Analysis of dosimeter badge data

A. James Liley[1], J. Ben Liley[2]

[1]University of Otago, Dunedin, New Zealand
[2]National Institute of Water and Atmospheric Research (NIWA), Lauder, Central Otago, New Zealand

Abstract. The study of personal UV exposure and vitamin D production using the Allen dosimeter badges required complete data series for all participants, but this was rarely obtained for the 8 or 10 weeks participation. Instead, a range of problems led to gaps in the data that had to be filled. The process required several techniques including heuristic search algorithms, statistical data processing, and supervised and unsupervised machine learning. The three steps in the analysis were: 1) Find and remove false and unreliable measurements; 2) Interpolate available data across areas where measurements were missing; 3) Reduce the data to a usable and presentable format. Ultimately, each participant's data were reduced to 1680 hourly values of average UV, which could then be combined with thrice daily records of clothing worn, daily dietary intake, and subsequent serum 25OHD.

Badge data series

The New Zealand electronic dosimeter badges are described in the paper by Allen (2010). In the New Zealand UV-vitamin D study in 2008 and 2009, the badges were worn by 517 participants for 8 or 10 weeks, and set to record every 8 seconds from 06:00 to 22:00. To quantify cumulative UV radiation scaled by skin exposed, the study required complete time series. Instead, a range of badge or operator errors interrupted the individual data series for hours, days, or even whole weeks. Even where data were apparently recorded, there are instances where they are clearly erroneous. Some are much larger than possible for the time of day, supposedly start at the wrong time, or are repeats of earlier data. This could happen because a badge failed to restart, or the data marker that separates new and old data in memory was missing. Static electricity, and poor or bounced battery connections, could also cause data gaps. All of these factors have been considered in updating the badge design (Sherman 2010), but for this study data correction was required.

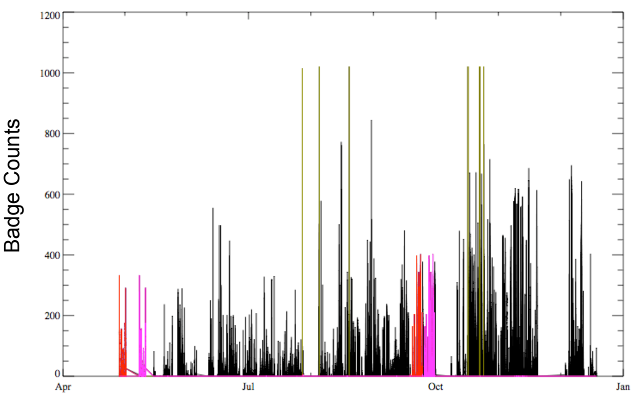

Figure 1. Detecting spikes (green) and repeats in data; first occurrences shown in red, repeats in purple.

All of the corrections described here were automated to work from the raw data, which are unchanged. The first step was to correct any time errors, and remove any time-data pairs that did not correspond to a real measurement. Interviewers' records of visit times and badge issuance were used to resolve instances of concurrent data where a new badge was started for a participant before the old one was stopped.

Heuristic search techniques were used to find repeats, and unsupervised learning algorithms used to scan for areas of malfunction and baseline elevation for each badge. The example in Fig 1 shows spikes and repeated data; the intervals shown in purple are recorded as containing invalid data for that badge and period. Where an elevated baseline was detected, it was subtracted and the data retained.

Participant time series

From data flagged or corrected as above, hourly values for each participant were calculated. For estimated data to fill gaps, a weighted sum of approximating variables was used. Predictor variables for any hour of data included average data values for the person for that hour of the day, average data values for all participants at that hour of the year, average values for that participant during that hour of the week, and average values for the participant over the nearest week in which they had recorded measurements. A single-layer perceptron was utilised to optimise weights, by training on predicting measured values. Evolution of the coefficients is shown in Fig 2.

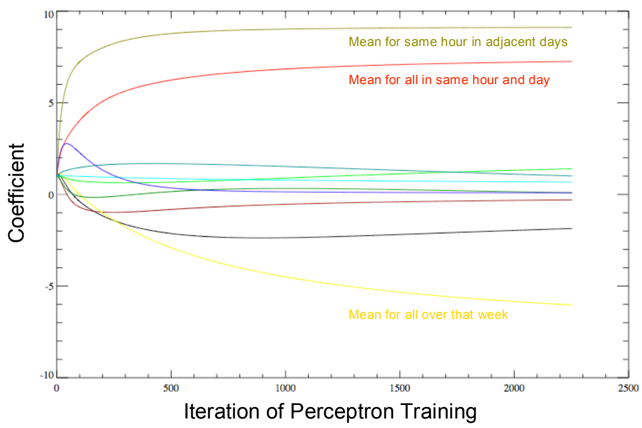

Figure 2. Evolution and stabilisation of the perceptron coefficients with iteration of the training algorithm.

In this instance, weights also gave an indication of the effectiveness of each predictor, though there was collinearity from using both mean and median as separate predictors. The best predictors, consistently across sites and years though coefficients differed, were the mean value for that time of day over the nearest week in which the participant had recorded measurements (9.1 in Fig 2) and the mean for all participants at that time and day (7.3).

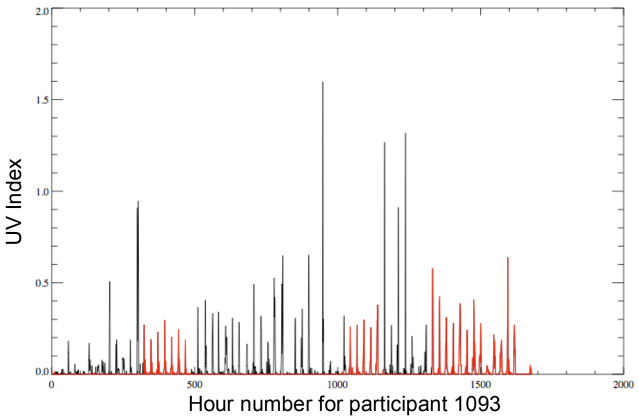

Figure 3. Example of gap filling for one participant.

Figure 3 illustrates how the gap filling completes the hourly data series, though this was an 8-week participant so the last two weeks were not required.

All error detection and correction was automated, as above, for objectivity and repeatability. On the other hand, recognising the types of data error and confirming their resolution required human intervention to review both individual readings and general trends. Figure 4 illustrates one such review, with data from Dunedin to November 2008. The hourly data have been combined into three periods per day (before 11:00, 11:00-16:00, after 16:00) corresponding to the participants' daily logs of clothing worn. In downstream analyses these two datasets are combined to scale the UV received by amount of skin exposed, as the integral of this product is used as a predictor of vitamin D production.

In Fig 4, participants appear as horizontal bars, with coloured bands showing the average UVI experienced over the period. Typically there will be one coloured band per day. The daily total exposure averaged across all participants is shown in red, expressed as Standard Erythemal Doses (SEDs, 1 UVI for 1 hour gives 0.9 SED) on the scale at right. Grey dashed vertical lines are Mondays, highlighting a weekly periodicity for many participants, consistent with spending more time outside in the weekend, especially in the warmer months. Average UV exposures dropped markedly in winter, as expected.

Figure 4 also illustrates other aspects of the study. The upper envelope of the data shows the rate of recruitment of participants, which achieved good uniformity in Dunedin for 2008. The same plot for Auckland (not shown) highlights a slowdown in late winter until more dosimeter badges were made, and the subsequent busy period in spring to achieve target numbers. In 2009, both Dunedin and Auckland achieved uniform recruitment.

Other measures of the UV exposure have also been derived from the badge data. From a knowledge of the UV irradiance spectrum for given solar zenith angle and ozone amount, the nearly erythemal response of the badges can be converted to CIE or other vitamin D action spectra.

References

Allen, M.W., McKenzie, R.L. 2010. Electronic UV dosimeters for research and education. Paper 11 at NIWA UV Workshop, Queenstown, 7-9 April.

Sherman, D. 2010. Personal UV dosimeter badges: Mark II. Paper 13 at NIWA UV Workshop, Queenstown, 7-9 April.

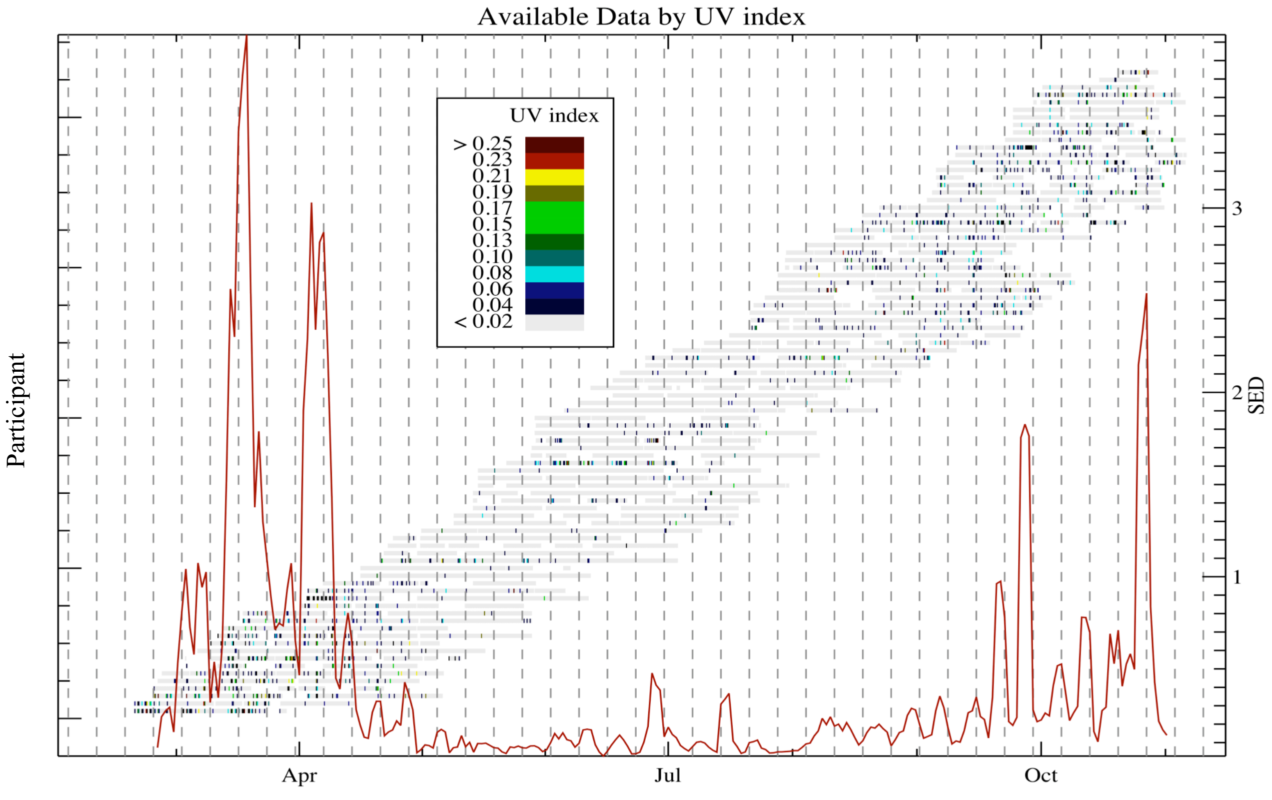

Figure 4. Data for Dunedin 2008, with participants as horizontal bars in three periods per day, gaps for missing data, and colours for average UVI. Average received SEDs for the day are in red, and grey dashed vertical lines are Mondays.