Mem. Natl Inst. Polar Res., Spec. Issue, **58**, 210–226, 2004

Technical report

An advance in geolocation by light

Philip A. Ekstrom

Lotek Wireless 114 Cabot St., St. John's, NL, A1C 1Z8 Canada
CEFAS, Pakefield Rd. Lowestoft, Suffolk NR33 0HT, England (phil.ekstrom@lotek.com)

(Received March 21, 2003; Accepted July 14, 2003)

Abstract: A new analysis of twilight predicts that for observations made in narrow-band blue light, the shape of the light curve (irradiance *vs*. sun elevation angle) between +3 and –5° (87 to 95° zenith angle) has a particular rigid shape not significantly affected by cloudiness, horizon details, atmospheric refraction or atmospheric dust loading. This shape is distinctive, can be located reliably in measured data, and provides a firm theoretical basis for animal geolocation by template-fitting to irradiance data.

The resulting approach matches a theoretical model of the irradiance *vs*. time-of-day to the relevant portion of a given day's data, adjusting parameters for latitude, longitude, and cloudiness. In favorable cases, there is only one parameter choice that will fit well, and that choice becomes the position estimate. The entire process can proceed automatically in a tag.

Theoretical estimates predict good accuracy over most of the year and most of the earth, with difficulties just on the winter side of equinox and near the equator. Polar regions are favorable whenever the sun crosses –5° to +3° elevation, and the method can yield useful results whenever the sun makes a significant excursion into that elevation range. Early results based on data taken on land at 48°N latitude confirm the predictions *vs*. season, and show promising performance when compared with earlier threshold-based methods.

key words: geolocation, geopositioning, animal tracking, twilight

Introduction

One of the important quantities to be measured and logged for a free-living animal is its location. While the Global Positioning System or Service ARGOS will serve for locating sufficiently large terrestrial animals or marine animals that surface sufficiently often and stay sufficiently long on the surface, pelagic fish like tuna and ocean-phase salmon live below salt water so that radio techniques cannot be used. They also may carry a tag for many years before recovery, imposing a tight power budget on any battery-powered instrument. For such animals, a completely passive geolocation method based on ambient light has become important (Arnold and Dewar, 2001). However, previously available threshold-based methods have persistent limitations (Musyl *et al.*, 2001). Against this background new methods for light-based geolocation that may have less severe limitations become interesting.

The roots of newer template-fitting approaches to geolocation lie in the earlier threshold methods, those earlier methods are needed to provide initial position estimates for fitting, and they also provide a convenient setting for introducing astronomical details. Therefore we begin with a brief discussion of threshold methods. Second we introduce a geophysical model published elsewhere (Ekstrom, 2002) that provides some insight and a needed template. Finally we discuss a new template-fitting method, trying it and a threshold method against a limited set of data to illustrate some of their similarities and differences.

Materials and methods

This entire work is intended as a methods paper, so the present section will focus on the field data used below to illustrate the behavior of two geolocation methods. These data were taken on land at a dark rural site located in a forest clearing at 48°35′ N, 122°57′ W (Shaw Island, WA, USA). The surrounding trees provide a screen several degrees high (a classically "bad" horizon) that helps to shield the site from artificial light. They also contribute shadows that can be seen at high sun angles in the data of Ekstrom (2002). It is a central result of that paper that such a horizon does not affect geolocation done based on blue-light measurements taken for sun elevation angles below 3°.

Data were taken by Lotek LTD750 archival tag serial number 1563. (This model tag is obsolete and no longer available. The Lotek 2000 series are successor tags with the same kind of clock, light detector, and geolocation algorithms.) The light detector was of a fluorescent conversion type where the excitation spectrum of a fluorescent dye is used as a filter to define the sensor's wavelength acceptance band. That band is centered near 470 nm and has a sharp red-side edge near 500 nm. Irradiance values were recorded as integers on a logarithmic scale that had 15 units change in reading per decade change of irradiance. This was transformed during later floating point calculations to a scale with 32 units to the decade for compatibility with later model tags and is plotted on that scale in the figures. The logarithmic slope was verified by manufacturer's production test to be constant over the entire usable irradiance range and the sensor is temperature-compensated so that above the sensitivity floor the irradiance value reported does not change with changes in temperature. However the level of the floor did vary by an estimated 1.5 decades with seasonal outdoor temperature variations, improving as the temperature decreased. In cool weather the intensity scale extended nearly ten decades below nominal clear-day noon sunlight intensity. Since the range of irradiances involved in the present study was limited, and since when the sensitivity floor could be seen in the data from very dark nights it was many decades below the range of interest for this study, no effort was spent to determine its exact location.

The threshold geolocation method described below was carried out internally by the tag every UT midnight based on its own private data record of 128-s observations that was erased each day after the results of the calculation were recorded. The threshold-angle pair used was that illustrated in Fig. 3 below, a threshold down from midday sun irradiance (more precisely the 92nd percentile of the day's data, approximately midday median) by a factor of 1000 and a corresponding assumed solar elevation angle of –3.44°. The time series log was configured to record exactly the same data used by the internal algorithm, making it available for parallel treatment by another method. The tag clock is automatically corrected for temperature, and is specified to drift less than 30 s per year. This particular tag's clock has drifted less than 15 s in the last several years. Accordingly the scheduled observation time of each measurement is used directly as the time-of-day associated with each observation. The common practice of linear interpolation between beginning and ending clock errors was not employed here.

The particular data used below were taken as part of a large reference data set being accumulated on an ongoing basis. Since none of the work which will be done with this data makes any use of absolute irradiance values (as opposed to irradiance changes over a day), no absolute sensitivity calibration was performed either before or after exposure. Time series data are read out approximately every 80 days, since this particular older design tag has time series memory limited to about 82 days sampling at the interval chosen.

The particular data segment used below corresponds to the period between two such read-out events and was initially chosen essentially at random (its name, the date 00Jan15, sorted first in a computer directory) as test data for an initial MathCAD (Version 2001Pro) implementation of the template fit method presented below. (MathCAD, Mathsoft Engineering and Education, 101 Main Street Cambridge, MA 02142-1521, USA, www.mathsoft.com.) It turns out that the particular read-out event defining the end of that observation period was late, and as a result of the way the tag was configured that caused a gap several days long in the data near the center of the period. When the method being tested proved interesting, it was noticed that the data set being used happened to include both favorable and unfavorable seasons, so was used for illustrating the method in this publication.

Template fit results shown are the output of MathCAD programs used for initial exploration of the method, and are among the very first results obtained with the method. A more extensive evaluation of the method awaits its implementation in a way more convenient for processing larger data sets. Please note that the initial results presented here are illustrative and not intended as a complete and definitive comparison of threshold *vs*. template methods, or even of these two particular methods. As one example, the possible greater fragility of the template method has not been tested against the highly correlated noise introduced by, for instance, depth-induced irradiance changes imperfectly corrected. That remains for the future.

Threshold methods

Historically, light-based geolocation aboard a diving animal has been done in four steps:
1) make a sequence of simultaneous light and depth measurements,
2) infer from these the sequence of surface solar irradiances throughout the day,
3) infer from that sequence the times of sunrise and sunset,
4) infer from those times the latitude and longitude.

Light-based animal geolocation can be traced at least as far back as Smith and Goodman (1986), with other published accounts by DeLong *et al.* (1992), Wilson *et al.* (1992), Hill (1994), Welch and Eveson (1999), Metcalfe (2001), Hill and Braun (2001), and Musyl *et al.* (2001).

In most cases, the time of sunrise or sunset has been taken as the moment when the inferred surface irradiance crosses some threshold value; thus we will refer to these methods collectively as threshold methods. To make the third inference from light to sunset time in a threshold method, one must choose a threshold intensity value to indicate that the sun is at some particular solar elevation angle $\mathbf{a}$ (or zenith angle $\mathbf{z} = 90°–\mathbf{a}$) that is taken to represent sunset. For reasons that will be explored below, this association is a weak link in the logical chain. Variants that infer from irradiance to angle first, and then apply a threshold to the inferred angles are equivalent as far as this difficulty is concerned.

In sharp contrast, the last step has a strong theoretical basis. It is all astronomy, and the relevant relationships applicable to all geolocation methods discussed here can be found in standard references (Astronomical Almanac, current year)*:

$$\sin \mathbf{a} = \sin \delta \sin \phi + \cos \delta \cos h \cos \phi, \qquad (1)$$

$$h = \mathrm{UT}/4 - L - E$$

where **a** is the solar elevation angle, δ is the solar declination, ϕ is the observer's latitude, h is the solar hour angle in degrees, L is the observer's longitude in degrees, E the astronomical "equation of time" in degrees and UT is the universal time expressed as minute of day. Note that $\sin \mathbf{a} = \cos \mathbf{z}$, where $\mathbf{z} = 90° - \mathbf{a}$ is the zenith angle. Since δ and E depend only on the date, if the elevation angle **a** is known then one can in principle solve for the latitude ϕ and the longitude L given two observations of the times (UT) when the sun is at the same elevation angle **a** and widely varying h (sunrise and sunset). The longitude one obtains depends primarily on the average of the two observed times (time of solar noon or midnight), the latitude on their difference (day length). This, in short, is the astronomical basis of all threshold methods.

In practice, however, there are difficulties, especially regarding latitude. When the assumed angle **a** corresponding to the threshold is nonzero and the season is near enough to the equinox so that $|\mathbf{a}| > |\delta|$, then there is an even number of solutions for the latitude, and no unique result for ϕ. If a solution can be found at all, then besides the true solution there is also an additional one that corresponds to the same day length but not to the correct position. Also whenever the declination δ is near zero (near one of the solar equinoxes) the day length depends only weakly on latitude. As a result, a small error in determining day length can lead to a very large error in the implied latitude. In extreme cases there may be no solution to the relevant equations. Equivalently, any small error in determining the actual sun elevation angle **a** that corresponds to an adopted threshold will be greatly magnified near the equinoxes. Figure 1 shows the characteristic curve of inferred latitude error that accompanies such an error in threshold or elevation angle. Reversing the sign of the error in elevation angle reverses the sign of the latitude error.

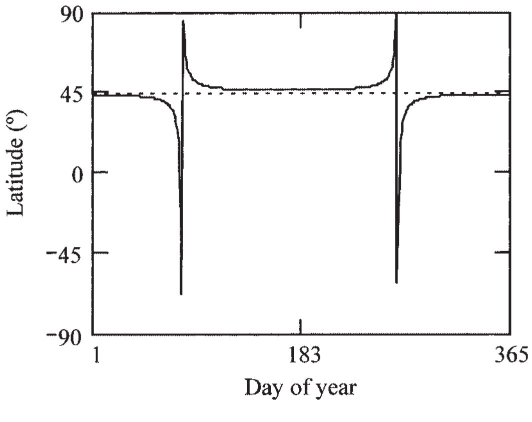

Fig.1. Latitude inferred from day length. Dotted curve: correct latitude. Solid curve: latitude deduced when the elevation angle assumed to correspond with the irradiance threshold is wrong by −1.5°.

* Copies are available from the US Government Printing Office or in England from Her Majesty's Stationery Office.

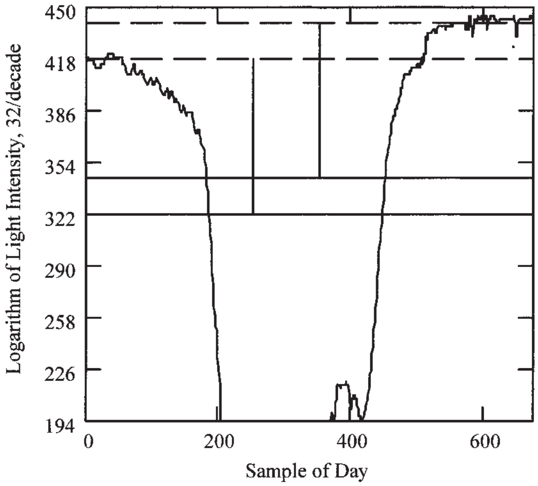

Fig. 2. Light data from April, 2000, Shaw Island, WA. Y-axis ticks indicate decades of irradiance. Dashed lines are example intensity references matched to the two partial days shown, solid lines three decades lower are corresponding thresholds. The lower threshold leads to an apparently longer day or shorter night.

Unfortunately, there is no universal relationship between the value of sea-surface irradiance and the solar elevation angle **a**, so that there is no universally valid way to associate a threshold with an elevation angle. The wide variety of existing threshold methods might be regarded as a symptom of this awkward fact. Most obviously, an overcast day will be darker than a clear one, and on such a day a given irradiance will be reached when the sun is higher in the sky than it would be if the sky were clear. Clouds advance the apparent time of sunset and they delay apparent sunrise.

Clouds are more harmful to determinations of latitude than of longitude. Even in partly cloudy weather, as displayed in Fig. 2, the sunrise and sunset transients will be reasonably symmetrical (not perfectly so, of course, because of weather). Therefore an erroneous shift in threshold that moves a sunset later will move the corresponding sunrise earlier. When the two are averaged to form midnight or noon, those shifts tend to average out. In contrast, the times of sunrise and sunset are subtracted to obtain day length. A threshold error induces shifts in these times that are in opposite directions and because of this subtraction they will combine to strongly affect apparent latitude. Choosing the threshold deep into twilight where the variation of irradiance with time is as steep as possible can minimize these shifts, but cannot eliminate them.

Figure 3 shows 80 days worth of surface irradiance data measured at the Shaw Island site and described above. The data were first divided by the 92nd percentile of each day's data (a robust measure of midday irradiance) to remove effects of instrument sensitivity and average cloudiness, and then plotted *vs.* solar elevation angle calculated from the known location, time-of-day and date. The traces cross the two index lines at points spanning at least a decade in irradiance and several degrees in elevation angle. We can readily see that even when normalized to midday irradiance, there is no unique relationship between elevation angle and irradiance. Although not shown here, the situation is worse before normalization.

Within the framework of a threshold method, there is no clear best thing to do about this

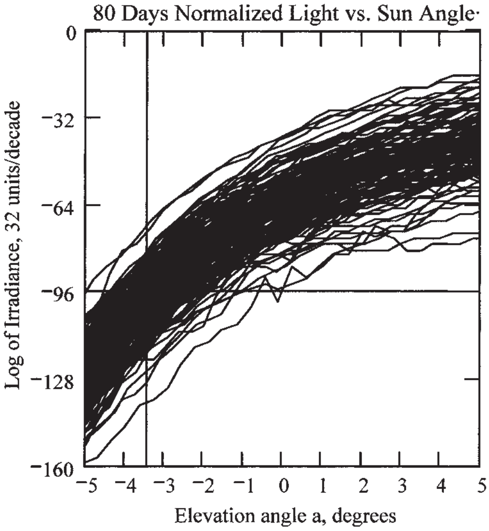

Fig. 3. Irradiance data normalized to the 92nd percentile of the day's data, a measure of noon irradiance. See text regarding index lines. The curves cross the irradiance index line at **a** = –3.36 ± 0.75°.

situation. One possible approach is the following one: First choose an intensity threshold with respect to some reference intensity. We note from Fig. 3 that the average curve has become about as steep as it will by the time the irradiance has fallen a factor of 1000 from its noon value, taken as the day's 92nd percentile of irradiance data. A steep curve is desirable since it translates a given error in irradiance or threshold position into a relatively small error in angle. The sun elevation angle corresponding to that threshold can be seen to be about **a** = –3.5°. This value can be refined by taking a mean of calculated elevation angles at irradiance threshold crossing for data regarded as typical and taken at a known location.

Another possible refinement procedure makes use of field data that spans an equinox taken at a fixed site, but does not require knowledge of the site's latitude. When a threshold method is applied to such data, if the assumed elevation angle **a** corresponding to the threshold is not correct, the distinctive variations in apparent latitude shown in Fig. 1 will appear in the result. If the summer results are more northerly than the winter results (as shown in Fig. 1), the angle has a negative error. In the reverse situation, the angle is too high. One adjusts the assumed value of **a** and re-runs the analysis for all days in the data set seeking to minimize the amplitude of the error. When this process was carried out for a three-decade irradiance threshold using data from the Shaw Island site taken in 1994, the value of **a** = –3.44 resulted. That is the source of the threshold-angle pair used in this work.

Figure 4 shows position results calculated by a threshold method using the threshold-angle pair illustrated by the cross-lines in Fig. 3 and using only the latter half of the data plotted in Fig. 3 (avoiding the equinox for the moment, which occurs in the first half). The dotted diamond represents the mean ± one sample standard deviation in both latitude and longitude. Note that graduations on the latitude scale are each worth two degrees, compared to one degree on the longitude scale; as we might expect, latitude performance is typically poorer than longitude performance. Crossed index lines here represent the known actual location where the measurements were made. Note that the corresponding diamond for standard deviations in the mean would be smaller by a factor of $\sqrt{40} = 6.3$. The usual 3σ extreme error bar would not include the true position in latitude, indicating that there is a systematic error in the

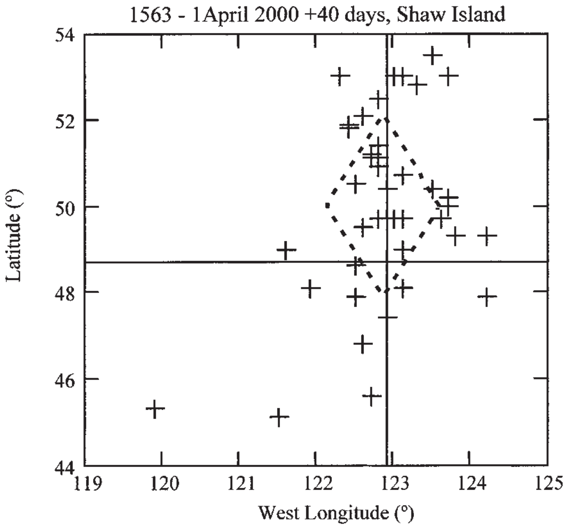

Fig. 4. Geolocation results using the threshold indicated. Index lines indicate the known location on land, the dotted box shows the mean ± 1 standard deviation.

latitude determination. In contrast, the longitude mean is very near to the true value.

These results will eventually serve as the "before" part of a "before and after" comparison, representing the performance of one particular threshold method, and by implication providing some indication about those methods as a class. It is a relatively simple method with a fixed threshold-angle pair chosen once for all situations, and may be considered representative of fixed-threshold methods. Hill and Braun (2001) have pointed out that there is advantage to be gained in latitude determination by using different threshold values during different times of the year and at different latitudes. This method does not in general avail itself of that advantage. However the comparison below is performed on the summer side of equinox when the low elevation angle (large zenith angle) that corresponds to the threshold chosen is in the direction Hill and Braun (2001) would recommend. Since that comparison is intended to be illustrative rather than quantitative, we leave the issue there after noting that the sensor characteristics and sampling interval employed have been chosen so that the noise they introduce is normally dominated by expected day-to-day natural variations in the data being gathered.

A geophysical model of twilight

While the many data curves of Fig. 3 do not lie on top of each other as we might wish, most of them do have a remarkably similar shape. In fact, they may be brought much more closely into coincidence by simply shifting them up and down (on a logarithmic scale, equivalent to multiplication by a constant) to compensate for differences in cloudiness. The transients tend to have a stereotyped shape, and it turns out that one can understand this shape.

A simplified radiative transfer model (Ekstrom, 2002) predicts that for sufficiently strong light scattering (by atmospheric dust, etc.) the sunrise-sunset transient should have a particular rigid shape vs. solar elevation angle for angles in $-5° < \mathbf{a} < +3°$. This is the approx-

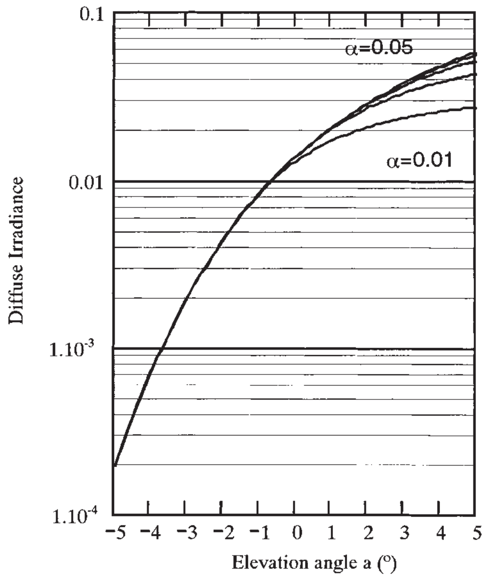

Fig. 5. Theoretical curves for blue irradiance during twilight. For elevation angles less than 3° and scattering stronger than $\alpha = 0.03$, the curves merge.

imate range of angles for which surface irradiance is dominated by light that has been singly scattered in the stratosphere and which conforms to a simple model. For higher angles, light coming directly from the sun begins to contribute significantly; for lower angles, multiply scattered light (ducted over the horizon) will contribute. In neither of those bounding cases does the curve of irradiance *vs.* angle exhibit such an invariant shape. One could wish that the universal part of the sunset curve extended to lower sun angles; Hill and Braun (2001) point out that valuable latitude information can be found at sun angles lower than –5°.

The relevant theoretical curve for the single-scattering regime is plotted in Fig. 5 for several values of the scattering parameter α. The small particles that are important in the upper atmosphere scatter light more strongly at shorter (bluer) wavelengths. The statement $\alpha = 0.03$ km^{-1} means that an incident pencil beam of light will lose 3% of its irradiance for every km of surface-equivalent path through the atmosphere. One "surface equivalent km" of path length is actually about 10 km long in the low atmospheric density found at the altitude of peak scattering. For research purposes the scattering parameter α can be determined by fitting the complete model to a complete day's data. The value one obtains is most strongly determined by data taken when the sun is higher than 3° elevation angle that is not normally used for geolocation. Data taken near 470 nm, as dictated by the needs of correction for depth, is normally well fit by the comfortably large value $\alpha \geq 0.03$ km^{-1}. As a result, analysis of twilight field data taken in the blue can simply use the common curve that applies for all large α. For less-strongly-scattered light of longer wavelength, the expected value of α is smaller and the situation is more complicated. Thus the properties of both the atmosphere and the ocean fortunately agree in rewarding the choice of blue light measurements as a basis for geolocation.

The physical picture may be summarized as follows: when the sun is near the horizon (for solar elevation angles **a** near zero), light on a direct route from the sun to the earth's surface must travel "sideways" through the atmosphere nearly parallel to the surface. That light sees an effective (density weighted) atmospheric path length of ~160 km, compared with ~4 km for light traveling directly downward. Rays traveling such long paths through the dense lower atmosphere lose essentially all of their blue light (thus the sunset horizon is red). Rays that pass overhead within a certain range of altitudes have traveled very nearly the same geometric distance but through much less dense high-altitude air. They are just then losing their blue light. About half of that light is scattered downward toward the earth's surface (thus the sky is blue). Light passing even higher overhead encounters air even less dense. It proceeds unhindered and unnoticed.

In such a situation, most of the light that is scattered through any large angle will travel only a short additional distance before being lost either into space or by absorption at the surface. As a consequence, the path geometry and the exponential density structure of the atmosphere dominate the situation in a direct way, and a relatively simple model can capture the dependence of surface irradiance on sun elevation angle.

During twilight a sensor filtered to accept only blue light will see only light from the sky overhead. That light is scattered at high altitude: the strongest scattering occurs at 10 km altitude when **a** = 3°, rising above 16 km when **a** = 0°. Incident light that is scattered there has never been near the earth's surface, never low enough to encounter any horizon feature. It will pass over the tallest mountain and the tallest tropospheric cloud on its way to the scattering point. Nor will this incident light be significantly refracted in the low-density air along its path at those altitudes. Thus a tag that geolocates based on measurements of blue light made for sun angles less than 3° will not be subject to either "horizon effect" or atmospheric refraction. This benefit is available to both old and new geolocation methods; however, there is one more feature that principally benefits a method based on the shape of the twilight transition.

It is only on the last few km of the path, as the light is on its way down to the surface after scattering overhead, that we expect any significant effect due to weather. To the extent that the average effective sky coverage by separate clouds or the effective attenuation due to a deck of clouds is constant during the sunrise or sunset transient, we can expect the transient's shape to approximate the theoretical ideal, whatever its overall intensity. This, then, is what underlies the stereotyped shapes in Fig. 3. What use can we make of it?

A comment on the literature

These results affect the interpretation of previous work, in particular the important paper of Hill and Braun (2001) referred to several times here. Since atmospheric refraction is suppressed by a factor near ten at the height of peak scattering when compared with the surface values they use, the refraction errors assumed in that work should be reduced by that factor, and then cannot account for the errors observed in practical geolocation. The apparent agreement found was coincidental.

In fact, it appears that weather dominates geolocation errors through its effect on surface irradiance rather than through its effect on refraction. In order to make use of the Hill and Braun analysis, irradiance variations can be approximately translated into apparent angle variations using the slope of the light curve (Fig. 5) but we find that the slope at **a** = –5° (**z** = 95°) is about a factor of 5.4 steeper than it is at **a** = 3° (**z** = 87°) leading to an apparent angular variation corresponding to any given irradiance variation that is smaller by that factor. The Hill and Braun (2001) paper assumes equal angular errors for all thresholds (themselves expressed in terms of angles, not measurable irradiances). If we are to apply their analysis to errors that arise in irradiance changes, the error curve for high sun angles must be inflated by that factor of 5.4 before being compared with a curve for a low sun angle. Intermediate angles require intermediate adjustments.

Nothing in the above calls into question the seasonal variation of error pointed out in that work, and it is to this aspect that we ordinarily refer here. Indeed, the fact that during the difficult time just on the winter side of equinox the best latitude information is to be found in data at high sun angles, combined with the fact that irradiance data at high sun angles are so badly obscured by noise when seen in terms of angle, provides a possible way to interpret the difficulty that all methods seem to experience in determining latitude at that season.

An initial example of template-fitting to irradiance

A common approach to data analysis begins with a model and fits it to a set of data, adjusting model parameters to minimize some error measure. In favorable circumstances, there is only one choice of parameters that produces a good fit, and that choice of values is taken to represent that data set. When the model expresses a distinctive shape for a data curve, a template, the process can be called a template fit.

Such methods are not new to geolocation; the "reflection method" or "dawn and dusk symmetry method" for longitude introduced by Roger Hill (Musyl *et al*., 2001) uses a reflected copy of the irradiances making up a given day's sunrise transient as a template that is fit to the same day's sunset transient. The parameter being adjusted is the assumed time of noon. While it deals with longitude only it does fit directly to irradiances, was known to the author during formative stages of this work and can be considered an ancestor of the method developed here.

Musyl *et al*. (2001), briefly described another template method for latitude that fits to angle instead of to irradiance. In the method described, some particular day's light data were used to translate irradiance to sun angle; how the possible order-of-magnitude variations of overall illumination due to day-to-day variations in weather were handled in making the angle inference is not stated. The method then chooses a latitude value so as to minimize the difference between theoretical and inferred sequences of angles. Unfortunately the more detailed explanation "elsewhere in [that] volume" is not to be found.

The approach here is different in three major ways. First, we fit directly to measured irradiances taking as our template the theoretical shape of the sunrise-sunset irradiance transient discussed above, second we limit the range of angles used to those for which both theory and field data agree on the template's validity, and third we take overall illumination as a free parameter that may be called "cloudiness", thus removing the effect of any slow variations in weather. The shape as a function of sun elevation, $F(\sin \mathbf{a})$, arises explicitly in the twilight model. The full model, with parameters determined by a fit to a variety of field data suggests $-5° < \mathbf{a} < 3°$ as the range in elevation angle over which the curve can be considered a robust model of nature. One substitutes in $\sin(\mathbf{a})$ from eq. (1) and includes two intensity multipliers to re-express the template as a function of latitude, longitude, and morning and evening cloudiness, plus the independently determined (not fitted) parameters δ and E that depend only on the date of the observations.

It is the logarithm of F that is actually used in the fit, not F itself; one fits to the data expressed on the same logarithmic scale in which it is measured and logged. The fundamental reason for this is that all the processes giving rise to intensity behavior, both signal and noise, are multiplicative. As we pass though the relevant portion of twilight, irradiances range over more than two orders of magnitude. Over this entire range the data's rate of change, the natural variations, and the measurement noise are all roughly proportional to the intensity being measured at any given moment. The template's sensitivity to parameter values is also roughly proportional to its local value. When formulated on a linear scale, straightforward application of least-squares fitting theory (Press *et al.*, 1988) to this situation leads to a weighting function near inversely proportional to the template value squared. Changing to a logarithmic scale flattens the optimum weighting function to an approximate constant, thus identifying that latter scale as the natural way to express the situation. This simplification in weighting function, which can then be implemented as unity, is one of several practical benefits of working on a logarithmic scale that are especially valuable when doing a fit in the field inside the small processor aboard a tag.

The result of such a fit is shown in Fig. 6 where the data are the same as we saw in Fig. 2. Note that the morning cloudiness value differs from the evening one. The horizontal lines connecting the ends of the template curves to midnight and midday are there to guide the eye and to make that offset obvious. The data are from one of the days contributing to Fig. 3, chosen to show the method noticing a cloudy sunset and a brighter sunrise. The fortunate fact that the template is curved makes this distinction possible. Because of that curvature, the template will ordinarily fit best at only one place in a data record despite the stretching that is available to the fit-finder through changes in assumed latitude.

As a result one can separately model the cloudiness of morning and evening, which means that the method is robust against slow changes in weather. It is only the rapid changes occurring actually during twilight that can distort the shape of the light curve and cause errors. Such changes do, of course, occur and they do represent a danger to this shape-sensitive method.

In addition to matching the correct day length, a good fit must also match the slope of the sunrise and sunset transients. The sun appears to set more slowly at high latitudes than at the equator; this additional information is ignored by a threshold method but as an inherent

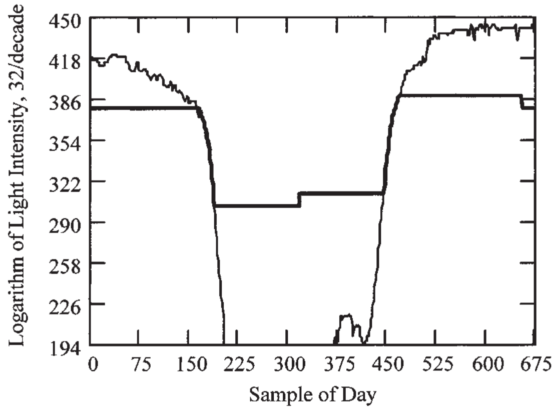

Fig. 6. The heavy line is a template based on the curve of Fig. 5 fit to the data of Fig. 2 (lighter line) over the range $-5° < \alpha < 3°$.

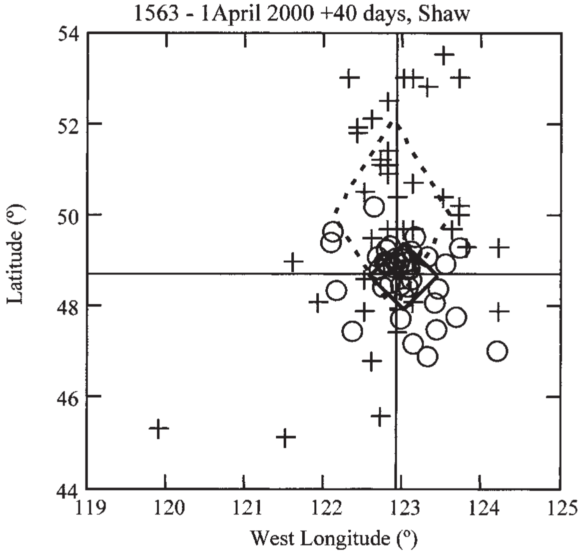

Fig. 7. Crosses, dotted box, and index lines are as in Fig. 4, circles are geolocation results from a template-fit method, solid box represents mean ± 1 standard deviation.

feature of a template fit method it will be used to help determine latitude. However this use of slope information means that the light sensor must have an accurately known and stable, including temperature-stable, logarithmic slope.

Figure 7 shows the threshold-method results of Fig. 4, along with added template-fit results generated by re-processing the same time series data with a semi-automatic template fit program implemented in MathCAD. The new smaller solid diamond represents the mean ± one standard deviation of the template fit results. As one can see, not only are the data variations substantially smaller, but also the latitude mean is closer to the known correct answer. In fact in this example the agreement in the mean appears to be fortuitously good as it is better than the sample standard deviation in the mean would lead one to expect very often.

The new method has no adjustments. The program used to generate the results of Fig. 7 gave somewhat better results with this data when days were taken to begin and end at noon instead of at midnight, and the reason for that is not clear. Other than making the more favorable choice in that regard, neither method was "tuned" in any way for this comparison. Indeed there is no obvious way to "tune" the template fit; if the results had turned out poorly there is little that could have been done to improve them other than making sure that the fit had converged properly. The template fit method with a geophysically based template differs most from any of the threshold methods in just that regard, and its rigidity is an important advantage. The whole puzzle about how to choose thresholds and their corresponding angles is gone, replaced by a physical model requiring as input only the atmospheric scale height and the radius of the earth. Both of those quantities are independently measurable; the method's results are not adjustable to accommodate our preferences, whether intentionally or unintentionally.

Theoretical expectations

So far in this paper the method has been tested against data measured at a single location on land at a favorable time of year. The season was restricted deliberately to postpone some difficulties that we must now consider. In this investigation we can use our geophysical model to predict the light data curves to be expected under other circumstances, and can combine these curves with the error estimates that are associated with the least-squares algorithms used in the fit. Doing so lets us explore other situations theoretically. It turns out that when determining latitude, the new method has its own blind spots at the equator and slightly to the winter side of each equinox.

Standard approaches for estimating the errors in least-square fits are fully applicable and have simple interpretations only when the data being analyzed are statistically independent and have Gaussian-distributed errors of known variance. Irradiance data from the field measured aboard a diving animal and corrected to estimate surface irradiance satisfy none of these conditions. They are affected by medium-scale non-Gaussian variations that affect in a correlated way a sequence of measurements that might include some but not all of a twilight transient. These can be due to both clouds and to imperfectly compensated animal motion. Thus what we can hope to obtain from the error treatment below is only a kind of uncalibrated error sensitivity function, not an absolute error estimate. Further, the sensitivity function is strictly valid only for a class of data errors different from the ones we expect. Even the shape of the plots we can generate will not be precisely correct, but they will be a useful qualitative indication of where to expect the accuracy of the method to be especially good or especially bad.

Proceeding as usual (Press *et al*., 1988) with error estimates for a nonlinear least-squares fit, we take as a measure of error sensitivity for each parameter the corresponding diagonal element of the covariance matrix of the locally linearized approximation, evaluated for some particular choice of latitude and sun declination (representing season). An important assumption we make in doing this is that the fit process producing the parameter values has converged well enough so that we are in a region of parameter space that includes the true parameter value and over which the model is locally linear. As we shall see, there are situations where the fit does not converge correctly and where, as a consequence, this assumption is not justified. We will find that the regions of parameter space where this can occur are indicated in a general way by large values of the error sensitivity measures. We can thereby locate them, but we must experiment to learn their exact extent.

Finally, the presence of the cloudiness parameters and the longitude in the model profoundly affects the structure of the covariance matrix and the behavior of all the matrix elements. No parameter determined from the data may be omitted from the model either when fitting or when generating error estimates. Doing so would amount to claiming some independent way of measuring the omitted parameters that did not use the same data; it would produce much more favorable error estimates that would not apply. However, when all parameters are included in the model, we find that the resulting matrix elements still depend on the values of only the latitude and declination, so the particular values of the other parameters can be ignored when plotting the results. Thus we can plot the error measures against only two parameters, latitude and sun declination.

The error measure for longitude is shown in Fig. 8, where to avoid obscuring our view

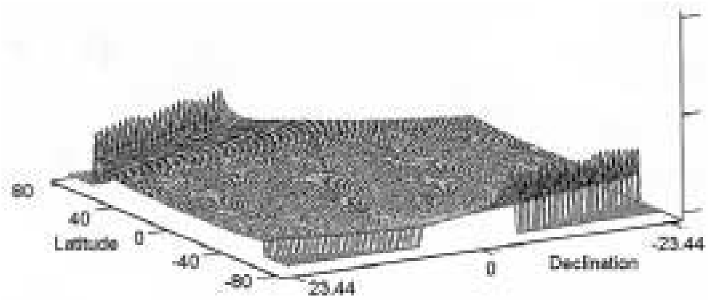

Fig. 8. Error measure for longitude *vs*. latitude and declination.

of the surface, the measure has been set to zero in corners of the parameter space where it begins to rise rapidly. These regions correspond to the polar summer and winter when and where the sun does not traverse a substantial portion of the range of elevation angles $-5° < \mathbf{a} < 3°$; that is, for conditions of midnight sun or midday dark. Except in those extreme regions, the error measure is nearly constant. As a result, we expect that our longitude estimate will be well behaved for all latitudes and seasons, so long as the sun actually does rise and set.

The corresponding measure for latitude is shown in Fig. 9. It is drawn to the same vertical scale, but in neither case are divisions shown on that scale because these curves do not apply, strictly speaking, to our case. There is reason to hope that their shapes are approximately correct, but a valid numerical scale cannot now be established for the non-Gaussian, correlated data we will find in practical applications. Still, even without an absolute calibration we can see that the error becomes substantially larger near the equinox and near the equator, and that the smallest errors in all seasons are found in the polar regions. The large peak centered at the intersection of equinox and equator is flattened, and is twisted toward the winter side of the equinox.

A more involved analysis indicates that in the regions where large latitude errors are predicted, the change in expected light data corresponding to a variation in latitude becomes very small and the pattern of that variation becomes very similar to the pattern of a change in cloudiness. In this situation the method has difficulty distinguishing changes in latitude from changes in cloudiness. Indeed, if the cloudiness parameters are removed from the fit, the broad column obvious in Fig. 9 shrinks to a narrow spike and the analysis predicts good performance almost everywhere. If only there were another way to determine cloudiness, one

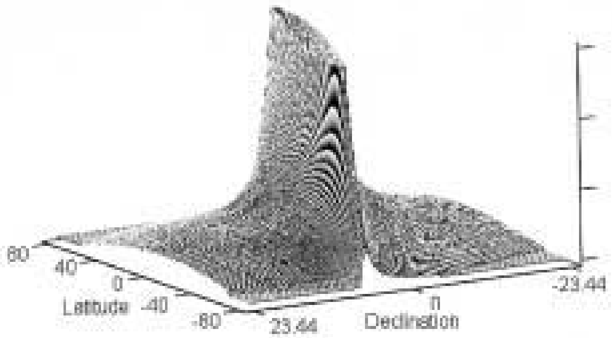

Fig. 9. Error measure for latitude *vs*. latitude and declination.

could look forward to great improvements; however, within the present analysis of the present data, the problem is inescapable. The present version of the method avoids data taken at elevation angles above 3° that might provide help in determining cloudiness for the excellent reason that the direct beam begins to be important and singly-scattered light with its stereotyped light curve is no longer the only significant contribution to irradiance. The situation becomes far more complex and much of the robustness against weather is lost. Still, that portion of a day's data does have additional information that might be used to bound the cloudiness, and perhaps future developments will allow its use to improve the latitude performance in those situations where the present method by itself is least effective.

A fully automated approach

A practical template-fit implementation should operate without human intervention at least to the point of producing its position estimates. If it is to run in the tag itself and serve as a field data compressor, then it must operate without help. As part of autonomous operation, it must find some way to start the fit with parameter values that are near enough to the correct solution so that the fitting process will converge.

The only relevant position information likely to be available is that derived from a threshold method and, as we have seen, the latitude values derived from such a method are subject to large errors in some cases. Fortunately, just when the latitude performance of a threshold method is at its worst, the template method requires only a longitude value to align the template in time-of-day with the relevant portion of the data curve, and a threshold methods can still determine longitude. To guarantee that the relevant equations in the threshold method have unique solutions for (nearly) all latitudes and seasons, the irradiance threshold employed should correspond to $\mathbf{a} = 0$.

Still, the closer the starting point is to the correct latitude, the more likely is the fit to converge correctly. In particular, near the equinox there is little variation in day length with latitude; the best latitude information available to the template-fit method comes from the slope of the sunrise and sunset transients. Since the sunrise-sunset slope is an even function of latitude, near the equinox there will be two values of the latitude parameter, one in each hemisphere, at which a locally optimum fit can be found. If the solution is started in the wrong hemisphere, it may converge to the local minimum in fit error that is to be found in that hemisphere. This would not be a problem if one knew that the animal would stay in one hemisphere and were willing to pre-set the method to fit in only that hemisphere, but a general method should be able to follow an animal across the equator.

One possible response to this situation begins by solving for latitude in the threshold method using the largest and smallest values of day length (or equivalently of elevation angle) that would be consistent with the expected data errors, including those variations illustrated in Fig. 3. If the resulting two threshold-method latitudes lie in different hemispheres, then start a fit in each hemisphere and report both results for later resolution, perhaps on the basis of either fit residual or track continuity.

One experiment whose results are shown in Fig. 10 always started two solutions, putting one starting point in each hemisphere when that was indicated, and putting both at different latitudes in the same hemisphere whenever there was no uncertainty regarding hemispheres. This starting strategy was combined with the Levenberg-Marquardt method (Press *et al.*,

1988) for non-linear least squares fitting and applied to the entire run of 80 days data we have been considering. The right-hand half of the figure corresponds to the data used to generate Fig. 7, the left hand half corresponds to the period at and just before (on the winter side of) the solar equinox (see trace a). The results of the two separately-started fits for each day are shown as curves c and d, which we hope will coincide and which in favorable cases usually do so.

In the well-behaved right half expected to be typical of most of the year, the two fits almost always converge to the same values, while in the left half where theory leads us to expect trouble, and where for some dates the two solutions were started in different hemispheres the two results are sometimes quite different. There is a region where we can see the southern-hemisphere solution following the predicted spurious southern-hemisphere error minimum as it moves south in response to changing sun declination. In this difficult region, even when the two solutions agree, sometimes neither one is a very good approximation to the true position.

Trace b is the same error measure plotted in Fig. 9, but evaluated for the latitude determined by the fit. An error estimate evaluated at the true position would be preferable, but the true position would not be independently known when processing real animal data. A convenient vertical scaling has been chosen for it to make its variations display well. This error measure becomes large as a warning in just that season when the misbehavior occurs. Thus it promises to be a useful tool in evaluating the reliability of fit results, once experience is acquired with it when applied to real animal data.

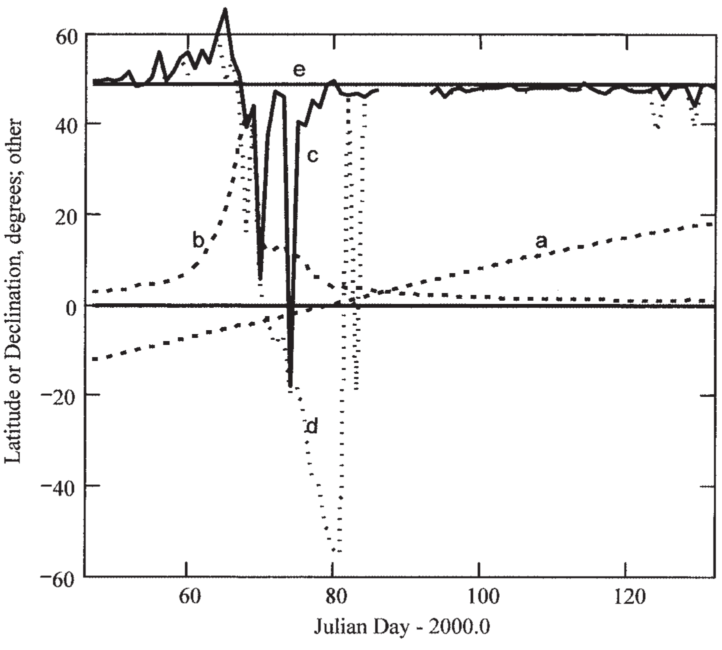

Fig. 10. Results of a fully-automated fitting experiment. a: solar declination, b: uncalibrated error sensitivity measure, c: latitude from fit 1, d: latitude from fit 2, e: known latitude of observations. Note gap in the latitude traces near day 90 caused by a gap in the test data.

Discussion

The method is not fully developed in its practical aspects, and has not been tried against either surface light data from more than one location or against field data from a diving animal. Its use of nonlinear least-squares as an integral part of each day's calculation can be expected to result in brittle behavior not part of past experience with threshold methods, and this will need to be explored and understood. The calculations reported here have all been done using MathCAD, a tool more suited to initial exploration of the method than to its practical implementation. Work is underway on a more general program for reprocessing field data records on a desktop computer. Further development and testing of the method will take place using that program. A version is also being pursued that will do the daily fits in real time aboard an archival (data storage) tag, allowing the method to be used in a small-memory miniature tag, for long missions at high data rates in any tag, and where data must be transmitted—situations where a complete time-series record is not always available on shore.

The availability of a simple geophysical model of twilight has led to a new geolocation approach that seems broadly applicable but is especially attractive for use at high latitudes. The method is not fully developed; early results are encouraging.

References

Arnold, G. and Dewar, H. (2001): Electronic tags in fisheries research: a 30-year perspective. Electronic Tagging and Tracking in Marine Fisheries, ed. by J. Sibert and J. Nielsen. Boston, Kluwer, 7–64.

DeLong, R.L, Stewart, B.S. and. Hill, R.D (1992): Documenting migrations of northern elephant seals using day length. Mar. Mammal Sci., **8**, 155–159.

Ekstrom, P.A. (2002): Blue Twilight in a Simple Atmosphere. Proceedings of SPIE, conference 4815, paper 14. SPIE Conference proceedings can be ordered from the Society of Photo-Optical Instrumentation Engineers web bookstore, www.spie.com. The single article is also available at http://www.lotek.com /library.htm#irrad

Hill, R.D. (1994): Theory of geolocation by light levels. Elephant Seals: Population Ecology, Behavior, and Physiology, ed. by B.J. Le Boeuf and R.M. Laws. Berkeley, Univ. California Press, 227–236.

Hill, R.D. and Braun, M. (2001): Geolocation by light level—The next step: Latitude. Electronic Tagging and Tracking in Marine Fisheries, ed. by J. Sibert and J. Nielsen. Boston, Kluwer, 315–330.

Metcalfe, J. (2001): Summary report of a workshop on daylight measurements for geolocation. Electronic Tagging and Tracking in Marine Fisheries, ed. by J. Sibert and J. Nielsen. Boston, Kluwer, 331–342.

Musyl, M., Brill, R., Curran, D., Gunn, J., Hartog, J., Hill, R., Welch, D., Eveson, J., Boggs, C. and Brainard, R. (2001): Ability of archival tags to provide estimates of geographical position based on light intensity. Electronic Tagging and Tracking in Marine Fisheries, ed. by J. Sibert and J. Nielsen. Boston, Kluwer, 343–368.

Press, W., Flannery, B., Teukolsky, S. and Vetterling, W., ed. (1988): Numerical Recipes in C: The Art of Scientific Computing. New York, Cambridge Univ. Press, Chapter **14**, 540–558.

Smith, P. and Goodman, D. (1986): US National Oceanic and Atmospheric Administration, technical memorandum NOAA-TM-NMFS-SWFC-60.

Welch, D. and Eveson, J.P. (1999): An assessment of light-based geoposition estimates from archival tags. Can. J. Fish. Aquat. Sci., **56**, 1317–1327.

Wilson, R.P., Ducamp, J.-J., Rees, W.G., Culik, B.M. and Niekamp, K.(1992): Estimation of location: global coverage using light intensity. Wildlife Telemetry, ed. by I.G. Priede and S.M. Swift. New York, Ellis Horwood, 131–134.