precise leveling in the United States 1900-1903, with a readjustment of the level net and resulting elevations: U.S. Coast and Geodetic Survey, Report for 1903, Appendix 3, p. 195–809.

Heer, Rainer, and Niemeir, Wolfgang, 1985, Theoretical models, practical experiments and the numerical evaluation of refraction effects on geodetic leveling, *in* D.B. Zilkoski, ed. Proceedings, 3rd International Symposium on North American Vertical Datum: National Geodetic Information Center (NOAA), Rockville, Maryland, p. 321–342.

Heroux, P., Gale, W., and Faucher, G., 1985, Field test on the systematic effect of refraction in precise leveling, *in* Zilkoski, D.B., ed., Proceedings 3rd International Symposium on North American Vertical Datum: National Geodetic Information Center (NOAA), Rockville, Maryland, p. 311–320.

Holdahl, S.R., 1982, Recomputation of vertical crustal motions near Palmdale, California, 1959–1975: Journal of Geophysical Research, v. 87, p. 9374–9388.

Kukkamäki, T.J., 1938, Über die nivellitische refraktion: Publication no. 25, Finnish Geodetic institute, Helsinki, 48 p.

Mark, R.K., Gilmore, T.D., and Castle, R.O., 1987, Evidence of suppression of the unequal refraction error in geodetic leveling: Journal of Geophysical Research, v. 92, p. 2767–2790.

Mark, R.K., Tinsley, J.C., III, Newman, E.B., Gilmore, T.D., and Castle, R.O., 1981, An assessment of the accuracy of the geodetic measurements that define the southern California uplift: Journal of Geophysical Research, v. 86, p. 2783–2808.

Rappleye, H.S., 1948, Manual of geodetic leveling: U.S. Coast and Geodetic Survey Special Publication 239, 94 p.

Schomaker, M.C., and Berry, R.M., 1981, Geodetic Leveling: NOAA Manual NOS NGS 3, National Oceanic and Atmospheric Administration, Washington, D.C., p. (1-1)–(5-27).

Shaw, R.H., and Smietana, P.J., Jr., 1983, Temperature stratification and refraction errors in geodetic leveling: Journal of Geophysical Research, v. 88, p. 10,485–10,494.

Stein, R.S., Whalen, C.T., Holdahl, S.R., Strange, W.R., and Thatcher, W., 1986, Saugus-Palmdale, California, field test for refraction error in historical leveling surveys: Journal of Geophysical Research, v. 91, p. 9031–9044.

Strange, W.E., 1981, The impact of refraction correction on leveling interpretations in southern California: Journal of Geophysical Research, v. 86, p. 2809–2824.

Teledyne Geotronics, 1981, Leveling survey data of the southern California uplift 1981 for U.S. Geological Survey, Rep. 14-08-0001-19847: U.S. Geological Survey, Menlo Park, California, various p.

Whalen, C.T., 1980, Refraction errors in leveling-NGS test results, *in* Gerard LaChapelle, ed., Proceedings, 2nd International Symposium on Problems Related to the Redefinition of North American Vertical Geodetic Networks, Ottawa, Ontario, Canada: Canadian Institute of Surveying, p. 757–782.

———1981, Results of leveling refraction tests by the National Geodetic Survey: NOAA Technical Report NOS 92 NGS 22, National Oceanic and Atmospheric Administration, Washington, D.C., 20 p.

Whalen, C.T., and Strange, W.E, 1983, The 1981 Saugus to Palmdale, California, leveling refraction test: NOAA Technical Report NOS 98 NGS 27, National Oceanic and Atmospheric Administration, Washington, D.C., 13 p.

Wilson, R.C., 1983, Abstract, Highway 46 from east of Shandon at 144 FMK east to C 285, for U.S. Geological Survey, Rep. 14-08-0001-21350, U.S. Geological Survey, Menlo Park, California, various p.

☆U.S. GOVERNMENT PRINTING OFFICE: 1994-683-048/10007

SELECTED SERIES OF U.S. GEOLOGICAL SURVEY PUBLICATIONS

Periodicals

Earthquakes & Volcanoes (issued bimonthly).
Preliminary Determination of Epicenters (issued monthly).

Technical Books and Reports

Professional Papers are mainly comprehensive scientific reports of wide and lasting interest and importance to professional scientists and engineers. Included are reports on the results of resource studies and of topographic, hydrologic, and geologic investigations. They also include collections of related papers addressing different aspects of a single scientific topic.

Bulletins contain significant data and interpretations that are of lasting scientific interest but are generally more limited in scope or geographic coverage than Professional Papers. They include the results of resource studies and of geologic and topographic investigations, as well as collections of short papers related to a specific topic.

Water-Supply Papers are comprehensive reports that present significant interpretive results of hydrologic investigations of wide interest to professional geologists, hydrologists, and engineers. The series covers investigations in all phases of hydrology, including hydrogeology, availability of water, quality of water, and use of water.

Circulars present administrative information or important scientific information of wide popular interest in a format designed for distribution at no cost to the public. Information is usually of short-term interest.

Water-Resource Investigations Reports are papers of an interpretive nature made available to the public outside the formal USGS publications series. Copies are reproduced on request unlike formal USGS publications, and they are also available for public inspection at depositories indicated in USGS catalogs.

Open-File Reports include unpublished manuscript reports, maps, and other material that are made available for public consultation at depositories. They are a nonpermanent form of publication that may be cited in other publications as sources of information.

Maps

Geologic Quadrangle Maps are multicolor geologic maps on topographic bases in 7 1/2- or 15-minute quadrangle formats (scales mainly 1:24,000 or 1:62,500) showing bedrock, surficial, or engineering geology. Maps generally include brief texts; some maps include structure and columnar sections only.

Geophysical Investigations Maps are on topographic or planimetric bases at various scales; they show results of surveys using geophysical techniques, such as gravity, magnetic, seismic, or radioactivity, which reflect subsurface structures that are of economic or geologic significance. Many maps include correlations with the geology.

Miscellaneous Investigations Series Maps are on planimetric or topographic bases of regular and irregular areas at various scales; they present a wide variety of format and subject matter. The series also includes 7 1/2-minute quadrangle photogeologic maps on planimetric bases that show geology as interpreted from aerial photographs. Series also includes maps of Mars and the Moon.

Coal Investigations Maps are geologic maps on topographic or planimetric bases at various scales showing bedrock or surficial geology, stratigraphy, and structural relations in certain coal-resource areas.

Oil and Gas Investigations Charts show stratigraphic information for certain oil and gas fields and other areas having petroleum potential.

Miscellaneous Field Studies Maps are multicolor or black-and-white maps on topographic or planimetric bases on quadrangle or irregular areas at various scales. Pre-1971 maps show bedrock geology in relation to specific mining or mineral-deposit problems; post-1971 maps are primarily black-and-white maps on various subjects, such as environmental studies or wilderness mineral investigations.

Hydrologic Investigations Atlases are multicolor or black-and-white maps on topographic or planimetric bases presenting a wide range of geohydrologic data of both regular and irregular areas; principal scale is 1:24,000, and regional studies are at 1:250,000 scale or smaller.

Catalogs

Permanent catalogs, as well as some others, giving comprehensive listings of U.S. Geological Survey publications are available under the conditions indicated below from the U.S. Geological Survey, Books and Open-File Reports Sales, Federal Center, Box 25286, Denver, CO 80225. (See latest Price and Availability List.)

"Publications of the Geological Survey, 1879-1961" may be purchased by mail and over the counter in paperback book form and as a set of microfiche.

"Publications of the Geological Survey, 1962-1970" may be purchased by mail and over the counter in paperback book form and as a set of microfiche.

"Publications of the Geological Survey, 1971-1981" may be purchased by mail and over the counter in paperback book form (two volumes, publications listing and index) and as a set of microfiche.

Supplements for 1982, 1983, 1984, 1985, 1986, and for subsequent years since the last permanent catalog may be purchased by mail and over the counter in paperback book form.

State catalogs, "List of U.S. Geological Survey Geologic and Water-Supply Reports and Maps For (State)," may be purchased by mail and over the counter in paperback booklet form only.

"Price and Availability List of U.S. Geological Survey Publications," issued annually, is available free of charge in paperback booklet form only.

Selected copies of a monthly catalog "New Publications of the U.S. Geological Survey" are available free of charge by mail or may be obtained over the counter in paperback booklet form only. Those wishing a free subscription to the monthly catalog "New Publications of the U.S. Geological Survey" should write to the U.S. Geological Survey, 582 National Center, Reston, VA 22092.

Note.--Prices of Government publications listed in older catalogs, announcements, and publications may be incorrect. Therefore, the prices charged may differ from the prices in catalogs, announcements, and publications.