Journal of Hydrology (2006) **328**, 432–452

available at www.sciencedirect.com

SCIENCE DIRECT®

journal homepage: www.elsevier.com/locate/jhydrol

Optimal sampling schemes for estimating mean snow water equivalents in stratified heterogeneous landscapes

Fred G.R. Watson *, Thor N. Anderson, Wendi B. Newman, Susan E. Alexander, Robert A. Garrott

California State University Monterey Bay, 100 Campus Center Seaside, CA 93955-8001, United States

Received 11 May 2004; received in revised form 21 December 2005; accepted 21 December 2005

KEYWORDS
Snow hydrology;
Nested sampling;
Hierarchical bootstrap;
Landsat;
Landscape;
Yellowstone;
Wildlife

Summary Snow was sampled across a heterogeneous landscape within 40 strata defined by time, elevation, vegetation, and solar radiation (i.e. terrain). The objective was to efficiently develop accurate estimates of the mean snowpack properties of each stratum, for later use in testing the accuracy of a snowpack simulation model. Sources of variance were assumed to be 'fixed' (i.e. predictable) or 'random' (i.e. unpredictable). A nested sampling design was used to minimize random effects relative to sampling cost, and theoretically shown to be optimal or near optimal. Within the strata that were sampled, time and elevation effects on snow water equivalent (SWE) were greater than random effects, but random effects were greater than vegetation and radiation effects. Random effects were greatest at small scales (<100 m).The results indicated that because of this large variance at small scales, seasonal snowpacks must be sampled very intensely (up to 54 cores per estimated mean) in order to yield confidence intervals for the mean that are narrow enough to determine vegetation and radiation effects on SWE; and thus also, to enable meaningful comparisons between models and data with respect to vegetation and radiation effects. The resulting data have low sampling variance, and thus are ideally suited for testing of snowpack simulation models.

Introduction

Landscape-scale spatial snowpack dynamics are poorly understood relative to other areas in snow hydrology (Tarboton et al., 2000). Most studies in snow hydrology are in some way driven by the need to estimate basin-wide runoff (e.g. Elder et al., 1991, 1998; Shook and Gray, 1996; Luce et al., 1998; Skaugen, 1999; Yang and Woo, 1999; Pomeroy et al., 2002) or to provide operational forecasting for snow-affected industries (Smith and O'Brien, 2001) or to improve climate forecasting (Liston et al., 1999; Greene et al., 1999; Bergengren et al., 2001; Essery, 2003). Studies relating to snow ecology at the landscape scale are relatively rare

* Corresponding author. Tel.: +1 831 582 4452.
E-mail address: fred_watson@csumb.edu (F.G.R. Watson).

0022-1694/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2005.12.032

and less well developed from the perspective of the hydrologist (e.g. Walsh et al., 1994). Snowpack dynamics in large heterogeneous landscapes are well known to be important ecological drivers (Halfpenny and Ozanne, 1989; Marchand, 1996), but this knowledge is only qualitative.

Snowpack depth and snow water equivalent (SWE) cannot be remotely sensed in large heterogeneous landscapes (Cline et al., 1998; Mote et al., 2003). Further, they are not well estimated by snow cover (Faria et al., 2000), requiring snow cover maps to be supplemented with direct SWE measurements in modelling studies (e.g. Blöschl et al., 1991). Thus, we rely on models to estimate or predict SWE at large scales. However, testing of these models is limited by a scarcity of spatial field measurements at scales larger than small research catchments (e.g. Anderton et al., 2004).

Our work addresses this limitation — that of inadequate field measurement of snowpack heterogeneity and its covariance with predictable sources of environmental heterogeneity. It is situated within a broader study that makes novel links between wildlife ecology and snow hydrology (Garrott et al., 2003), by understanding at fine spatial scales the mechanisms that govern the spatial distribution of large herbivores (Bergman et al., in press).

Aims and methodology

The underlying aim of the study was to provide accurate and precise validation data for a physically based deterministic spatial simulation model (Watson et al., in press). The comparisons are intended to be made between individual fixed-effects strata, where a fixed-effect stratum is defined as an area with uniform attributes such as can be specified a priori over large landscapes (e.g. vegetation, terrain, climate). This is an appropriate scale for comparison because finer scale comparisons between model and data would suffer unfairly from the very high random sampling variance of snowpacks at (*circa* 30 m) pixel scales (Anderton et al., 2004).

Another way of putting this is that no matter how well can map landscape parameters, and how well we can model their influence on snowpacks, there will always be some degree of random variation that cannot be mapped or modeled. This variation should be quantified, so that it maybe separated out from variation in model-data comparisons that may be truly attributable to mapping and modeling error.

The focus of this paper is thus the design and development a geostatistical sampling scheme for ensuring accuracy, precision, and efficiency of snowpack measurements within specific homogeneous strata of a heterogeneous landscape. The general methodology was to repeatedly measure the snowpack at numerous points, using a sampling design that allowed quantification of both random variance and different spatial scales and covariance with key landscape characteristics. A Federal Sampler was used for all snow measurements, yielding point estimates of SWE (m), snowpack depth (m), and snowpack density (kg/m^3).

Study area

The study area (Fig. 1) is the home range of the Madison–Firehole elk herd in Yellowstone National Park, Wyoming, USA (314 km^2, centered on 44°38′ N, 110°52′ W). This area corresponds to the home range of a non-migratory population of about 600 elk (*Cervus elaphus*), and is a critical component of the home range of about 3000 migratory bison (*Bison bison*) (Garrott et al., 2003). Elevation varies between 2030 and 2620 m. The terrain is mountainous with high, forested plateaus on the periphery, incised by canyons and valleys containing a mosaic of broad meadows, undulating forests, and geothermal basins. Most of the forest is lodgepole pine (*Pinus contorta*), much of which is regenerating after large fires in 1988. Previous attempts to characterize SWE variation in the Study Area with respect to elevation, terrain, and vegetation effects lacked a formal sampling design and quantification of error (Farnes and Romme, 1993).

Stratified nested design for sampling SWE

Fixed and random effects

It is well known that snow properties exhibit high spatial variability and that point measurements can be highly unrepresentative of areal means (Skaugen, 1999; Yang and Woo, 1999; Kuchment and Gelfan, 2001; Pomeroy et al., 2002). What is not known is the number and spatial arrangement of samples necessary to achieve specific confidence intervals for the mean. Therefore, considerable attention was given to the design and execution of an adequate sampling design. The majority of previous sampling designs have been primitive, relative to the complex spatial structure of snowpacks. The staple of the discipline has been the (sometimes randomly located) snow course, a usually linear transect of about 5–25 point measurements spaced between about 5 and 100 m apart (NRCS, undated; OMNR, 1997; Yang and Woo, 1999; Pomeroy et al., 2002). Such designs should not be used to estimate SWE variance because of underestimation due to spatial autocorrelation. Occasionally, variance estimates are reported from unstated sampling designs (Pomeroy et al., 1998a), which should also be avoided. Only recently have sophisticated sampling designs emerged that are appropriate for unbiased estimation of areal means and accurate characterization of sampling variance and covariance (e.g. Erxleben et al., 2002). Completely randomized sampling designs are also rare (e.g. Elder et al., 1989, 1991).

We assert that the SWE at a given point and time in the landscape can be described as the sum of both predictable, fixed effects and unpredictable, random effects. We use the term 'predictable' within the context of landscape-scale spatial modeling studies to refer to effects that correlate with covariates that are measurable at the landscape scale (e.g. using digital elevation models and remote sensing). We restricted our sampling design to four predictable influences: time, elevation, vegetation, and solar radiation (analyzed hierarchically in this order). These effects are well known to influence SWE (Elder et al., 1989, 1991). We ignored for the time being potentially predictable variation due to steep terrain effects (e.g. avalanches), large-scale wind effects, saturated soils, geothermal activity, and longwave radiation from surrounding cliffs (Olyphant, 1986). However, we were able to map the likely location of these

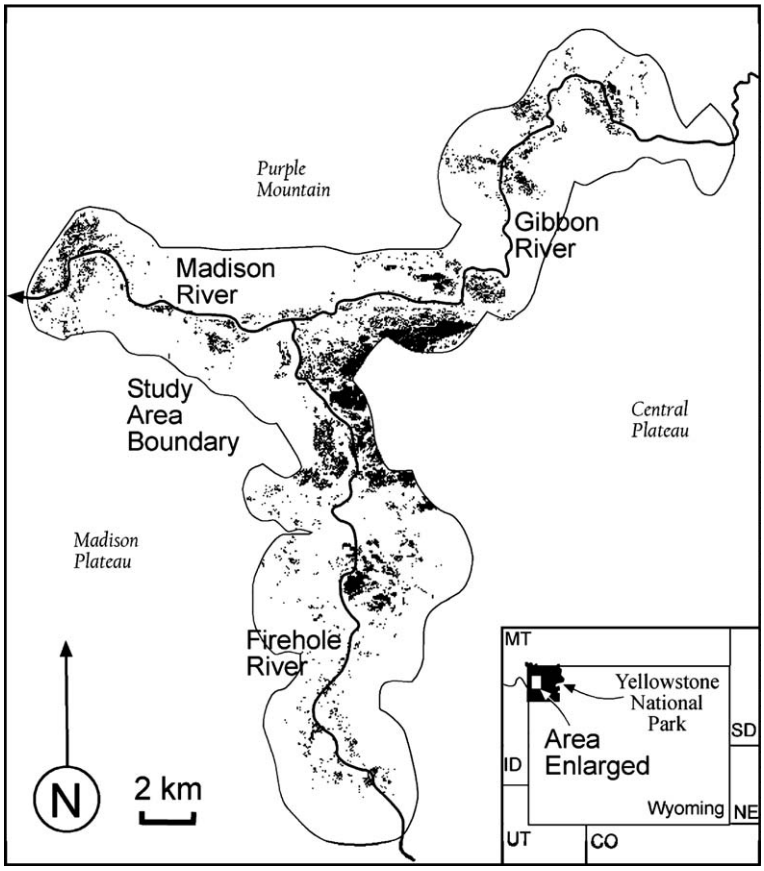

Figure 1 Study area in western central Yellowstone National Park, USA. Black areas illustrate the discontinuous spatial extent of one of the sampling strata—flat, unburned forest.

effects and define the sampling domain to avoid them. The term 'fixed stratum' is used hereafter as a collective term referring to all (possibly discontiguous) locations in the study area that have similar elevation, vegetation, and radiation and that might be sampled at approximately the same time.

All other influences were assumed to be 'random' and nested within the fixed hierarchy above. The word random is used in the sense that although these variables may have covariates, these are in turn assumed to be unbiased and unpredictable. An example is variation in snow depth due to topography, which appears to be dependent on scale — being unpredictable at small-scales (~40 m), and only partly predictable at medium-scales (~1000 m) (Lapen and Martz, 1996). There are no true random variables in spatial statistics (Isaacs and Srivastava, 1989), but many spatial variables can be regarded as random for all intents and purposes (Webster and Oliver, 2001). Apart from measurement error, examples of influences that were considered to be random include micro-topography, fallen logs, distance from hollows around tree trunks, and small-scale wind effects. In geostatistical terms, the variogram within a given fixed stratum was hypothesized to exhibit a well-defined sill at a range smaller than the geographic extent of the stratum.

Practicalities

A number of practical constraints influenced the sampling design. Sampling cost was proportional to separation distance. A single 'time' in the hierarchical design could span a maximum of 2—3 days. Strata were highly discontiguous in space, which limited the spatial patterns that could be used to sample them. Sampling could not be conducted in the same place at two different times (within a radius of about 10 m). These constraints were brought about by factors such as difficulty of access, probability of heavy snowfall, snowshoe compaction, and winter safety.

Design for sampling fixed effects

Fixed effects were strata were organized within a 4-stage hierarchy described by discrete ranges of time, elevation, vegetation, and solar radiation (denoted respectively by the letters *t*, *e*, *v*, and *r*) (Table 1, Figs. 2 and 3). The design was adaptively developed during 2001, but a fixed configuration was used during 2002.

Design for sampling random effects

Random spatial variation was sampled using a 5-stage nested sampling design with random starting points. Combined with the four fixed-effect stages, this yielded a 9-stage hierarchy. Nested analysis of variance (NANOVA, Sokal and Rohlf, 1981) was then used to quantify the relative magnitude of fixed effects and random effects at a range of spatial scales (see below). The five stages of the nested design embodied variation at five approximate scales: 1000 m, 250 m, 100 m, 10 m, and

Table 1 Definition of classes within systematic strata

Time		Elevation		Vegetation		Radiation	
Class	Range	Class	Range	Class	Description	Class	Description
T1	23–25 Feb 2001	E1	<2100 m	V1	Meadow	R1	Sunny (~20° slope, ~south-facing)
T2	3 March 2001	E2	2100–2300 m	V2	Burned forest	R2	Flat
T3	18-19 March 2001	E3	>2300 m	V3	Unburned forest	R3	Shady (~20° slope, ~north-facing)
T4	10–11 May 2001						
T5	19–21 Feb 2002						
T6	17–20 March 2002						
T7	23–25 April 2002						

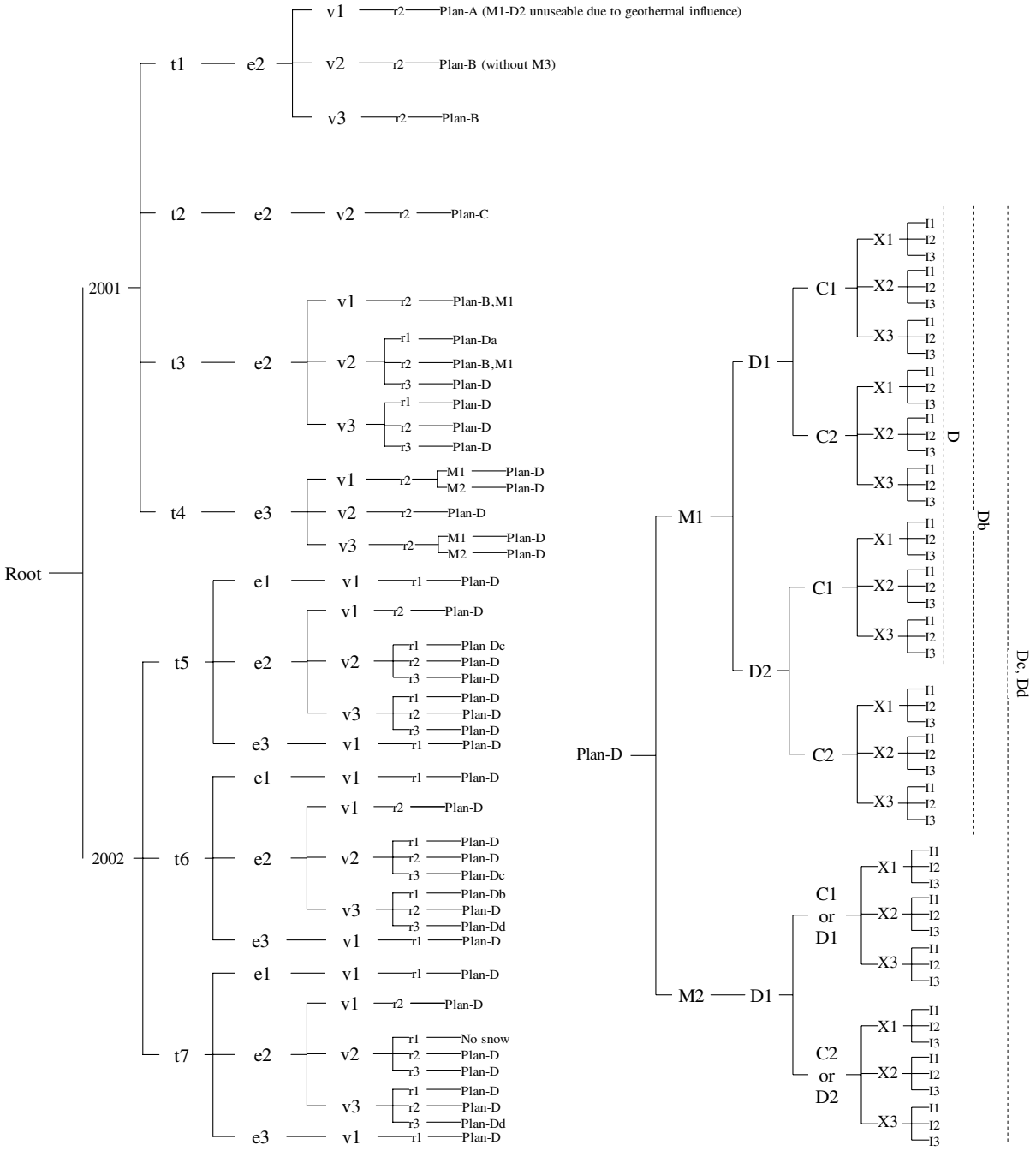

Figure 2 Nested random sampling designs. The letters M, D, C, X, V, I denote decreasing sampling scales.

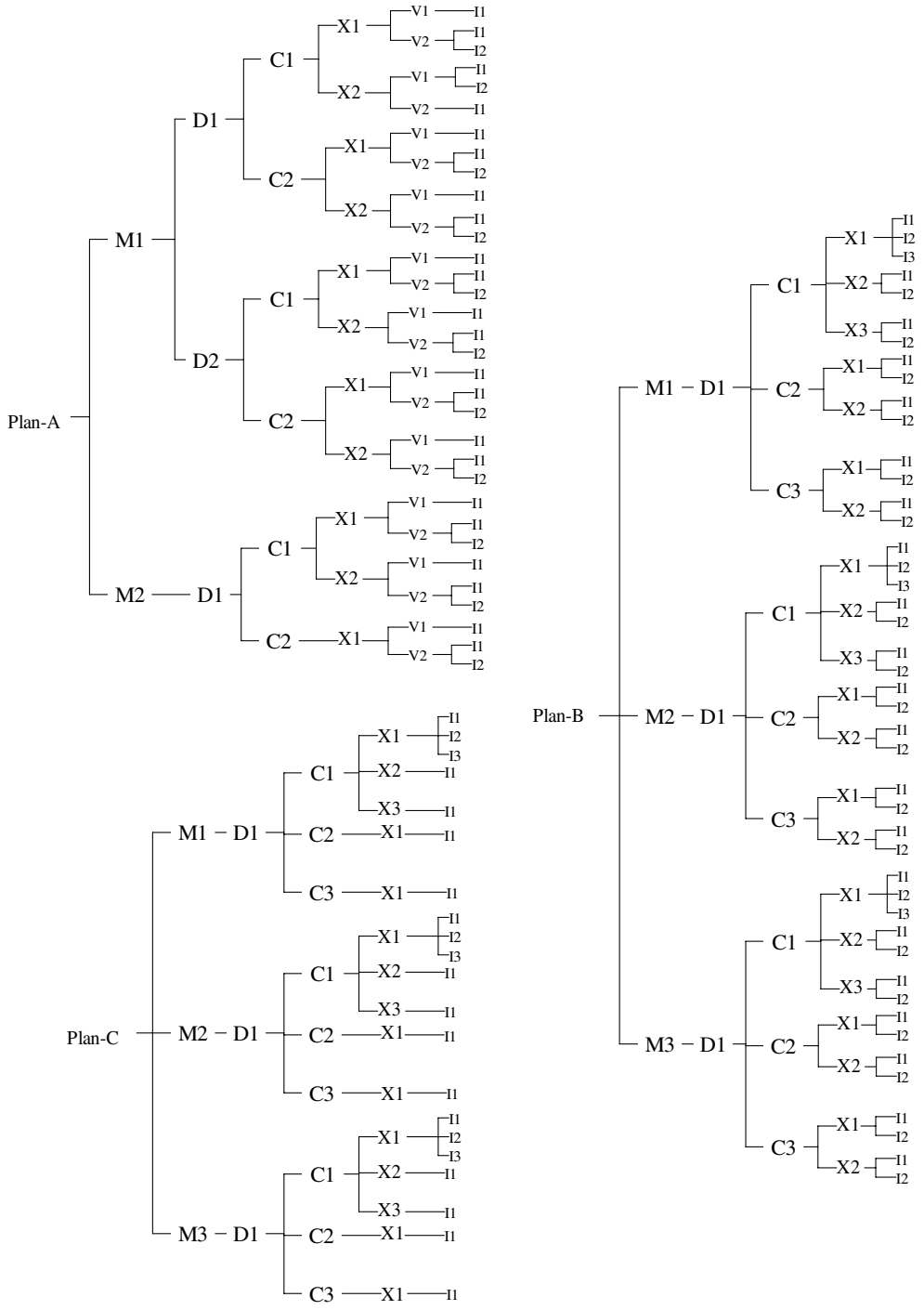

Figure 3 Fixed sampling design. The letters t, e, v, r denote specific fixed effects.

1 m (denoted by the letters M, D, C, X, I — taken from approximately corresponding roman numerals). In order to avoid over-sampling at small scales, the nesting was unbalanced — the number of subgroups sampled varied from group to group within a given stage (Figs. 2 and 3).

Early sampling events revealed that most variation within a systematic stratum occurred at length scales less than 100 m. Random variation at length scales longer than 300 m was usually negligible. These considerations ultimately favored a hybrid random-systematic sampling design within each fixed stratum that resembled a line of triangles (Fig. 4). The core of the design was a 10 m equilateral triangle with 1 m equilateral triangles at each corner. Sampling occurred at the 9 corners of these triangles, thus repeatedly quantifying spatial variability at 1 m and 9 m scales (Stages *I* and *X*). The triangular design was intended to minimize sampling bias due to anisotropy. Three 10 m triangles were then arranged so that two of them were approximately 100 m apart, and the third was approximately 250 m from the midpoint of the first two (Stages *C* and *D*) — a total of 27 cores.

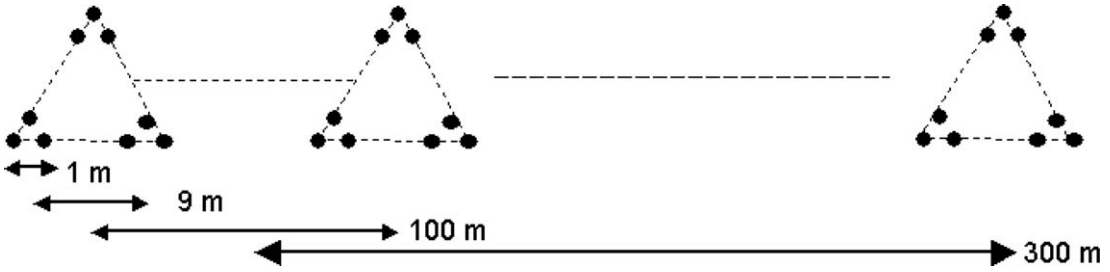

Figure 4 Schematic layout of typical late-study sampling design for random effects.

This design was adaptively augmented with additional groups of 9 cores whenever the mean SWE of the most recent 9 cores fell outside the range of the previous two means of 9-core groups. This was a precaution against the possibility of large-scale variation. However, a retrospective Monte Carlo bootstrap simulation on all stagewise deviances showed that adaptations beyond 27 cores actually only offered negligible reduction of the sampling error (0.2% at Stage *D*, and 2.0% at Stage *M*).

Randomness was considered in two ways. Firstly, it was assumed that at a given scale, the underlying spatial field was random – i.e. that from a given starting point, any point some distance away from that point was equally as representative as any other point *the same distance away*. Thus the effect of random sampling was achieved by systematic sampling of the random field (Lohr, 1999, p. 43) at the requisite scales. Secondly, the starting point for the entire design was chosen at random. Subjective bias was removed by using GPS receivers, laser range finders, tape measures, and sighting compasses to locate all coring sites.

Nested analysis of variance

NANOVA was used to understand the relative magnitude of random and fixed effects in the data, and to estimate a variogram for random variation (Miesch, 1975) that could be used to determine optimal sampling designs. Following Webster and Oliver (2001), based on (Youden and Mehlich, 1937), the general model for nested variation is based on the idea that a population can be divided into classes at a distinct number of stages to form a hierarchy. Each single observation embodies variation contributed from each level of the hierarchy. For a design with *M* fixed and *N* random stages, the most general model of variation is:

$$z_{f_1\ldots f_M s_1\ldots s_N} = \mu + F_{f_1} + \ldots + F_{f_1\ldots f_M} + S_{f_1\ldots f_M s_1} + \ldots + S_{f_1\ldots f_M s_1\ldots s_N} + \varepsilon_{f_1\ldots f_M s_1\ldots s_N} \quad (1)$$

where $z_{f_1\ldots f_M s_1\ldots s_N}$ is the value of the s_Nth unit in group $f_1\ldots f_M s_1\ldots s_{N-1}$ at stage $M+N-1$ in … the f_1th group at stage 1, and μ is the general mean. *F* and *S* are zero-mean variables denoting fixed environmental and independent random spatial effects respectively, and ε represents measurement error.

Assuming homogeneity of variance within fixed strata, (1) reduces to:

$$z_{f_1\ldots f_M} \approx \mu + F_{f_1} + \ldots + F_{f_1\ldots f_M} + S_{f_1\ldots f_M,1} + \ldots + S_{f_1\ldots f_M,N} \quad (2)$$

with population variances:

$$\sigma^2_{f_1}, \sigma^2_{f_1 f_2}, \ldots, \sigma^2_{f_1\ldots f_M}, \sigma^2_{f_1\ldots f_M,1}, \ldots, \sigma^2_{f_1\ldots f_M,N}$$

So the total random variance within a given fixed stratum is:

$$\sigma^2_{f_1\ldots f_M,\mathrm{RAND}} = \sum_{s=1}^{N} \sigma^2_{f_1\ldots f_M,s} \quad (3)$$

or within a given fixed stage, simply:

$$\sigma^2_{\mathrm{RAND}} = \sum_{s=1}^{N} \sigma^2_s \quad (4)$$

These variances are estimated as follows. At each stage *s*, the degrees of freedom, d_s, are $d_s = G^s - G^{s-1}$ (Gower, 1962), with G^s equal to the number of groups at the *s*th stage, and $G^0 = 1$ (note, *s* is a superscript here, not a power). The sums of squares, SS_s, and mean squares, MS_s, at each stage are:

$$\mathrm{SS}_s = \sum_{h=1}^{G^r} \sum_{i=1}^{g_h^{r,s}} c_{h,i}^{r,s} (\overline{z_{h,i}^{r,s}} - \overline{z_h^r})^2, \quad \mathrm{MS}_s = \mathrm{SS}_s / d_s \quad (5)$$

where G^r is the number of groups at the *r*th stage, with $r = s-1$; $g_h^{r,s}$ is the number of subgroups at stage *s* within the *h*th group at stage *r*; $c_{h,i}^{r,s}$ is the number of individual cores in the *i*th subgroup at stage *s* within the *h*th group at stage *r*; $\overline{z_h^r}$ is the mean value of all individual cores within the *h*th group at stage *r*; and $\overline{z_{h,i}^{r,s}}$ is the mean value of all individual cores within the *i*th subgroup at stage *s* within the *h*th group at stage *r*. Eq. (5) represents a direct computation of sums of squares, based on the more computationally efficient but less intuitive equations given by Sokal and Rohlf (1981).

Because of the unbalanced design, the estimated variances, $\hat{\sigma}_s^2$, at each stage are not simply the mean squares (as in normal ANOVA), but are recursively dependent on estimated variances from lower stages (Gower, 1962; Sokal and Rohlf, 1981; Webster and Oliver, 2001):

$$\hat{\sigma}_s^2 = \frac{\mathrm{MS}_s - \sum_{t=s+1}^{n} u_{s,t} \hat{\sigma}_t^2}{u_{s,s}}, \quad \hat{\sigma}_{N+1}^2 = 0 \quad (6)$$

$$U_{s,t} = \frac{1}{d_s} \left\{ \sum_{i=1}^{G^s} \sum_{j=1}^{g_i^{s,t}} \frac{(c_{i,j}^{s,t})^2}{c_i^s} - \sum_{i=1}^{G^{s-1}} \sum_{j=1}^{g_i^{s-1,t}} \frac{(c_{i,j}^{s,t})^2}{c_i^{s-1}} \right\} \quad (7)$$

where $u_{s,t}$ is the coefficient of variance at stage *s* for the *t*th component, with $s \leqslant t$, and c_i^s is the number of individual cores in the *i*th subgroup at stage *s*.

The study resulted in 1222 SWE samples (1058 non-zero snow cores) organized within 40 systematic strata spanning two winters. The variation in SWE measured within each stage of the sampling design is summarized in Table 2. and Figs. 5–7. Although the analysis was done in terms of variance, the data are plotted as standard deviations (i.e.

Table 2 Variance in SWE at each stage of the nested sampling design

Effect	Stage	Description	2001 data only			2002 data only			2001 and 2002 data combined		
			Variance	Fraction of total random (%)	Standard deviation (mm)	Variance	Fraction of total random (%)	Standard deviation (mm)	Variance	Fraction of total random (%)	Standard deviation (mm)
Fixed	T	Time	419.65		20.5	1743.05		41.7			
Fixed	E	Elevation	0.00		0.0	8947.66		94.6			
Fixed	V	Vegetation	0.00		0.0	175.27		13.2			
Fixed	R	Radiation	2174.10		46.6	76.33		8.7			
Random	MDCXI	Sum	1138.32	100.0	33.7	1263.21	100.0	35.5	1184.63	100.0	34.4
Random	M	700–7000 m	140.96	12.4	11.9	214.98	17.0	14.7	155.03	13.1	12.5
Random	D	250 m	7.40	0.7	2.7	0.00	0.0	0.0	3.28	0.3	1.8
Random	C	100 m	1.31	0.1	1.1	224.90	17.8	15.0	150.26	12.7	12.3
Random	X	9 m	447.12	39.3	21.1	199.84	15.8	14.1	278.31	23.5	16.7
Random	I	1 m	541.53	47.6	23.3	623.49	49.4	25.0	597.75	50.5	24.4

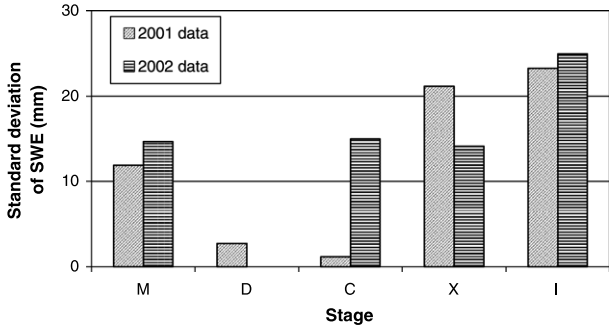

Figure 5 Random SWE variation at different spatial scales. The 2002 data were based on a consistent sampling design, with even representation from each of the 9 sampled strata. The 2001 data were biased by representing some fixed effects more than others (see Fig. 3). The variance at the M-stage may be overestimated due to the adaptive nature of the design.

square rooted) in order to have meaningful measurement units. Note that the estimated variation shown in Fig. 5 is *biased* toward systematic strata that were sampled more frequently — i.e. mid-season snowpacks (as opposed to late-season ones).

The small-scale (1-m, and point measurement error) component of the random variation is clearly the greatest. In the combined 2002 data, at least 83% of the random variance in SWE occurs at length scales of 100 meters or less. The 2002 data are superior because a consistent sampling design and set of fixed effects were used throughout. The 2001 data provide a useful comparison, agreeing with the 2002 data for the most part. In general, the analysis suggests that the most efficient sampling designs for quantifying mean SWE are those that frequently sample variation at short length scales.

The spatial variance structure varies between strata, as shown in Fig. 6 for the 2002 data. Variance structure is consistent during the accumulation phase (February and March data) and of course reduces to negligible variance in late melt out (April data). Burned forest has higher variance than unburned forest or meadow, particularly at the I-scale (probably due to fallen logs) and the C and D-scales (likely due to burn heterogeneity). Some large-scale (M) variation was observed in unburned forest, perhaps due to variation in forest cover. Meadows exhibited very low moderate scale (D) variation, and thus were not sampled at any larger scales (i.e. M-scale) in 2002 (based on the assumption that a spectral gap does not occur, i.e. a sill, followed by an increase in variance at scales larger than our largest sampling scale). Variance structure was consistent throughout the topographic radiation gradient. Anomalously high variation was observed across large-scales on shady slopes, which we interpret as a result of confounding vegetation effects as opposed to inherent random variation at this scale on shady slopes in general. It was difficult to find accessible shady slopes with homogeneous unburned or burned forest extending for long distances.

Fixed and random stages are compared in Fig. 7. As expected, time and elevation were the dominant fixed effects present in the data. Vegetation and radiation effects were approximately equal, but were much smaller than time,

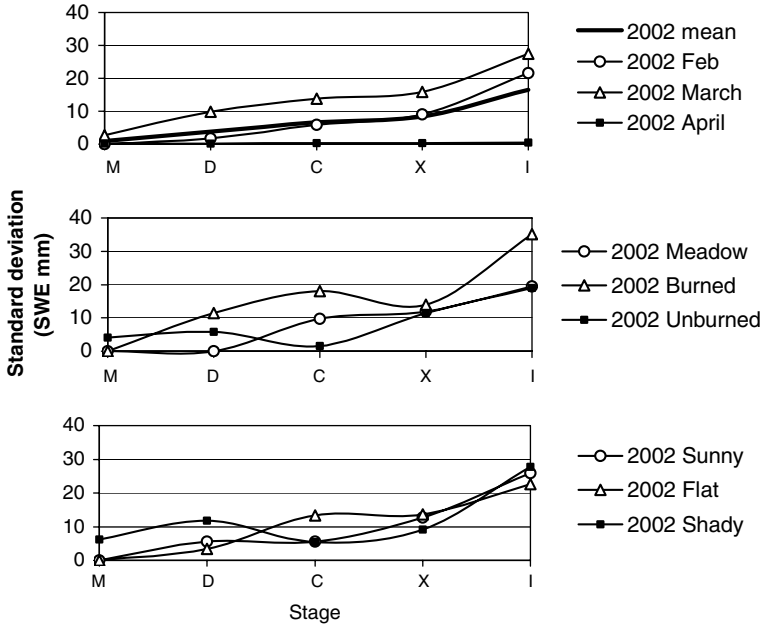

Figure 6 Random SWE variation at different spatial scales, broken down by time, vegetation, and radiation effects.

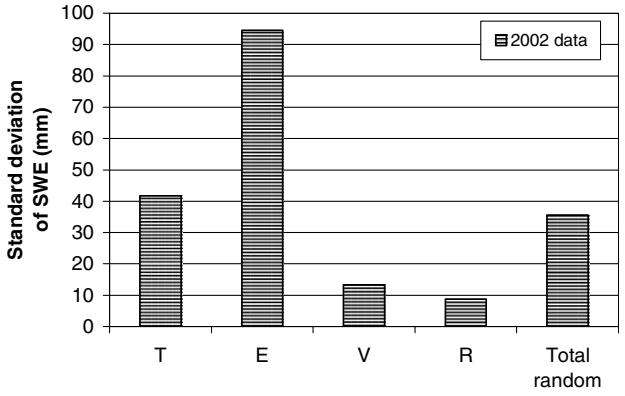

Figure 7 Fixed and random SWE variation. The standard deviation shown for fixed effects is based on a systematic sampling design, with no randomization. It is thus not a true indication of variance in the population. It is shown in order to give an approximate comparison to the magnitude of random effects. Only the 2002 data are shown, because the 2001 data did not sample the fixed strata consistently.

elevation, and random effects. This implies that vegetation and radiation effects, while thought to be important wildlife habitat selection parameters, can only be detected after a considerable amount of random sampling effort is directed toward narrowing the confidence intervals for random variation.

Estimation of spatial means

Ultimately, we desire an estimate of the mean SWE over the entire area corresponding to each fixed stratum of the sampling design, and of the confidence intervals of that estimate. Two estimators of the spatial mean suggest themselves: the mean of all samples; and the hierarchical mean based on the nested design — defined below. In balanced nested designs, the two are equivalent. However, in unbalanced designs, they differ, and we show that the superiority of one over the other depends on both the population characteristics and the sampling design. We illustrate this using a simple two-stage example of measurements y_{ij} sampled within n groups at a primary stage, p, and an independent secondary stage, s, comprising m_i measurements in each primary group $i = 1 \ldots n$. The variances of each stage, σ_p^2 and σ_s^2, are assumed to be known.

Perhaps the most common estimator is what we term here as the 'hierarchical' estimator (the 'unbiased' estimator of Thompson (2002, p. 145) and Lohr (1999, p. 147)):

$$\hat{\mu}_{\mathrm{H}} = \frac{1}{n}\sum_{i=1}^{n}\hat{\mu}_i, \quad \hat{\mu}_i = \frac{1}{m_i}\sum_{j=1}^{m_i} y_{ij} \tag{8}$$

An alternate estimator simply takes the mean of all measurements, as if they were not organized within a nested design at all. We term it the 'flat' estimator:

$$\hat{\mu}_{\mathrm{F}} = \frac{1}{m}\sum_{i=1}^{n}\sum_{j=1}^{m_i} y_{ij}, \quad m = \sum_{i=1}^{n} m_i \tag{9}$$

Assuming unbiasedness, the better of the two estimators is the one with the least estimated variance. In turn, the variance of the estimator depends on the degree of unbalancedness, and the proportion of the overall variance that is exhibited at lower rather than higher stages. This is shown as follows.

The expected variance of the weighted mean $\hat{\mu}_1$ of n independent realizations (each with weight, w_i) of a random variable with variance σ^2 is (Sokal and Rohlf, 1981, p. 135):

$$\mathrm{var}(\hat{\mu}) = \sigma^2 \sum_{i=1}^{n} w_i^2 \Big/ \left(\sum_{i=1}^{n} w_i\right)^2 \tag{10}$$

which for two independent, nested stages gives:

$$\mathrm{var}(\hat{\mu}) = \sigma_p^2 \sum_{i=1}^{n} w_i^2 \Big/ \left(\sum_{i=1}^{n} w_i\right)^2 + \sigma_s^2 \sum_{i=1}^{n}\sum_{j=1}^{m_i} w_{ij}^2 \Big/ \left(\sum_{i=1}^{n}\sum_{j=1}^{m_i} w_{ij}\right)^2 \quad (11)$$

$$\mathrm{var}(\hat{\mu}_H) = \sigma_p^2/n + \sigma_s^2 \sum_{i=1}^{n} \frac{1}{m_i} \Big/ n^2 \quad (12)$$

$$\mathrm{var}(\hat{\mu}_F) = \sigma_p^2 \sum_{i=1}^{n} m_i^2 \Big/ \left(\sum_{i=1}^{n} m_i\right)^2 + \sigma_s^2/m \quad (13)$$

since for the hierarchical estimator, $w_i = w = 1$, $w_{ij} = \overline{w_i}$, and $w_{ij} = \overline{w_i} = 1/m_i$; and for the flat estimator, $w_i = m_i$ and $w_{ij} = \overline{w_i} = 1$. The best estimator is determined not by the actual variances, but by their relative proportions, p_s, which we represent by making the substitutions:

$$p_s = \frac{\sigma_s^2}{\sigma^2}, \quad \sigma^2 = \sigma_p^2 + \sigma_s^2, \quad \sigma_s^2 = p_s\sigma^2, \quad \sigma_p^2 = (1-p_s)\sigma^2 \quad (14)$$

which, after equating $\mathrm{var}(\hat{\mu}_H) = \mathrm{var}(\hat{\mu}_F)$ leads to a criterion value of p_s that can be used to determine which estimator is best:

$$p_s^* = \left(\frac{\sum_{i=1}^{n} m_i^2}{m^2} - \frac{1}{n}\right) \Big/ \left(\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{m_i} - \frac{1}{n} - \frac{1}{m} + \frac{\sum_{i=1}^{n} m_i^2}{m^2}\right) \quad (15)$$

For the simplest unbalanced case where $n = 2$, $m_1 = 1$, $m_2 = 2$, and $m = 3$ we obtain $p_s^* = 57.1\%$. We confirmed this result using 23 repeats of 30,000 Monte Carlo simulations on two-stage normal distributions of varying p_s. These simulations also confirmed the earlier assumption of negligible bias, and thus that the estimators may be compared solely on the basis of their variances.

All of the sampling designs used in the present study were balanced in the lowest three stages (C, X, and I) (see Figs. 2 and 3). Since the two estimators are equivalent for balanced designs, the choice of estimator should thus be based on means computed within the unbalanced stages (M and D). Further, unbalancedness always conformed to the 'simplest unbalanced case' above. From the nested analysis of variance above, we estimate that at least $p_s = 83\%$ of the total random variance in SWE occurs at scales of 100 m or less. This is higher than the criterion ($p_s^* \cong 57\%$), which indicates that, for our purposes, and contrary to the customary use of nested data, the flat estimator is better than the hierarchical estimator, and was thus used forthwith.

Analysis of optimal sampling design

Contrary to the ideals of sampling, the optimal sampling design for the present study could not be identified *a priori* for practical reasons. This does not preclude, however, a retrospective analysis of the optimality of the design used in 2002. The optimal sampling design is that which minimizes the variance of the estimated mean SWE subject to practical constraints. Specifically, strata had to be sampled within 2.5–3 h each; and the time cost of visiting sites in a stratum was proportional to separation distance. This was summarized as a vector of estimated time costs (in minutes) associated with locating a site within each stage of the nested sampling design:

$$\mathbf{c} = [\,c_M \quad c_D \quad c_C \quad c_X \quad c_I\,] = [c_1 \cdots c_5] = [\,30 \quad 15 \quad 10 \quad 2 \quad 0.5\,] \quad (16)$$

These costs sum to approximately 1 h, which we would also estimate as the time required to core a single randomly located site in a non-nested design.

In the unconstrained, uncorrelated case with population variance σ^2 the optimal design simply uses a very large number, n_R, of randomly located samples, with the resulting estimated mean having variance σ^2/n_R and standard error $\sqrt{\sigma^2/n_R}$. Applying the time cost constraint, only two, or at most three samples could be taken using this design in 2.5 h. Assuming the population variance is equal to the total random variance, 1263 mm^2, from Table 2, the standard error of a mean estimated from three random samples is 20.5 mm. We would like to do better, considering that if the data were normal, this would imply a wide 95% confidence interval of ±40 mm (compare this with the standard deviation due to vegetation and/or radiation effects ≅10 mm, see Table 2).

Given that time cost increases with distance traveled, the easiest way to obtain more cores for a given cost is to deliberately take cores closer together. However, we know that SWE is spatially auto-correlated, so we would expect to under-estimate the variance if we used the simple random sampling equation above (although simple estimation of variances and standard deviations from potentially spatially auto-correlated data is not uncommon in the snow sampling literature, e.g. König and Sturm, 1998; Pomeroy et al., 2001). A systematic random design with strata placed farther apart than the correlation range would be less affected by auto-correlation, but would cost as much or more than a random design, as cores would be mandated to be a considerable distance apart. A nested design allows us to explicitly take account of auto-correlation and travel cost, using a sampling structure that can be optimized to minimize the estimated variance given knowledge at multiple scales of: (a) travel cost; and (b) variance exhibited at different scales (provided by the nested analysis of variance above). Note that the inferential cost of the nested approach is that it relies on an assumption of homogeneous spatial variance structure within each fixed-stratum.

Assuming a *balanced* nested design for the moment, the structure of a 5-stage nested design can be specified by a 5-element vector:

$$\mathbf{n} = [\,n_M \quad n_D \quad n_C \quad n_X \quad n_I\,] = [\,n_1 \quad \cdots \quad n_5\,]$$

where the elements of $\mathbf{n}$ are integers n_s corresponding to the number of groups at each random-effects stage within each respective parent group. The total number of groups at each random effects stage is then

$$\boldsymbol{\eta} = [\,\eta_M \quad \eta_D \quad \eta_C \quad \eta_X \quad \eta_I\,] = [\,\eta_1 \quad \cdots \quad \eta_5\,], \quad \eta_s = \prod_{j=1}^{s} n_j \quad (17)$$

Note that η_s is the same as g_h^{M+s} in the notation of Eq. 5, for a given fixed stratum h. For each of these designs, the total cost is: $C = \boldsymbol{\eta}\mathbf{c}^T$ and the variance and standard error of the estimated mean are:

$$\mathrm{var}(\hat{\mu}_{BAL}) = \sum_{s=1}^{5} \frac{\hat{\sigma}_s^2}{\eta_s} \quad \mathrm{SE}(\hat{\mu}_{BAL}) = \sqrt{\mathrm{var}(\hat{\mu}_{BAL})} \quad (18)$$

Only 253 designs satisfy $C \leqslant 180$ min, and of these, the design with minimum variance is (determined by exhaustive search): $\mathbf{n} = [2 \quad 1 \quad 2 \quad 3 \quad 4]$. This is a 48-core, 178-min design that is practically very similar to the balanced version of the design that was actually used (36-cores, 172-minutes): $\mathbf{n} = [1 \quad 2 \quad 2 \quad 3 \quad 3]$. These two designs have very similar estimated standard errors of 13.9 and 14.1 mm respectively. The balanced nested approach is thus an improvement over purely random sampling (and most likely also systematic random sampling).

Our standard second-year design was an *unbalanced* 27-core, 152-min version of the 36-core design above, where the first group at the second stage contained two third-stage groups, but the second group at the second stage only contained one third-stage group. It could be written as: $\mathbf{n} = [1 \quad 2 \quad 1.5 \quad 3 \quad 3]$, where the non-integer 1.5 denotes the unbalancedness between group sizes of 1 and 2. We did not do a global search of all possible unbalanced designs to find the optimum, but we did compare our chosen unbalanced design with the optimum balanced design, as follows.

The variance of estimated means from unbalanced designs generalizes from (**18**) for multi-stage designs as (for the *h*th group at Stage M):

$$\mathrm{var}(\hat{\mu}_{\mathrm{UNBAL}})_h^M = \sum_{s=1}^{5} \left[\sigma_s^2 \times \left(\sum_{i=1}^{g_h^{M,s}} w_{s,i}^2 \right) \div \left(\sum_{i=1}^{g_h^{M,s}} w_{s,i} \right)^2 \right] \qquad (19)$$

where $w_{s,i} = 1$ for the flat estimator of the mean, and $w_{s,i} = 1/g_h^{M,s}$ for the hierarchical estimator of the mean. Using the above equation, the estimated standard errors of the hierarchical and flat estimators respectively were, 15.6 and 15.1 mm. As expected, these are slightly worse than for their balanced counterpart with 9 extra cores. However, they are 30 minutes quicker. Indeed, if the absolute measure of optimality is given by minimizing the *product* of estimated cost and estimated standard error, then the unbalanced 27-core design is superior among all that have been discussed.

We thus reach the following conclusions: nested designs are desirable, if not, optimal designs for estimating the mean of a homogeneous, locally stationary spatial field; unbalanced designs can produce better estimates of the mean than balanced designs, with respect to minimization of estimated variance and estimated sampling cost; and the adaptive 27-core unbalanced design used in the present study was close to the optimal, if not, the optimal design given the stated practical constraints.

Empirical distribution functions

Confidence intervals were estimated for the estimates of the mean SWE of each fixed stratum. Beforehand, this required a characterization of the empirical distribution functions (EDFs) corresponding to the variation exhibited at each random effects stage. If the EDFs were normal, then classical parametric statistics — i.e. the Student's *t*-distribution — could be used to estimate confidence intervals. The starting point for compilation of the EDFs were the deviations of each group mean from its parent group mean:

$$\delta_{h,i}^{*r,s} = \overline{z_{h,i}^{r,s}} - \overline{z_h^r}, \quad s = r + 1 \qquad (20)$$

These deviations are not solely and exactly representative of the population variance σ_s^2 at their nominal stage, *s*. This is both because the group means $\overline{z_{h,i}^{r,s}}$ contain expected variance σ_{s+1}^2/n_{s+1} introduced from the stage below, and because the parent group mean $\overline{z_h^r}$ is itself a variable with expected variance σ_s^2/n_s. Specifically, within a single fixed stratum of a balanced design, the expected sample variance of the deviations is:

$$\frac{E(\mathrm{SS}_s)}{n_1 \times \ldots \times n_{s-1}(n_s - 1)} \cong \sigma_s^2 - \frac{\sigma_s^2}{n_s} + \frac{\sigma_{s+1}^2}{n_{s+1}}, \quad 1 \leqslant s \leqslant N \qquad (21)$$

with σ_s^2 estimated using Eq. (6), $\sigma_{N+1}^2 = 0$, and $\mathrm{SS}_s = \sum_i (\delta_{h,i}^{r,s})^2$ summed over all groups *i* at stage *s* within all parent groups *h* at stage *r* throughout the fixed stratum.

The naive deviations $\delta_{h,i}^{*r,s}$ can be scaled by a correction factor λ_s in order to make them representative of solely the variance at stage *s*. This factor (squared) is simply the ratio of some estimate of the variance at each stage $\hat{\sigma}_s^2$ to the actual (population) variance of the deviations $\mathrm{Var}_p(\delta_{h,i}^{r,s}) = \mathrm{SS}_s/\eta_s$:

$$\lambda^2 = \eta_s \hat{\sigma}_s^2 / \mathrm{SS}_s \qquad (22)$$

which leads to the corrected deviation:

$$\delta_{h,i}^{r,s} = \lambda_s \delta_{h,i}^{*r,s}, \quad s = r + 1. \qquad (23)$$

Because these steps lead to sets of sample deviations representing the variance at each stage of a nested design, they facilitate a general method for multistage balanced hierarchical non-parametric bootstrapping called for by Davison (1999) and paralleled by Carpenter et al. (1999). Eq. (21) appears to be novel in the general form given here. It can be seen as an extension of Davison and Hinkley (1997, Eqn 3.26 etc.), after re-writing it as a recursive estimate of $\hat{\sigma}_2^2$, substituting into (22), and noting that Davison and Hinkley ignored the inflation term, $-\sigma_s^2/n_s$, presumably by assuming large n_s.

As described earlier, $\hat{\sigma}_s^2 | f_1 \ldots f_M$ was estimated for all five random effects stages within each of the 40 fixed effects strata using (6) (40 NANOVA runs yielding a total of 200 variance estimates, and 200 sums of squares). These values were used to compute 5×40 correction factors, which are plotted in Fig. 8. The correction factors do not vary greatly between fixed strata. Therefore, a single set of *N* factors λ_s was used to inflate/deflate all 1228 deviations. This amounted to slight inflation of all deviations at Stage I, and deflation (or 'shrinkage') of all deviations at coarser stages: $\lambda_s = \{0.22, 0.31, 0.43, 0.51, 1.25\}$, $s = 1, \ldots, 5$.

The corrected deviations were pooled and compared in various ways in order to examine questions of homogeneity *between* fixed effects strata. Fig. 9 shows quantile—quantile (Q—Q) plots for each random effects stage, with deviations pooled across the entire sampling design. Normally distributed data would form a straight line of points on these plots. The solid, curved lines indicate 95% confidence limits for normality. The I-deviations and to a lesser extent the X-deviations are clearly kurtotic, and thus non-normal. The larger scale deviations also appear non-normal, but there are too few of them for this to be characterized as statistically significant. The Q—Q plots do not curve uniformly upwards, and thus show no evidence of being log-normally distributed or otherwise positively skewed, contrary to the findings of other authors (e.g. Shook, 1995), cited by Shook and Gray (1996).

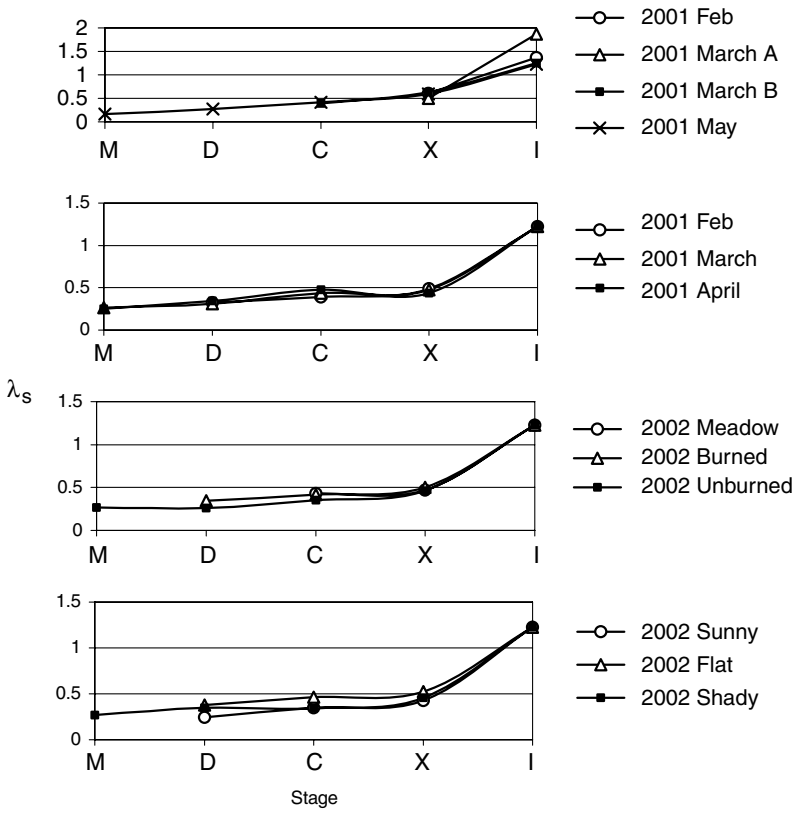

Figure 8 Correction factors (λ_s), for estimating unbiased deviations representing the variance specific to each stage — grouped and averaged by time (top), vegetation (middle), and radiation (bottom).

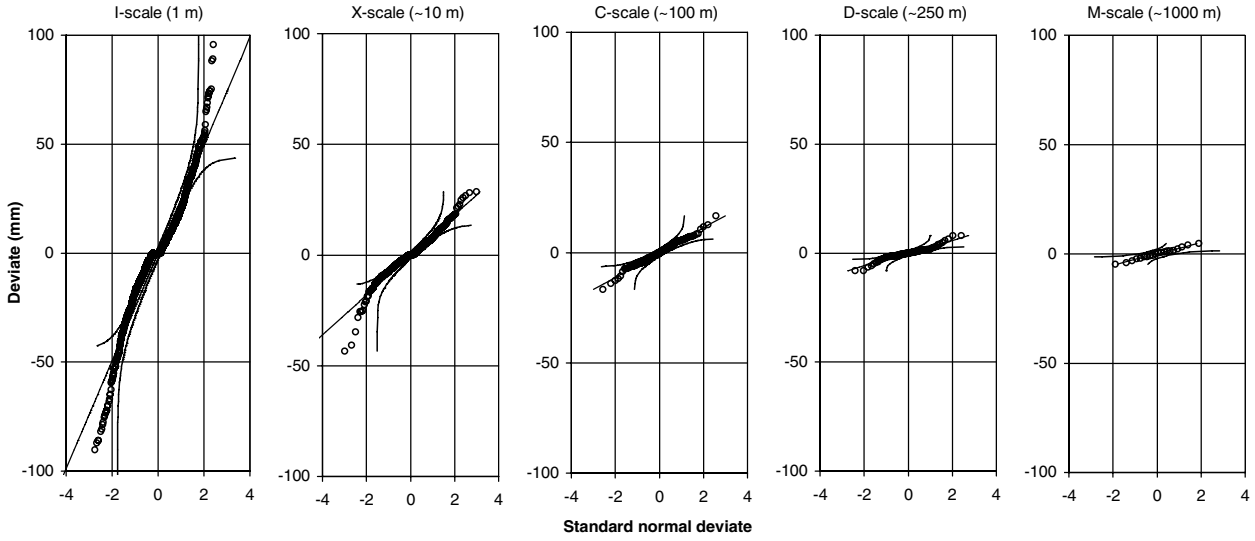

Figure 9 SWE deviations at each stage, corrected to exclude variance due to neighboring stages.

The corrected deviations are broken up into subsets grouped according to specific fixed effects in Fig. 10. Fine-scale variation in April 2002 is much more kurtotic than data from preceding months. This is because of patches of zero-snow extending for at least 1 m and often at least 10 m, leading to frequent zero-deviations at these scales. Fine-scale deviations in unburned forest differ markedly from those in burned forest and meadows, in ways not easily characterized in parametric terms. As a check of this interpretation, we performed M-tests for variance heterogeneity (Crow et al., 1960), which proved positive for vegetation-specific heterogeneity at I, C, and M scales (negative for X and D scales). Vegetation effects on variance are likely to be due to differing frequencies of fallen logs and other physical anomalies. Differences due to solar radiation are present, although more so at the X-scale than the I-scale. These effects are possibly due to interactions between solar radiation and vegetation, leading to differential shading of the snow surface

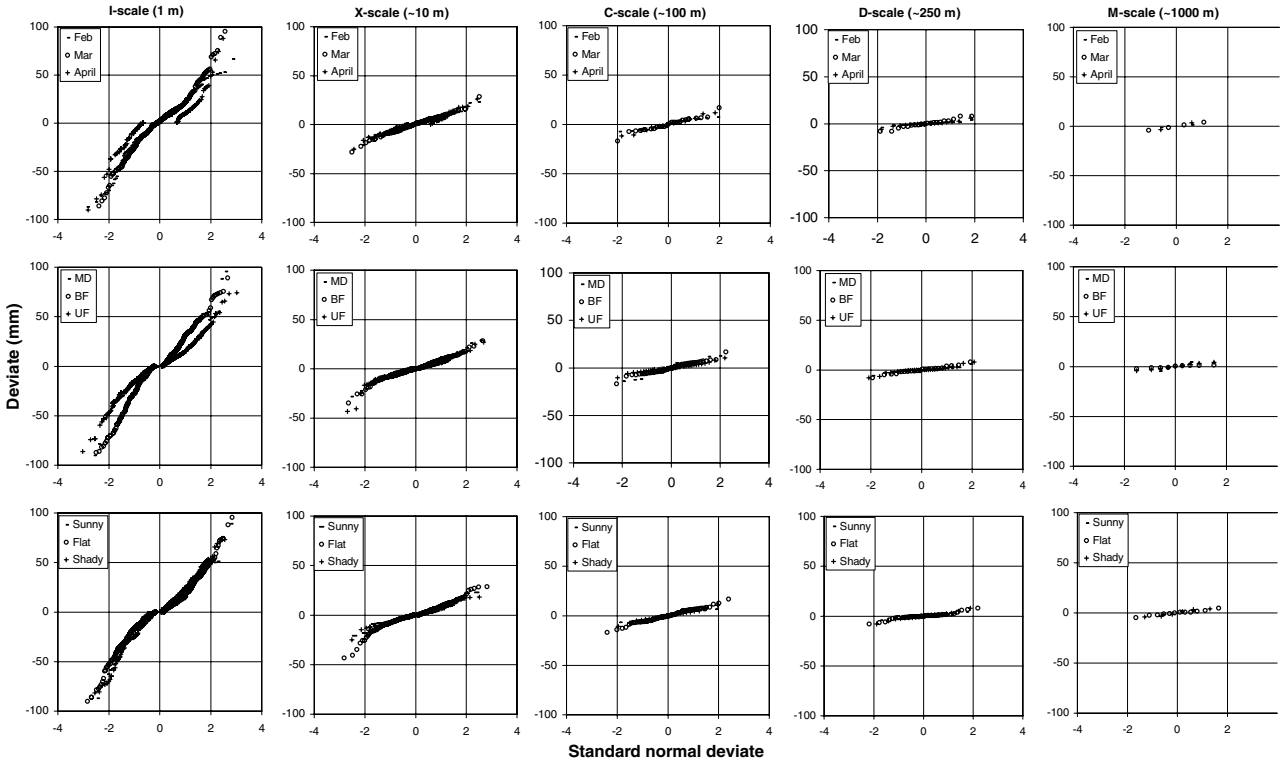

Figure 10 Corrected SWE deviations at each stage, pooled by various fixed effect groupings.

by trees. In summary, the distributions of deviations are non-normal and differ between fixed-effects strata, and so methods of estimation of confidence limits should allow for this.

Confidence intervals

We used a hierarchical bootstrap to estimate confidence intervals for each estimate of the mean SWE of a fixed effects stratum. Bootstrapping (Efron and Tibshirani, 1998) was chosen over classical parametric statistical methods because of evidence for non-normality in the data, and because the relative ease with which the bootstrapping can be used to estimate confidence intervals for a complex sampling design. Hierarchical bootstrapping is relatively new area (Davison, 1999), with initial methods suggested by Davison and Hinkley (1997) and Carpenter et al. (1999).

Our approach for the estimation of a confidence interval for a given fixed stratum was to replicate the actual nested sampling design used in the field for that stratum by drawing in a structured, random manner from stage-specific pools of (corrected) deviations. Deviations for the three coarsest scales (C, D, M) were pooled across all fixed strata, since there was minimal evidence in Fig. 13 for differences in the distribution of deviations at these scales that could be related to specific fixed effects. The idea of drawing bootstrap samples from pools larger than the original sample is unconventional, and rests on the demonstration of approximately identical distribution of deviations across all the original samples to be placed in a single pool. The pools for the two finest scales were simply the original sampled deviations for the fixed stratum in question. All draws were made with replacement, following Carpenter et al. (1999), but differing from the recommendation of Davison and Hinkley (1997) since it is unclear how drawing without replacement should proceed in multi-stage designs with very small n_s.

For each of the 40 fixed strata, 800 bootstrap replicates were created, each based on between 27 and 54 random draws from the appropriate pools of deviations. These deviations were averaged using a flat estimator, as defined and discussed earlier, giving 800 bootstrap replicate estimates of the mean SWE of the fixed stratum (expressed as net deviations about the original estimate). No more than 800 replications were needed, since the bootstrap standard error was shown to be convergent after about 100—500 replications for all fixed strata.

Confidence intervals were needed for two purposes. Firstly, we needed to know the accuracy of our mean SWE estimates prior to their use in validation of the snowpack simulation model. Secondly, we wanted to attach some degree of statistical significance to comparisons between SWE estimates for different fixed strata. Although not as strict as the use of formal simultaneous confidence interval methods, non-overlapping confidence intervals offer sufficient discriminatory power for the purposes of our discussion. We thus chose 90% two-sided limits (taken as the 40th and 760th values in the sorted listed of 800 replicate means), because each of these is equivalent to a 95% one-sided limit, making the comparison between two intervals approximately equal to a statistical test with 95% confidence.

Results

Table 3 summarizes all site characteristics and SWE results. The following sections examine variation in SWE, snow

Table 3 Summary of results of nested sampling of SWE, snow depth, and snow density at 40 strata in western central Yellowstone National Park – defined along gradients of time, elevation, vegetation, and radiation

Watershed	Location	Date/s	Mean elevation (m)	Veg	Radiation	Mean slope (degree)	Mean aspect (degree true)	Mean annual precipitation (mm)	Est. top of trees (m)	Median diameter of trees (cm)	Density (trees /ha)	Est. basal area (m2/ha)	Number of cores	Flat mean SWE (mm)	LCL90 (−mm)	UCL90 (+mm)	Mean snowpack depth (mm)	Mean snowpack density (%)	Variance components – expressed as standard deviations (mm)							Coefficient of variance (C_v) (%)
																			M	D	C	X	I	RSS MDCXI	(CXI) / (MDCXI)	
G	E of Gibbon Falls	24-Jan-01	2238	BF	Flat	2	227	540					30	62	−12	11	412	15	0.0	0.0	0.0	15.0	27.1	31.0	0.7	50
F	E of Firehole Cascades	25-Jan-01	2230	UF	Flat	8	153	469					45	50	−6	7	344	14	0.0	0.0	0.0	0.0	12.7	12.7	1.0	25
M	Madison Ra Overlook	03-Mar-01	2088	BF	Flat	1	101	413					21	93	−10	10	503	18	0.0	0.0	0.0	8.3	14.7	16.9	0.7	18
G	Gibbon Geyser Meadow	18-Mar-01	2191	MD	Flat	0	0	597					21	151	−17	17	703	21			22.6	24.0	28.8	43.8		29
G	E of Gibbon Falls	18-Mar-01	2180	BF	Flat	7	228	549					21	125	−18	17	504	24			0.0	24.6	30.8	39.4		31
G	Ridge S of Secret Valley	19-Mar-01	2212	UF	Flat	2	63	529	15	10	7133	11.9	27	96	−9	9	440	22		0.0	0.0	0.0	8.5	8.5	1.0	9
G	Ridge S of Secret Valley	19-Mar-01	2207	UF	Sunny	9	178	529	12	13	5333	20.4	27	92	−9	9	398	24		0.0	0.0	0.0	12.6	12.6	1.0	14
G	Ridge S of Secret Valley	19-Mar-01	2205	UF	Shady	22	9	529	19	12	3133	10.5	27	91	−9	8	510	18		0.0	0.0	0.0	14.4	14.4	1.0	16
G	Tanker Bend	20-Mar-01	2238	BF	Shady	31	12	548					27	122	− 17	15	641	19		0.0	0.0	17.0	41.6	45.0	0.8	37
G	NE of Gibbon Falls	20-Mar-01	2216	BF	Sunny	28	188	537					27	5	−8	8	12	7		0.0	0.0	0.0	6.5	6.5	1.0	126
X	Mixed	10-May-01	2492	MD	Flat	1	114	539					54	105	−8	7	282	37	28.6	0.0	31.2	11.7	18.6	47.7	0.4	45
X	Mixed	10-May-01	2569	UF	Flat	3	250	535	18	23	941	40.4	54	131	−12	13	402	32	3.1	15.6	4.6	41.9	32.2	55.4	0.5	42
∼G	Mixed	10-May-01	2462	BF	Flat	1	229						27	53	−13	14	142	34		10.6	25.5	36.5	20.4	50.1	0.5	95
∼G	Norris Junction SE	10-May-01	2307	UF	Flat								0	0			0									
∼G	Power Line nr Solfatara Ck	10-May-01	2305	MD	Flat								0	0			0									
M	Madison Junction	19-Feb-02	2103	UF	Flat	4	197	430	17	16	1467	29.9	27	98	−9	9	488	20		0.0	0.0	0.0	12.4	12.4	1.0	13
M	Madison Junction	19-Feb-02	2128	UF	Sunny	22	184	438	15	17	1733	35.6	27	96	−10	9	476	20		0.0	0.0	5.4	14.4	15.4	0.8	16
M	Harlequin Hill	20-Feb-02	2117	UF	Shady	31	359	427	15	20	800	26.7	27	102	−12	11	582	17		0.0	9.7	16.3	23.2	29.9	0.6	29
M	Harlequin Meadow	20-Feb-02	2101	MD	Flat	1	331	424					27	133	−9	10	627	21		0.0	0.0	0.0	18.3	18.3	1.0	14
M	Seven Mile	20-Feb-02	2082	BF	Shady	24	64	399	22	11	67	5.7	27	139	−14	13	694	20		0.0	0.0	0.0	30.1	30.1	1.0	22
M	Seven Mile	21-Feb-02	2098	BF	Sunny	27	227	389	16	0	0	0.0	54	114	−11	11	488	23	0.0	15.1	10.4	21.8	34.6	44.8	0.5	39
G	Virginia Meadows	22-Feb-02	2384	MD	Flat	0	0	709					27	263	− 11	11	991	27		0.0	0.0	12.0	19.4	22.9	0.7	9
M	West Entrance	21-Feb-02	2072	MD	Flat	1	271	520					27	183	− 9	8	759	24		0.0	11.8	0.0	8.8	14.7	0.7	8
M	Soldier Pits	21-Feb-02	2079	BF	Flat	0	325	399	13	0	67	0.0	27	103	−15	15	560	18		0.0	20.7	25.6	33.4	46.8	0.6	45
M	Madison Junction	17-Mar-02	2110	UF	Flat	5	202	430	20	22	1000	35.5	27	120	−14	13	606	20		0.0	0.0	19.6	27.1	33.5	0.7	28
M	Madison Junction	17-18-Mar-02	2129	UF	Sunny	20	183	438	18	17	850	18.4	36	120	−10	9	619	19	0.0	7.0	0.0	16.4	14.2	22.8	0.6	19
M	Harlequin Hill/Tuff Cliffs	17-19-Mar-02	2126	UF	Shady	21	354	441	21	15	1570	29.8	54	128	−8	8	676	19	24.9	28.0	0.0	10.7	22.8	45.1	0.3	35
M	Harlequin Meadow	17-Mar-02	2100	MD	Flat	0	0	424					27	101	− 14	15	728	14		0.0	11.8	29.4	33.2	45.9	0.6	45
M	Seven Mile	18-19-Mar-02	2059	BF	Shady	26	55	392	3	3	0	0.0	54	161	−11	10	843	19	0.0	19.2	11.9	9.8	34.9	42.7	0.5	27
M	Seven Mile	18-Mar-02	2072	BF	Sunny	25	216	400					27	156	− 16	14	675	23		0.0	11.8	7.0	40.5	42.8	0.7	27
G	Virginia Meadows	19-Mar-02	2363	MD	Flat	0	0	709					27	295	− 13	14	1178	25		0.0	16.3	26.5	25.4	40.2	0.6	14
M	West Entrance	20-Mar-02	2055	MD	Flat	0	257	520					27	234	− 9	9	942	25		0.0	18.6	3.3	11.6	22.1	0.7	9
M	Soldier Pits	20-Mar-02	2086	BF	Flat	0	0	399					27	132	− 16	14	623	21		34.3	53.9	20.1	37.0	76.4	0.5	58
M	Madison Junction	23-Apr-02	2105	UF	Flat	7	210	430	17	14	733	20.0	27	2	−8	8	5	3		0.0	0.0	0.0	5.3	5.3	1.0	268
M	Madison Junction	23-Apr-02	2131	UF	Sunny	20	184	438	18	12	1467	12.7	27	0	−8	7	0	0		0.0	0.0	0.0	0.0	0.0	0.0	
M	Harlequin Hill / Tuff Cliffs	23-Apr-02	2108	UF	Shady	23	357	441	21	16	2000	36.4	54	58	−10	9	210	25	18.9	0.0	21.8	18.2	22.0	40.5	0.4	70
M	Harlequin Meadow	23-Apr-02	2095	MD	Flat	0	0	420					9	9	− 10	11	20	5		0.0	13.2	21.5	1.3	25.2	0.7	295
M	Seven Mile	24-Apr-02	2076	BF	Shady	29	68	399					27	8	−10	10	31	5		0.0	0.0	1.2	16.7	16.7	0.9	211
M	Seven Mile	25-Apr-02	2072	BF	Sunny	25	216	400					0	0			0									
G	Virginia Meadows	24-Apr-02	2375	MD	Flat	0	0	709					27	242	− 17	16	670	36		0.0	36.7	26.6	39.6	60.2	0.6	25
M	West Entrance	24-Apr-02	2064	MD	Flat	0	0	520					27	25	−8	9	82	18		22.8	14.8	0.0	10.2	29.0	0.4	114
M	Soldier Pits	25-Apr-02	2086	BF	Flat	0	0	399					27	9	−11	11	27	6		0.0	0.0	7.9	22.1	23.4	0.8	249
Overall average or total													1209		−11.3	11.0			*7.5*	4	9	12	21	30.5	0.7	
2001 average or total													408		−11.3	11.2			*6.3*	2	6	14	21	29.5	0.8	
2002 average or total													801		−11.2	10.9			*8.7*	5	10	12	21	31.1	0.7	

Note: confidence limits are expressed as differences from the mean.

depth, and snow density due to each of the four fixed effects. Note that in each of the associated figures, the confidence limits are not drawn for every stratum. Instead, a single confidence interval is drawn in the upper left corner, corresponding to limits of ±11 mm. This is a typical value, whereas the actual stratum-specific values are given in Table 3.

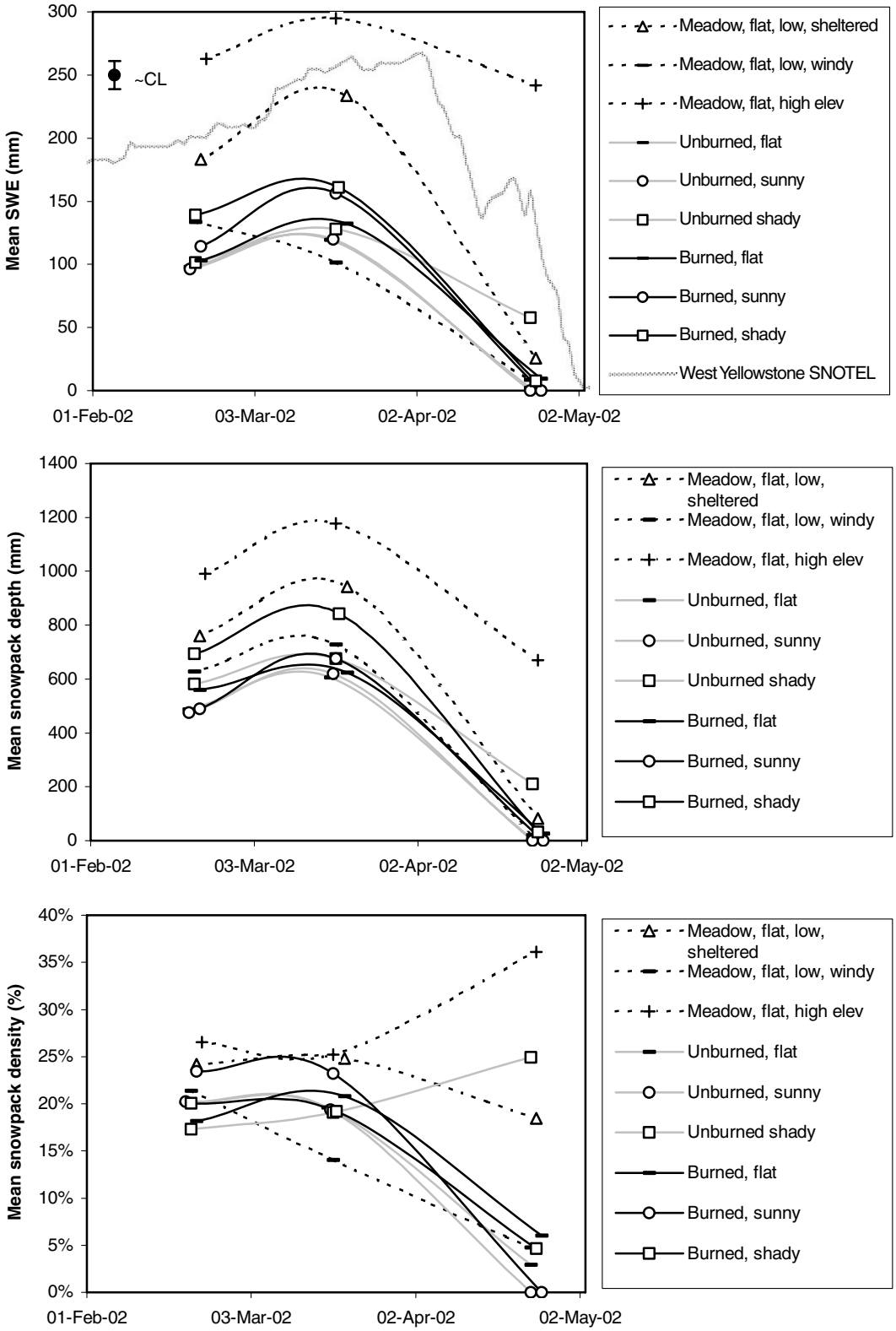

Figure 11 Snowpack variation over time (2002 data). For comparison, daily SWE data are shown from the West Yellowstone SNOTEL site (2042 m, 564 mm MAP, 21 km W of Madison Junction).

Temporal patterns

Looking at just the 2002 data, the February and March sampling campaigns occurred during the accumulation phases, and the April campaign occurred during late melt-out (Fig. 11). All fixed strata followed the same accumulation and ablation progression except the higher meadow stratum, which had only just begun ablating during the April

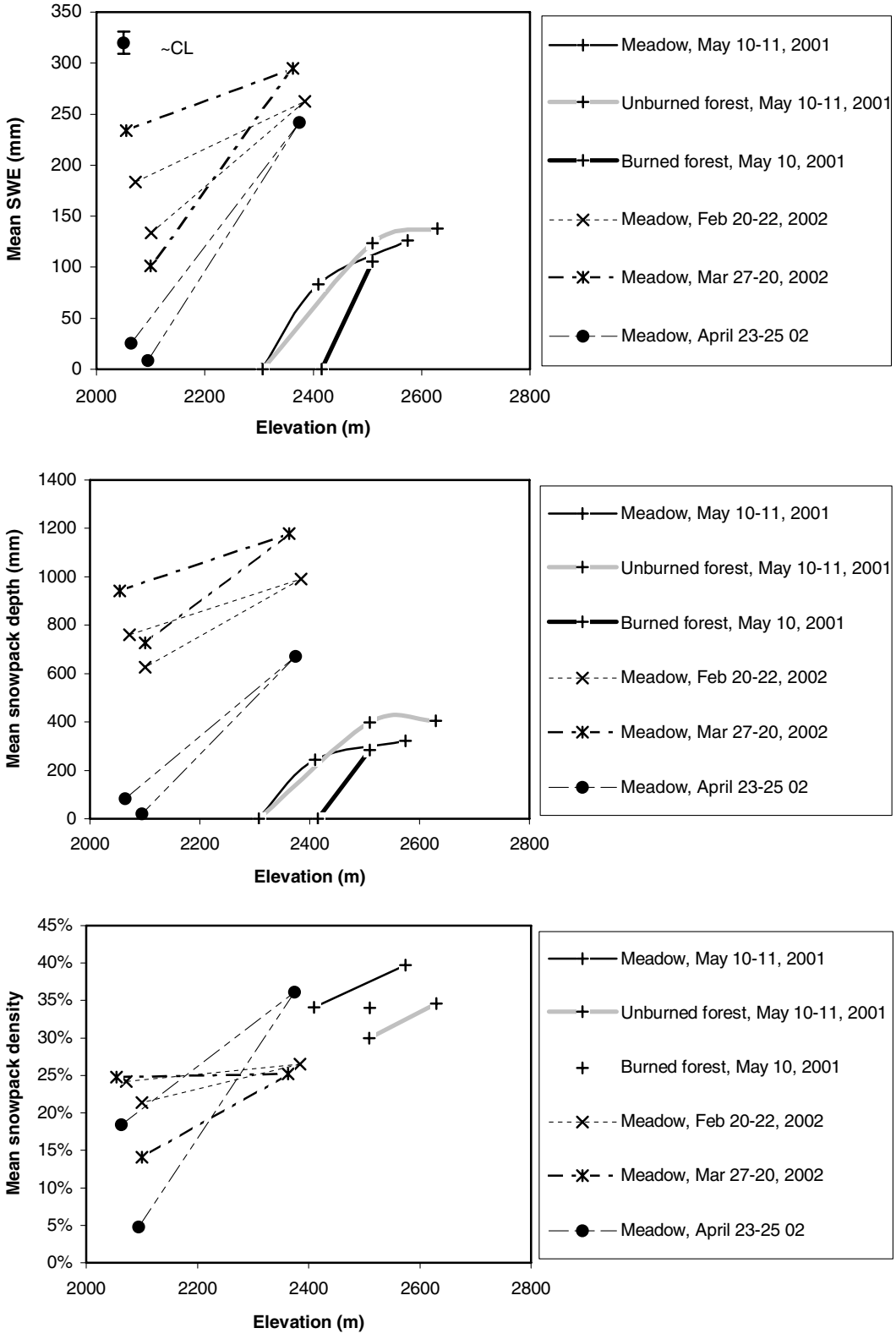

Figure 12 Snowpack variation with respect to elevation.

sampling campaign, and the low windy meadow stratum, which experienced intense wind-driven reductions in SWE prior to the March sampling.

Snowpack density remained relatively constant during accumulation, with the exception of the low windy meadow, which experienced a curious decline in density. Most sites were heavily ablated by April, but the two that were not ablated exhibited an expected increase in density during the early to mid ablation period.

Elevation effects

Elevation effects were not ideally sampled by our design, since suitable higher and lower elevation sites could not be located close enough to the other strata so as to eliminate the effects of regional precipitation trends. The original 'low' elevation site was later determined to have been located in an area that experiences higher mean annual precipitation (MAP) than all other parts of the study area at

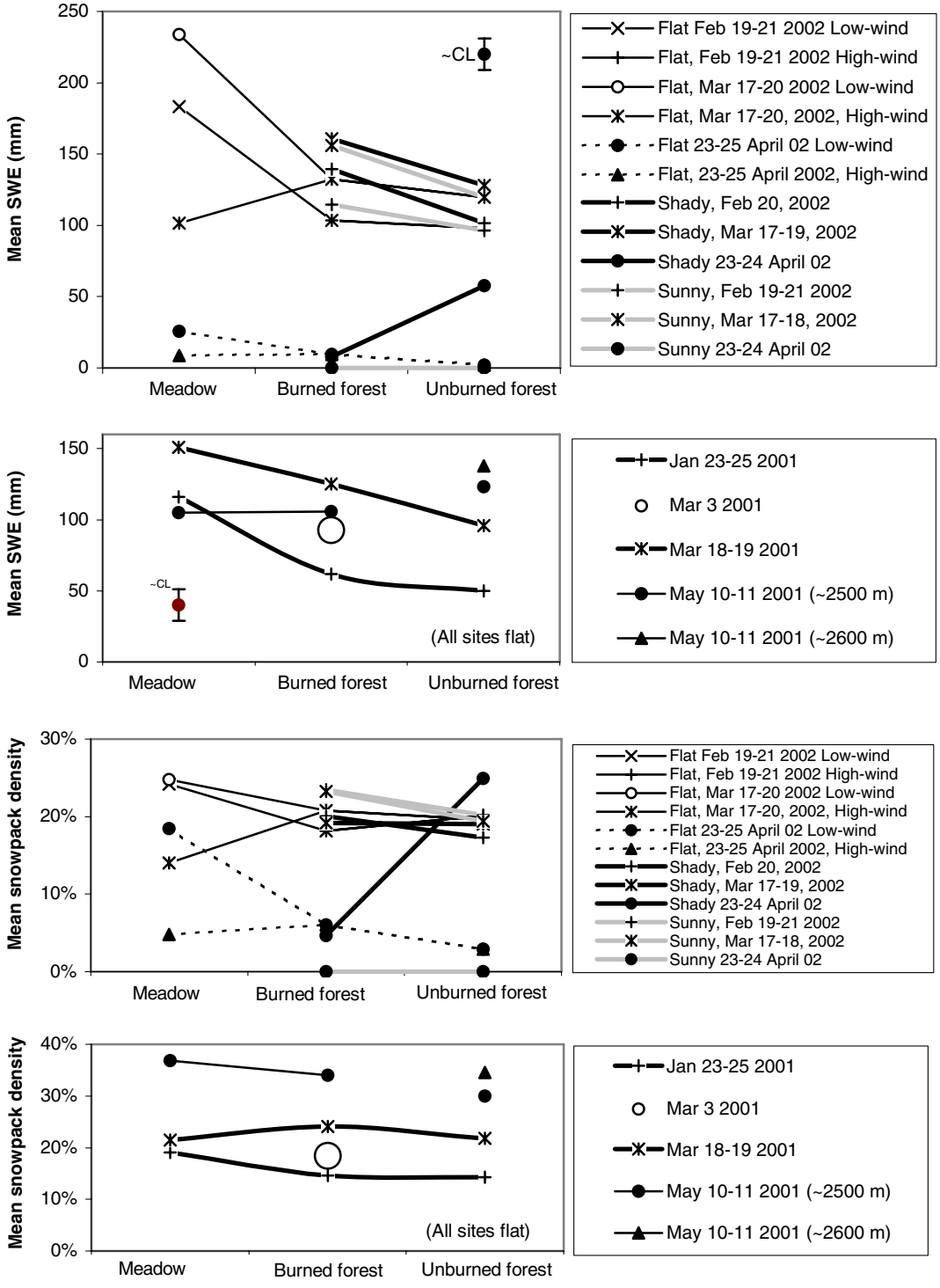

Figure 13 Snowpack variation with respect to vegetation.

comparable elevations (at least 18% higher MAP than all other strata, except the high elevation meadow). Conversely, the high elevation stratum was in an anomalously low portion of the regional precipitation pattern. Despite these problems, the data clearly confirm that the higher elevation strata supported deeper snowpacks with higher SWE (Fig. 12). There is also some evidence for higher snowpack density at higher elevations, particularly during the melt-out.

Vegetation effects

Vegetation effects are more subtly revealed in the data than temporal or elevation effects, and at times, are at the margin of 95% statistical confidence implied by barely non-overlapping confidence intervals. A priori, we expected unburned forests to accumulate the shallowest snowpacks (due to snowfall interception), but also to retain the snowpacks the longer than burned forests and meadows (due to solar radiation interception).

These expectations were reflected in the 2002 data (Fig. 13). The difference between unburned and burned forest was slight during the accumulation phase, with only occasional non-overlapping confidence intervals. This indicates that the effect of differing forest structure, although slight, is not a sampling artifact. The effect was strongest on both sunny and shady slopes, and lesser in flat terrain (not expected). Meadow data followed expectations in three SWE-comparisons and all four depth-comparisons. The SWE pattern was reversed in the fourth SWE comparison, apparently due to wind-driven sublimation that was not accompanied by a reduction in depth. Densities during the accumulation phase were also generally higher in meadows, which is consistent with their surfaces being exposed to sun-driven melt-freeze cycles. Sampling during the ablation phase revealed almost complete ablation at most sites, except unburned forest on shaded slopes — consistent with expectations. The 2001 data agree with the 2002 data, confirming a reversal of vegetation effect between accumulation and ablation phases, and also revealing strong mid-ablation densification, more so in solar-exposed areas.

Meadow SWE was least variable, with a mean coefficient of variation (C_v) of 25% for strata in February and March of 2002. Not surprisingly, given the heterogeneity of strewn logs, burned forest was most variable, with a C_v of 61% (Feb., Mar., 2002). Note that logs were considered to be a

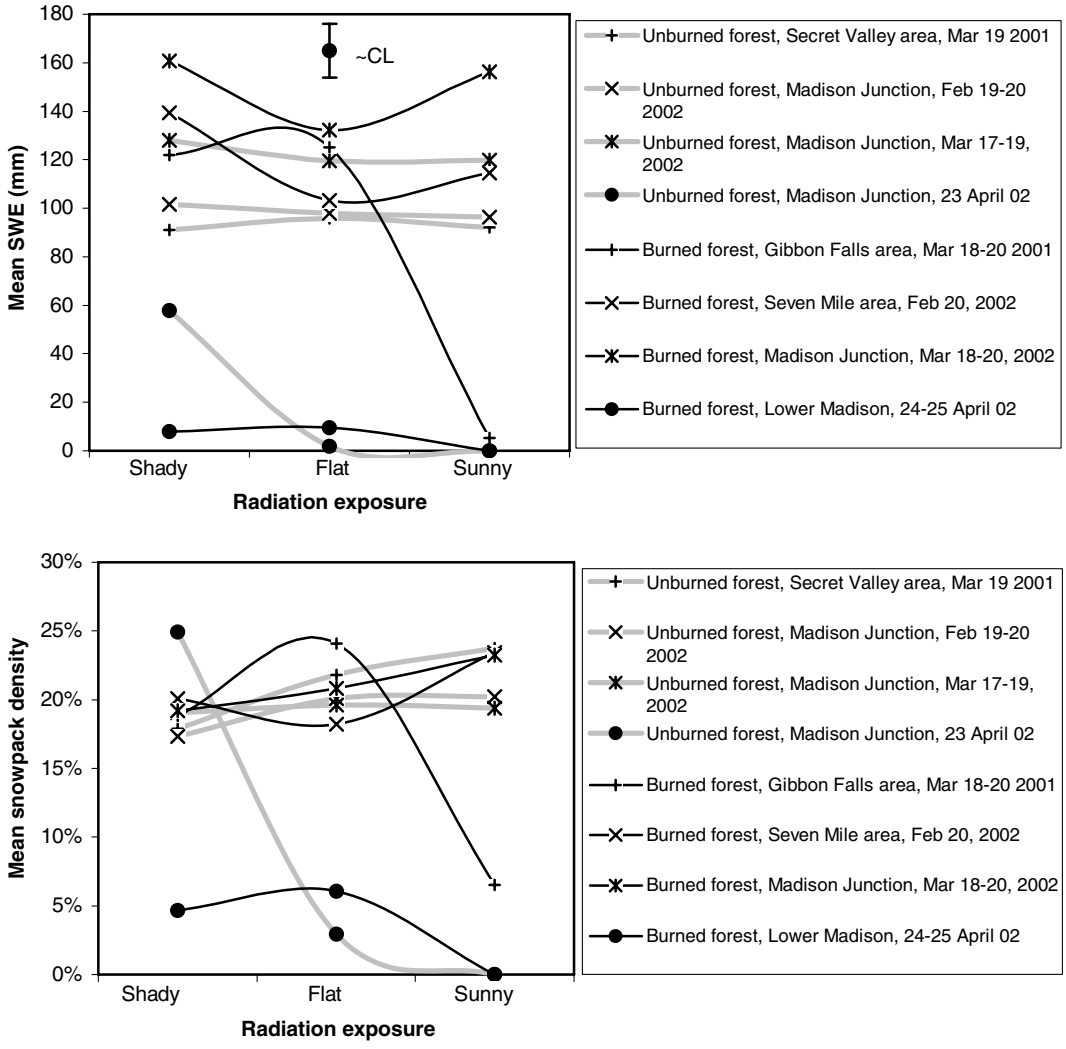

Figure 14 Snowpack variation with respect to terrain-induced solar radiation effects.

source of random variation just like any other, and thus were not censored or adjusted for in any way, since their presence imparts a real effect on mean SWE. Unburned forest had an intermediate C_v of 37%. These values follow the opposite trend to that reported by Pomeroy et al. (1998a). These authors used an unstated, and most likely different spatial sampling design that we suggest may have influenced their data as much as actual vegetation effects.

Radiation effects

At the outset, radiation effects were expected to result in a progression of reducing SWE and snow depth as one moves from shady, to flat, to sunny slopes. The data revealed this progression, but only during ablation (Fig. 14). Radiation effects were expected to be more pronounced in the burned forest than in the unburned forest, both due to increased snowpack insolation, and due to longwave radiation from standing dead timber. Longwave radiation from dead timber is likely to be more pronounced than from living trees, since the latter can also cool themselves and thus balance their energy surplus by losing latent heat through transpiration. The data agree with these expectations. The snowpack beneath burned forests on sunny slopes was ablated by March in 2001, whereas no radiation effect was observed in unburned forests at the same time. In similar agreement, the burned forest snowpack was ablated on all slopes by April in 2002, but only ablated on sunny and flat slopes in unburned forest. Excluding ablation-phase data, snowpack density was highest on sunny, followed by flat slopes. This is consistent with points made previously that insolation increases density.

Discussion

The discussion of our results is organized in sections on overall sampling approach, random effects, spatial variance structure, fixed effects, and limitations.

Sampling approach and random effects

Nested sampling to estimate the mean of strata is a novel approach to snowpack measurement. Stratification is of course not a new concept in snow hydrology (Stepphun and Dyck, 1974); cited by Shook and Gray (1996); Elder et al. (1991). However, while various forms of stratified and systematic sampling are well understood to lead to better estimators of the spatial population mean than purely random sampling (Dunn and Harrison, 1993), few universal truths have emerged about estimator dependency on spatial variance structure (Haining, 2003). We have shown that nested sampling of statistically stationary SWE fields can produce optimal or near optimal accuracy of estimated means for a given cost.

Both our model-driven approach and the empirical interpolation approach suffer in their ability to quantify the total error in estimates of either basin-wide averages or spatial patterns. This is because regardless of whether stratification is done at the outset (our approach) or after the fact (e.g. the regression tree-based approach), the assumption that residual variance is random cannot be tested. The stratification may have excluded a key fixed effect upon which the residual variance is dependent, or it may have sampled a known fixed effect in a non-random manner.

Pomeroy et al. (2002) considered SWE variation within a uniform forest stand to be 'unpredictable'. There is a need to better understand when variance is unpredictable, and if so, how its randomness is structured. The few studies that have sampled snowpacks at a large number of locations have done so with minimal attention to characterizing the considerable scale-dependent random variance (Yang and Woo, 1999), the way in which estimators and optimal sampling designs may be dependent on this (Elder et al., 1991), and the degree to which sampling for fixed effects may be obscured by random effects (Cooley, 1988; Luce et al., 1998; Tarboton et al., 2000).

Spatial variance structure

A number of studies have examined the spatial statistical distribution (i.e. variograms) of snowcover on the grounds that this facilitates better estimation of the effects of spatial variability on basin-averaged snowmelt dynamics. The present study measured variograms with non-zero nugget and a well-defined sill — random SWE variation was considerable over short distances ($\leqslant$10 m), but commonly decreased in individual strata to effectively zero at distances larger than about 100 to 1000 m. This implies within-stratum stationarity. We occasionally observed larger scale variance, but in each case this was thought to be due to the unintentional introduction of larger scale fixed effects (such as precipitation shadows) that violated the theoretical boundaries of our strata. Exponential or spherical variogram models could most likely be fitted to our data. The correspondence of these observations with those of others requires some clarification of the literature. Firstly, the presence of a sill is generally conditional on controlling for non-stationary effects. Studies that do not do this, do not describe variogram sills, and should not be naively compared to stationary data. Kuchment and Gelfan (2001) assert a (non-silled) power-law variogram model, but their data include marked fixed effects at length scales large enough to be orographic. Fractal data exhibit power-law variograms, and the popularity of the search for self-similarity has apparently led to a tendency to perceive power-law variograms to the exclusion of other possibilities. Shook and Gray (1996) fitted a power-law variogram models with truncation (i.e. a sill) beyond a certain range to their data, which they then referred to as 'semi-fractal'. A similar suggestion was made by Essery et al. (1999). In both cases, the possibility of an inherently silled, equally plausible model such as the spherical model (Balk and Elder, 2000) was not considered. Working in the monotony of the arctic tundra, Holmgren et al. (1998) measured well-defined sills in snow depth at a range of about 10 m, using both ground-based radar and hand-probe measurements. To synthesize these and our studies, we observe that fixed effects on snowpacks tend to have easily predictable covariates (wind is an exception) — and once these have been accounted for, the residual spatial variance has a well-defined sill and a short (1 to 100 m) range. These should be characterized by spherical or exponential variograms, and if

the scale of interest transcends local stationarity, then nested exponential variograms should be considered equally as much as power-law (fractal) variograms (Blöschl, 1999).

It is also worth giving careful consideration to assumptions about normality or log-normality of spatial variance in SWE, for this has implications on scaling between point-scale measurements, areal cover measurements, and mean aerial SWE (Skaugen, 1999). For example, log-normality was hypothesized by Pomeroy et al. (1998), cited by Pomeroy et al. (2002) see also (Essery et al., 1999). However, confidence limits for their Q–Q test for log-normality were not given, nor were alternate hypotheses considered. We suggest that log-normality – or more specifically, the positive skewness of previously reported snowpack data – is actually determined by dataset-specific fixed effects that could be modeled deterministically, rather than treated as residual variance represented by a frequentist model like a normal or log-normal distribution. Our data are controlled as much as possible for fixed effects. The residual (random) variance appears to be non-skewed, and thus not lognormal. There is evidence that it is even non-normal, given the positive kurtosis in our Q–Q plots.

Fixed effects

Our fixed-effect results generally agreed with expectations. In agreement with previous authors (Tarboton et al., 2000; Murray and Buttle, 2003), terrain-driven solar radiation effects to only manifest in SWE comparisons during melt out, most likely due both to high snowpack albedo during the snowfall season (USACE, 1956), and effects on snowpack energy content that do not manifest as SWE variations until meltout.

A similar intra-season shift occurs for vegetation effects. As summarized by Faria et al. (2000), it is well established that SWE accumulation is lower under dense canopies than in neighboring open areas, due to interception of precipitation (McKay and Gray, 1981; Pomeroy et al., 1998b). The magnitude of these effects is greatest between coniferous forests and meadows, with typical SWE-ratios (meadow:forest) ranging in the literature between 1.09 and 4, with a median of about 1.4 (Murray and Buttle, 2003). The ratios we measured were in general agreement with the literature, ranging from 0.82 to 2.32, with an average of 1.63 (values derived from Table 3.). It is also well established that, conversely, SWE is retained longer under denser canopies (Metcalfe and Buttle, 1995, 1998; Davis et al., 1997) due to attenuation of solar radiation (Ni et al., 1997). Note that this positive covariance applies at the stratum level only – since radiation feedback effects cause a negative covariance at smaller scales (Faria et al., 2000). Our data agree with these understandings.

The general model for density is for an increasing trend over time, especially near the end of the season (e.g. Anderton et al., 2004). Many of our data tend to agree with this model, but there are some notable data points where reductions in density occurred, apparently as a result of mid-season interactions with dense low vegetation, very late meltout interactions with ground vegetation, and perhaps a sublimation effect due to strong winds. Higher densities were observed in sunny strata (meadows and south-facing slopes) and at higher elevations (perhaps due to compression of deeper snowpacks, Tabler et al., 1990).

Although not included as such in the present study, in hindsight we believe wind to be a predictable fixed effect to some extent. This has been shown in empirical studies at small scales, for example by Anderton et al. (2004) It has also been incorporated as an empirical factor in modeling studies (Luce et al., 1998). The challenge remains to achieve a priori prediction of terrain-driven and vegetation-driven wind effects at landscape scales.

Limitations

Our conclusions are notwithstanding certain limitations, which should be addressed in future. Firstly, wind should be included as a fixed effect following existing work on landscape indices for blowing snow models (e.g. Essery et al., 1999, Fig. 9). Further, sampling should be conducted in high snow years, in addition to the low snow years experienced during the present study.

Certain statistical assumptions may have been violated. The assumption of within-stratum stationarity of snowpack properties should be tested, and methods for randomly locating nested designs within strata should be evaluated within this context. The assumption of uniform covariates within strata should be tested – for example, leaf area index apparently varied within our unburned forest strata, with likely effects on SWE (Pomeroy et al., 2002). Bias may have been introduced by the practical sampling constraints described. For example, 'random' variation at long length scales due to wind turbulence may differ between the sheltered valley-bottom sites typically accessible near roads and the more remote plateau sites not visited during the study. Inter-dependence between samples may have been introduced by the use of equilateral triangles in the smaller-scale parts of the sample design, which may have lead to underestimation of confidence limits.

Conclusions

Optimal sampling designs for the mean are those that yield unbiased estimates with minimal sampling variance at minimal sampling cost. In spatial sampling, the optimality of a given sampling design depends on: relative sampling effort at each spatial scale, relative amounts of random variance exhibited at each spatial scale, and relative sampling cost at each spatial scale. Nested sampling provides a convenient structure to examine scale-dependent variation, and to tailor sampling effort to specific spatial scales. In our data, 65% of the random SWE variation occurred at spatial scales up to 10 m, and 83% occurred at scales up to 100 m. Thus, optimal designs should emphasize samples spaced at these scales. Within the nested framework, the coarsest stage should also sample variation up to 300-m scales, but only sparsely, since – depending on the constraints of the study – sampling cost may be become prohibitive for more than one or two groups separated by these distances. Random variation appears to be negligible at scales greater than about 300 m. Spatial variance structure of snowpacks is dependent on seasonal timing and on landscape characteris-

tics such as degree of forest cover — but perhaps not to so great an extent as to warrant separate sampling designs for different landscapes within the same study.

Vegetation and solar radiation effects were present in roughly equal amounts, but in some cases were only statistically significant once random effects were sufficiently reduced by judicious sampling at between 27 and 54 spatially nested locations. The effects of sampling time and elevation were more obvious. The effects of wind — retrospectively — appeared to be significant.

The resulting data are suited for validation of deterministic spatial snowpack models in heterogeneous landscapes. They are an improvement over previous data because they minimize the variance due to random spatial fields that would otherwise unfairly obfuscate comparisons between model estimates and field data.

Acknowledgements

This project was conducted in partnership with the Yellowstone Center for Resources, Yellowstone National Park (PJ White). The stratified nested sampling approach was originally suggested by Matt Ferari. We are also indebted to J. Coughlan, S. Alexander, D. Ouren, A. Messer, E. Bergman, C. Gower, R. Jaffe, D. Kozlowski, J. Casagrande, J. Casagrande, R. Burton, K. Uschyk, J. Beard, A. Davison, M. Keator, K. Dimmick, D. Young, and two peer-reviewers. Funding was provided by the NASA Intelligent Data Understanding program (NCC2-1186, Watson), the NASA Office of Earth Science 'Research Education Applications Solutions Network' (NCC13-03009, Watson), NPS, and the National Science Foundation (DEB 0074444 and DEB-0413570, Garrott).

References

Anderton, S.P., White, S.M., Alvera, B., 2004. Evaluation of spatial variability in snow water equivalent for a high mountain catchment. Hydrological Processes 18, 435—453.

Balk, B., Elder, K., 2000. Combining binary decision tree and geostatistical methods to estimate snow distribution in a mountain watershed. Water Resources Research 36, 13—26.

Bergengren, J.C., Thompson, S.L., Pollard, D., Deconto, R.M., 2001. Modeling global climate-vegetation interactions in a doubled CO_2 world. Climatic Change 50, 31—75.

Bergman, E.J., Garrott, R.A., Creel, S., Borkowski, J.J., Watson, F.G.R., Jaffe, R.M., in press. Assessment of prey vulnerability through analysis of wolf movements and kill sites. Ecological Application.

Blöschl, G., 1999. Scaling issues in snow hydrology. Hydrological Processes 13, 2149—2175.

Blöschl, G., Kirnbauer, R., Gutknecht, D., 1991. Distributed snowmelt simulations in an alpine catchment 1. Model evaluation on the basis of snow cover patterns. Water Resources Research 27, 3171—3179.

Carpenter, J., Goldstein, H., Rasbash, J., 1999. A non-parametric bootstrap for multilevel models. Multilevel Modelling Newsletter 11 (1), 2—5. Available from: <http://multilevel.ioe.ac.uk>.

Cline, D.W., Bales, R.C., Dozier, J., 1998. Estimating the spatial distribution of snow in mountain basins using remote sensing and energy balance modeling. Water Resources Research 34, 1275—1285.

Cooley, K.R., 1988. Snowpack variability on Western Rangelands. In: Western Snow Conference Proceedings, Kalispell, Montana, April 18—20.

Crow, E.L., Davis, F.A., Maxfield, M.W., 1960. Statistics Manual: With Examples taken from Ordnance Development. Dover, New York, 288 pp.

Davis, R.E., Hardy, J.P., Ni, W., Woodcock, C., McKenzie, J.C., Jordan, R., Li, X., 1997. Variation of snow cover ablation in the boreal forest: a sensitivity study on the effects of the confider canopy. Journal of Geophysical Research 102, 29389—29395.

Davison, A.C., 1999. Bootstrap theory and resampling methods. In: Bulletin of the International Statistical Institute, Proceedings of 52nd session, Tome LVIII, Finland, 1999, 2pp.

Davison, A.C., Hinkley, D.V., 1997. Bootstrap Methods and their Application. Cambridge Univ. Press, Cambridge, UK, 582 pp.

Dunn, R., Harrison, A.R., 1993. Two dimensional systematic sampling of land use. Applied Statistics 42, 585—601.

Efron, B., Tibshirani, R.J., 1998. An Introduction to the Bootstrap. Chapman and Hall/CRC, Boca Raton, USA, 436 pp.

Elder, K., Dozier, J., Michaelsen, J., 1989. Spatial and temporal variation of net snow accumulation in a small alpine watershed, Emerald Lake basin, Sierra Nevada, California, USA. Annals of Glaciology 13, 56—63.

Elder, K., Dozier, J., Michaelsen, J., 1991. Snow accumulation and distribution in an alpine watershed. Water Resources Research 27, 1541—1552.

Elder, K., Rosenthal, W., Davis, R.E., 1998. Estimating the spatial distribution of snow water equivalences in a montane watershed. Hydrological Processes 12, 1793—1808.

Erxleben, J., Elder, K., Davis, R., 2002. Comparison of spatial interpolation methods for estimating snow distributions in the Colorado Rocky Mountains. Hydrological Processes 16, 3627—3649.

Essery, R., Li, L., Pomeroy, J., 1999. A distributed model of blowing snow over complex terrain. Hydrological Processes 13, 2423—2438.

Essery, R., 2003. Aggregated and distributed modelling of snow cover for a high-latitude basin. Global and Planetary Change 38, 115—120.

Faria, D.A., Pomeroy, J.W., Essery, R.L.H., 2000. Effect of covariance between ablation and snow water equivalent on depletion of snow-covered area in a forest. Hydrological Processes 14, 2683—2695.

Farnes, P.E., Romme, W.H., 1993. Estimating localized SWE on the Yellowstone Northern Range. Proc. 50th Eastern Snow Conference and 61st Western Snow Conference, Quebec City, 1993. pp. 59—65.

Gower, J.C., 1962. Variance component estimation for unbalanced hierarchical classifications. Biometrics 18 (4), 537—542.

Garrott, R.A., Eberhardt, L.L., White, P.J., Rotella, J., 2003. Climate-induced variation in vital rates of an unharvested large-herbivore population. Canadian Journal of Zoology 81, 33—45.

Greene, E.M., Liston, G.E., Pielke Sr, R.A., 1999. Relationships between landscape, snowcover depletion, and regional weather and climate. Hydrological Processes, 2453—2466.

Haining, R., 2003. Spatial Data Analysis: Theory and Practice. Cambridge Univ. Press., Cambridge, UK, 432 pp.

Halfpenny, J.C., Ozanne, R.D., 1989. Winter: An Ecological Handbook. Johnson Books, Boulder, Colorado, USA, 273 pp.

Holmgren, J., Sturm, M., Yankielun, N.E., Koh, G., 1998. Extensive measurements of snow depth using FM-CW radar. Cold Regions Science and Technology 27, 17—30.

Isaacs, E.H., Srivastava, R.M., 1989. Applied Geostatistics. Oxford Univ. Press, New York, 561 pp., Reg. Res. and Eng. Lab., Hanover, N.H.

König, M., Sturm, M., 1998. Mapping snow distribution in the Alaskan Arctic using aerial photography and topographic relationships. Water Resources Research 34, 3471—3483.

Kuchment, L.S., Gelfan, A.N., 2001. Statistical self-similarity of spatial variations of snow cover: verification of the hypothesis and application in the snowmelt runoff generation models. Hydrological Processes 15, 3343–3355.

Lapen, D.R., Martz, L.W., 1996. An investigation pf the spatial association between snow depth and topography in a Prairie agricultural landscape using digital terrain analysis. Journal of Hydrology 184, 277–298.

Liston, G.E., Pielke Sr, R.A., Greene, E.M., 1999. Improving first-order snow-related deficiencies in a regional climate model. Journal of Geophysical Research 104 (D16), 19559–19567.

Lohr, S.L., 1999. Sampling: Design and Analysis. Duxbury Press, Pacific Grove, CA, USA, 494 pp.

Luce, C.H., Tarboton, D.G., Cooley, K.R., 1998. The influence of spatial distribution of snow on basin-averaged snowmelt. Hydrological Processes 12, 1671–1683.

Marchand, P.J., 1996. Life in the Cold: An Introduction to Winter Ecology. University Press of New England, Hanover, NH, USA, 304 pp.

McKay, G.A., Gray, D.M., 1981. The distribution of snowcover. In: Gray, D.M., Male, D.H. (Eds.), Handbook of Snow: Principles, Processes, Management and Use. Pergamon Press, Canada, pp. 153–190.

Metcalfe, R.A., Buttle, J.M., 1995. Controls of canopy structure on snowmelt rates in the boreal forest. Proceedings of Eastern Snow Conference 52, 249–257.

Metcalfe, R.A., Buttle, J.M., 1998. A statistical model of spatially-distributed snowmelt rates in a boreal forest basin. Hydrological Processes 12, 1701–1722.

Miesch, A.T., 1975. Variograms and variance components in geochemistry and ore evaluation. Geological Society of America, Memoir 142, 333–340.

Mote, T.L., Grundstein, A.J., Leathers, D.J., Robinson, D.A., 2003. A comparison of modeled, remotely sensed, and measured snow water equivalent in the northern Great Plains. Water Resources Research 39, 1209. doi:10.1029/2002WR001782.

Murray, C.D., Buttle, J.M., 2003. Impacts of clearcut harvesting on snow accumulation and melt in a northern hardwood forest. Journal of Hydrology 271, 197–212.

Natural Resources Conservation Service (NRCS), (undated), Snow Surveys and Water Supply Forecasting. NRCS, National Water and Climate Center, Agriculture Information Bulletin 536. USA. http://www.wcc.nrcs.usda.gov/factpub/aib536.html.

Ni, W., Li, X., Woodcock, C., Roujean, J.L., Davis, R.E., 1997. Transmission of solar radiation in boreal conifer forests: measurements and models. Journal of Geophysical Research 102 (D24), 29555–29566.

Olyphant, G.A., 1986. Longwave radiation in mountainous areas and its influence on the energy balance of alpine snowfields. Water Resources Research 22, 62–66.

Ontario Ministry of Natural Resources (OMNR), (c. 1997), The Snow Network for Ontario Wildlife: The Why, When, What and How of Winter Severity Assessment in Ontario. Canada. 18 pp. http://www.mnr.gov.on.ca/MNR/pubs/snow.pdf.

Pomeroy, J.W., Gray, D.M., Shook, K.R., Toth, B., Essery, R.L.H., Pietroniro, A., Hedstrom, N., 1998a. An evaluation of snow accumulation and ablation processes for land surface modelling. Hydrological Processes 12, 2339–2367.

Pomeroy, J.W., Parviainen, J., Hedstrom, N., Gray, D.M., 1998b. Coupled modelling of forest snow interception and sublimation. Hydrological Processes 12, 2317–2337.

Pomeroy, J.W., Hanson, S., Faria, D., 2001. Small-scale variation in snowmelt energy in a boreal forest: an additional factor controlling depletion of snowcover? In: Proceedings of 58th Eastern Snow Conference, Ottawa, Ontario, Canada, 2001, 10pp.

Pomeroy, J.W., Gray, D.M., Hedstrom, N.R., Janowicz, J.R., 2002. Prediction of seasonal snow accumulation in cold climate forests. Hydrological Processes 16, 3543–3558.

Shook, K., 1995. Simulation of the ablation of prairie snowcovers. PhD Thesis, University of Saskatchewan, Sakatoon, SK, 189 pp.

Shook, K., Gray, D.M., 1996. Small-scale spatial structure of shallow snowcovers. Hydrological Processes 10, 1283–1292.

Skaugen, T., 1999. Estimating mean areal snow water equivalent by integration in time and space. Hydrological Processes 13, 2051–2066.

Smith, S.R., O'Brien, J.J., 2001. Regional snowfall distributions associated with ENSO: Implications for seasonal forecasting. Bulletin of the American Meteorological Society 82, 1179–1192.

Sokal, R.R., Rohlf, F.J., 1981. Biometry: The Principles and Practice of Statistics in Biological Research, second ed. W.H. Freeman and Co., New York, 859 pp.

Stepphun, H., Dyck, G.E., 1974. Estimating true basin snowcover. In: Adv. Concepts Tech. Study Snow and Ice Resour., Interdisciplinary Symp. US Nat. Acad. Sci., Washington, DC. pp. 314–328.

Tabler, R.D., Pomeroy, J.W., Santana, B.W., 1990. Drifting snow. In: Ryan, W.L., Crissman, R.D. (Eds.), Cold Regions Hydrology and Hydraulics. American Society of Civil Engineers, New York, pp. 95–146.

Tarboton, D., Blöschl, G., Cooley, K., Kirnbauer, R., Luce, C., 2000. Spatial snow cover processes at Kühati and Reynolds Creek. In: Grayson, R., Blöschl, G. (Eds.), Spatial Patterns in Catchment Hydrology: Observations and Modeling. Cambridge Univ. Press., Cambridge, UK, pp. 158–186.

Thompson, S.K., 2002. Sampling, second ed. John Wiley and Sons, Inc., New York, 367 pp.

United States Army Corps of Engineers (USACE), 1956. Snow Hydrology. Portland, Oregon, North Pacific Division, 437 pp.

Walsh, S.J., Butler, D.R., Allen, T.R., Malanson, G.P., 1994. Influence of snow patterns and snow avalanches on the alpine treeline ecotone. Journal of Vegetation Science 5, 657–672.

Watson, F.G.R., Newman, W.N., Coughlan, J.C., Garrott, R.A., in press., Testing a distributed snowpack simulation model against spatial observations. Journal of Hydrology.

Webster, R., Oliver, M.A., 2001. Geostatistics for Environmental Scientists. Wiley, Chichester, UK, 271 pp.

Yang, D., Woo, M.-K., 1999. Representativeness of local snow data for large scale hydrologic investigations. Hydrological Processes 13, 1977–1988.

Youden, W.J., Mehlich, A., 1937. Selection of efficient methods for soil sampling. Contributions of the Boyce Thompson Institute for Plant Research 9, 59–70.