Environmental Research 155 (2017) 141–166

Contents lists available at ScienceDirect

Environmental Research

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

Very short-term reactive forecasting of the solar ultraviolet index using an extreme learning machine integrated with the solar zenith angle

Ravinesh C. Deo[a,*], Nathan Downs[a], Alfio V. Parisi[a], Jan F. Adamowski[b], John M. Quilty[b]

[a] School of Agricultural, Computational and Environmental Sciences, International Centre for Applied Climate Sciences (ICACS), Institute of Agriculture and Environment (IAg & E), University of Southern Queensland, QLD 4300, Australia

[b] Department of Bioresource Engineering, Faculty of Agricultural and Environmental Science, McGill University, Québec H9X 3V9, Canada

ARTICLE INFO

Keywords:
Extreme learning machine
Solar ultraviolet index (*UVI*)
Multivariate adaptive regression splines
M5 Model Tree
Real-time solar forecasting

ABSTRACT

Exposure to erythemally-effective solar ultraviolet radiation (UVR) that contributes to malignant keratinocyte cancers and associated health-risk is best mitigated through innovative decision-support systems, with global solar UV index (*UVI*) forecast necessary to inform real-time sun-protection behaviour recommendations. It follows that the *UVI* forecasting models are useful tools for such decision-making. In this study, a model for computationally-efficient data-driven forecasting of diffuse and global very short-term reactive (VSTR) (10-min lead-time) *UVI*, enhanced by drawing on the solar zenith angle (θ_s) data, was developed using an extreme learning machine (ELM) algorithm. An ELM algorithm typically serves to address complex and ill-defined forecasting problems. UV spectroradiometer situated in Toowoomba, Australia measured daily cycles (0500–1700 h) of *UVI* over the austral summer period. After trialling activations functions based on sine, hard limit, logarithmic and tangent sigmoid and triangular and radial basis networks for best results, an optimal ELM architecture utilising logarithmic sigmoid equation in hidden layer, with lagged combinations of θ_s as the predictor data was developed. ELM's performance was evaluated using statistical metrics: correlation coefficient (r), Willmott's Index (*WI*), Nash-Sutcliffe efficiency coefficient (E_{NS}), root mean square error (*RMSE*), and mean absolute error (*MAE*) between observed and forecasted *UVI*. Using these metrics, the ELM model's performance was compared to that of existing methods: multivariate adaptive regression spline (MARS), M5 Model Tree, and a semi-empirical (*Pro6UV*) clear sky model. Based on *RMSE* and *MAE* values, the ELM model (0.255, 0.346, respectively) outperformed the MARS (0.310, 0.438) and M5 Model Tree (0.346, 0.466) models. Concurring with these metrics, the Willmott's Index for the ELM, MARS and M5 Model Tree models were 0.966, 0.942 and 0.934, respectively. About 57% of the ELM model's absolute errors were small in magnitude ($\pm$0.25), whereas the MARS and M5 Model Tree models generated 53% and 48% of such errors, respectively, indicating the latter models' errors to be distributed in larger magnitude error range. In terms of peak global *UVI* forecasting, with half the level of error, the ELM model outperformed MARS and M5 Model Tree. A comparison of the magnitude of hourly-cumulated errors of 10-min lead time forecasts for diffuse and global *UVI* highlighted ELM model's greater accuracy compared to MARS, M5 Model Tree or *Pro6UV* models. This confirmed the versatility of an ELM model drawing on θ_s data for VSTR forecasting of *UVI* at near real-time horizon. When applied to the goal of enhancing expert systems, ELM-based accurate forecasts capable of reacting quickly to measured conditions can enhance real-time exposure advice for the public, mitigating the potential for solar UV-exposure-related disease.

1. Background

Implementing sun-protection to mitigate the impact of erythemally-effective radiation is a strategic initiative of the Global Solar UV index (*UVI*) developed by the World Health Organisation (WHO), World Meteorological Office (WMO), United Nations Environment Programme (UNEP) and International Commission on Non-Ionizing Radiation Protection (ICNIRP) (Fernández-Delgado et al., 2014; WHO, 2002). Ranging from 0 to 11+ (the greater the value, the greater the risk), the *UVI* is a simple, clear numeric index used by health departments, national forecasting agencies and media outlets. Each integer increase in the *UVI* represents an increase of roughly 25 mW m^{-2} in erythematic ultraviolet radiation. Its calculation is based on the reference action spectrum for UV-induced erythema provided by the Commission internationale de l′éclairage (CIE, 1999). In Australia, where skin cancer incidence is high with about 1600 fatalities and

* Corresponding author.
E-mail address: ravinesh.deo@usq.edu.au (R.C. Deo).

http://dx.doi.org/10.1016/j.envres.2017.01.035
Received 20 December 2016; Received in revised form 27 January 2017; Accepted 28 January 2017
0013-9351/

270,000 keratinocyte skin cancer cases diagnosed every year the *UVI* was introduced in 1996 to facilitate public awareness, through such initiatives as the Sun Smart campaign (Dobbinson et al., 2008; Gies et al., 2004). Between 2006–2010, melanoma accounted for 14–23% of all cancers in the 15–19 and 20–24 year-old age groups (ACIM, 2016; Iannacone et al., 2015). Reflecting the low tropical latitudes and prevalence of fair skin types in Queensland — the geographical focus of the present study — melanoma cases have exceeded those reported in any other Australian states (Neale et al., 2010; Parkin et al., 2010). From a public health perspective, novel predictive techniques allowing accurate real-time estimates of erythemally-effective *UVI* are needed to better inform populations-at-risk and facilitate shifts in their attitude or behaviour with respect to sun exposure. Helping to mitigate the risks of exposure to solar ultraviolet radiation, *UVI* information systems for smartphones and mobile devices can be tailored to their specific environments (Igoe et al., 2013a, 2013b).

The Australian Bureau of Meteorology, Australian Cancer Council and Australian Radiation Protection and Nuclear Safety Agency (ARPANSA) recommend the Sun Protection Times, available through cellular phone and android system apps like SunSmart (SunSmart App, 2016), to advise the public on the *UVI* risk. The Cooperative Research Centre for Southern Hemisphere Meteorology has a forecasting system which issues predicted *UVI* values on a daily (hourly-step) basis (BOM, 2016). While in nine major Australian cities *UVI* is measured by agencies like ARPANSA, for rural and remote sites where meteorological stations are absent, *UVI* forecasts must rely on predictive models that employ minimum environmental monitoring efforts to pragmatically inform the public of the risk of solar *UVI* exposure.

Forecasting of ultraviolet solar radiation is achieved through either mechanistic or modelling-based methods. In a mechanistic method, a solar spectroradiometer or radiometer is mounted in an open space to monitor *UVI* (Gies et al., 2004; Kudish et al., 2005; Neale et al., 2010; Sharma et al., 2013) whereas in a modelling-based method, one employs either a radiative transfer equation (Allaart et al., 2004; Badosa et al., 2005; Schmalwieser et al., 2002) or a data-driven algorithm (Alados et al., 2007, 2004; Feister et al., 2013; Jacovides et al., 2015; Krzyscin, 2002a , 2002b; Krzyścin, 2003; Latosińska et al., 2015). Although mechanistic methods have been widely adopted by health departments, the cost of equipment (*e.g.*, spectroradiometers, radiometers and sky imagers) and accessibility issues can constrain their applicability in some regions. Radiative transfer models (*e.g.*, *LibRadTran* and *STARsci*) are considered the gold-standard for modelling *UVI* under cloud-free conditions (Mayer and Kylling, 2005; Ruggaber et al., 1994), though semi-empirical models (*e.g.*, *Pro6UV*) can also accurately predict *UVI* under such conditions (Green et al., 1974; Rundel, 1986). In radiative-transfer-based models, equations are forced within boundary conditions (*e.g.*, ozone concentration) (Junk et al., 2007); however, this has the disadvantage that boundary condition data must be acquired experimentally through a satellite source with bias corrections (Zhang et al., 2004).

In spite of the wide range of radiative transfer model applications (Neale et al., 2010; Parisi et al., 2008, 2007; Schmalwieser et al., 2002; Sudhibrabha et al., 2004), their practicality is limited by the requirement of boundary conditions for parametrically-forced equations. Solar irradiance is also modulated by many factors (*e.g.*, clouds, aerosols and ozone) (Sharma et al., 2013; Taylor et al., 2016) but their simultaneous measurement is either tedious or difficult to obtain simultaneously across a sufficiently large number of ground-based monitoring stations and in remotely populated areas (Allaart et al., 2004; Badosa et al., 2005; Feister et al., 2013). Furthermore, this data must be acquired from expensive monitoring devices like spectroradiometers and satellites. A radiative transfer model also has restricted applicability at sites where data are unavailable or fixed boundary conditions are implemented. The fact that such models have to apply a finite quantization range to the imprecisely fitted data restricts their ability to consider the dynamics of *UVI* on a continuous basis, or the nonlinear dependencies of solar irradiance on the predictors (Latosińska et al., 2015; Takenaka et al., 2011).

In addressing such non-linearity issues, data-driven models are gaining prominence over radiative transfer models. With a predictive power to extract attributes from antecedent data known to govern the *UVI* dynamics (Alados et al., 2007; Jacovides et al., 2015; Latosińska et al., 2015), data driven models are often applied on a set of predictors (*e.g.*, solar angle, θ_s) integrated in a regression algorithm for *UVI* forecasting. Especially, data-driven models' parsimonious nature has led to their widespread use in deducing boundary conditions in radiative models (Jacovides et al., 2006, 2015; Taylor et al., 2016). Free from distributional assumptions and not tied to boundary conditions, data-driven models are easy to implement, user-friendly and cost efficient compared to experimental approaches or radiative models (Elminir et al., 2008). Instead, data-driven models operate on antecedent observations and consider the non-linear (physical) processes which affect the dynamics of solar energy (Elminir et al., 2008; Latosińska et al., 2015), something which is impossible to implement in a radiative model based on fixed boundary condition data.

Data-driven models are developed with a wide range of machine learning algorithms: artificial neural network (ANN), support vector machine (SVM), multivariate adaptive regression splines (MARS) and M5 Model Trees, amongst the others. These are widely used in climate and atmospheric modelling (Deo et al., 2017; Deo and Şahin, 2015a, 2015b; 2016; Deo et al., 2015, 2016a, 2016b, 2016c; Krzyscin, 2002a, 2002b, 2003; Salcedo-Sanz et al., 2015). Using ANNs to model solar broadband UV, global spectral UV, and photosynthetically-active radiation (PAR) for a semi-urban region of Athalassa (Cyprus), Jacovides et al. (2015) showed the model precision to be dependent on input combinations and the spectral range of the forecasted data. Forecasting erythemal local noon UV irradiance (response variable) in Poland using a feedforward, multilayered supervised ANN drawing on experimental data: e.g., date, latitude, *UVI* measured 1, 2 or 365 days (1 year) in advance, Latosinska et al. (2015) obtained accurate forecasts of medium and high *UVI*. Having applied an ANN drawing on optical air mass, ozone, latitude, horizontal visibility and cloud (type, coverage and height) to forecast *UVI* at three locations on the Iberian Peninsula, Aldos et al. (2004) noted good accuracy based on statistical metrics of mean bias (*MB*) and normalized root mean standard error (*RMSE, %*): $-0.1\% \le MB \le 1.6\%$ and $14.6\% \le RMSE\ \% \le 19.6$. Alados et al. (2007) found an ANN model to exhibit strong predictive ability for *UVI* at 7 sites in Spain. Some studies have also integrated data-driven models into radiative-based models. For example, Taylor et al. (2016) found that an ANN-integrated radiative transfer solver yielded coherent maps of spectrally-integrated global horizontal irradiance within roughly one minute and Takenaka (2011) applied an ANN to approximate radiative transfer codes for ultraviolet A, ultraviolet B, PAR, as well as direct and diffuse solar radiation fluxes.

While two studies (Deo and Sahin, 2017; Deo et al., 2016a, 2016b, 2016c) developed an ANN and an SVM model, respectively, for global incident solar radiation forecasting in Queensland Australia, the study presented in the this paper is the first to assess the feasibility of data-driven models, particularly one implementing an ELM algorithm (Huang et al., 2015), to be used in forecasting *UVI* in an Australian context. The ELM algorithm requires less computational time than its counterparts (*e.g.*, ANN and SVM), and randomly (automatically) generates hidden layers' weights and biases from a given continuous probability distribution function (e.g., uniform, normal, etc.), thus reducing problems to sets of linear equations with analytically-determined hidden neuronal weights. Although the ELM was introduced as recently as 2006, its versatility has captured the attention of researchers in the health sector, and has led to its widespread use in biomedical applications: predicting human leukocyte antigen peptide bonding (Handoko et al., 2006), protein interaction (You et al., 2013), mammographic analysis (Qu et al., 2011) and diagnosis of lung cancer (Daliri, 2012). In the field of water resources, ELM has been applied to drought