Estimation of Leaf Wetness Duration Using Adaptive Neuro-Fuzzy Inference Systems

A. Ghobakhloo¹, F. Amir¹, J. Whalley¹ and P. Sallis²

¹ Geoinformatics Research Centre
² School of Computer and Mathematical Sciences
Auckland University of Technology, New Zealand
Email: akbar@aut.ac.nz

Abstract: The estimation of leaf surface wetness has received considerable attention in recent years from a diverse group of scientists. Leaf wetness is widely accepted as one of the most important input variables for modelling many biophysical processes such as the development and spread of fungal and bacterial diseases. Various types of sensors have been developed for measuring leaf surface wetness. The rapid development and varied nature of these sensors have contributed to a lack of standardisation and the lack of a single accepted protocol for the use of sensors. An alternative to the use of sensors is the simulation or modelling of leaf surface wetness (Huber & Gillespie, 1992; Weiss, 1990). Simulation enables surface wetness to be estimated from historical, forecast weather data, or both, rather than from monitoring and measurement using in-field leaf wetness sensors.

The objective of this work was to develop and evaluate Adaptive Neuro-Fuzzy Inference Systems (ANFIS) as an approach to modelling Leaf Wetness Duration (LWD). This paper reports on a comparative analysis of ANFIS with Classification and Regression Tree/Stepwise Linear Discriminant (CART), Number of Hours Relative Humidity Greater than 90% (NHRH>90%), Fuzzy Logic System (FLS) model, Penman-Monteith (P-M) model, and the Surface Wetness Energy Balance (SWEB) model. The experimental results in this study shows that the overall mean absolute error of ANFIS model was lower than all other models. It was also shown that ANFIS model resulted in higher estimation accuracy over FLS in all five stations.

Keywords: Leaf wetness duration, fuzzy logic, adaptive neuro-fuzzy inference system, CART
INTRODUCTION

The presence of water on plant surfaces influence many biophysical processes such as the development and spread of fungal and bacterial diseases. For a fungal disease to occur, certain environmental conditions have to be fulfilled, including a specific duration of wetness on the leaf surface. Leaf wetness duration (LWD) is the length of time that rain, dew, or fog droplets are retained on aerial plant surfaces. LWD is an important factor in the development of certain foliar diseases, but is difficult to measure since there is no observation standard (Sentelhas, 2004). Despite this difficulty, many plant disease management and decision support systems rely on LWD (Huber and Gillespie, 1992, Klemm et al., 2002). Various types of sensors have been developed for measuring leaf wetness. Mechanical sensors were developed from the 1950s (Hirst, 1954) and the trend shifted to electronic sensors after the 1970s (Sutton, et al., 1984). Electronic sensors use the principle of resistance or capacitance to measure the sensor’s surface exposure to droplets. The rapid development and varied nature of these sensors have contributed to a lack of standardisation and the lack of a single accepted protocol for the use of sensors. There are currently numerous different leaf wetness sensors available in the market. However, none of them have materials and deployment standards to verify them (WMO, 2008). In addition, the problems associated with leaf wetness measurement are not related only to the sensors themselves, but also to how the sensors are used (Rowlandson, et al., 2015). Using leaf wetness sensors can also be costly and labour intensive.

An alternative is to use mathematical models to simulate LWD using meteorological data. Simulation enables surface wetness to be estimated from historical and or forecast weather data, rather than from monitoring and measurement using in-field leaf wetness sensors (Huber & Gillespie, 1992; Weiss, 1990). Based on the input types LWD models are classified into two categories: empirical and physical models (Kim et al., 2002). Physical models simulate the processes of dew formation and evaporation between the plant surface and the atmosphere (Rao et al., 1998). Empirical models have simple inputs and often based on regression analysis and the relationship between leaf wetness and weather data (Huber and Gillespie, 1992).

Leaf wetness is a difficult variable to measure and cannot be considered a true atmospheric variable as it is related to structural and surface optical properties and microclimate (Sentelhas, 2005). Physical changes to the surrounding of a leaf has an important role in formation of leaf wetness. Various approaches such as fuzzy logic and neural networks, have been employed to model and characterise leaf wetness patterns. (Jang et al., 1997, Kim et al., 2004). Fuzzy logic and neural networks are characterisation techniques that have proven to achieve higher accuracy and precision than classic statistical approaches (Weiland & Mirschel, 2008). These are suitable methods to model complex nonlinear functions, dealing with prediction, classification and pattern recognition problems (Zadeh, 1994, Mellit & Kalogirou, 2008).

This paper proposes using Adaptive Neuro-Fuzzy Inference Systems (ANFIS) to estimate LWD. ANFIS is a hybrid system that combines fuzzy logic and neural network techniques that intend to take advantage of both paradigms to assess the condition of leaf surfaces. This paper reports on a comparative analysis of ANFIS with five existing LWD models, including Classification and Regression Tree/Stepwise Linear Discriminant (CART) (Gleason and Koehler, 1994), Number of Hours Relative Humidity Greater than 90% (NHRH≥90%) (Sentelhas et al., 2008), a Fuzzy Logic System (FLS) model (Kim et al., 2004), the Penman-Monteith model (Sentelhas et al 2006), and the Surface Wetness Energy Balance (SWEB) model.

RELATED STUDIES

A number of mathematical models have been developed to simulate LWD during the past years. Some of the models were involved in this study for proposed model validation. This section briefly reviews two empirical models, NHRH≥90% model, CART model, and two physical models, P-M model, SWEB model, also a hybrid empirical-physical model, FLS.

NHRH≥90% model is the simplest empirical model that uses only one variable to predict leaf wetness. This model assumption is that when Relative Humidity (RH) is greater or equal to a constant value (90% in most cases), then surface wetness is present (Sentelhas et al., 2008). The reason for this assumption is that when leaves are colder than surrounding air, and leaf temperature reached dew point of surrounding air, condensation begins to form.

CART model is an empirical model utilising a hierarchical decision tree based on threshold values of RH, dew point temperature, and wind speed as inputs to estimate LWD. The end of the regression tree is the branch of conditional statement that classifies whether a period is dry (no dew) or wet (dew present). CART model has relatively few input variables requirement, which is why this model is considered simple and easy to implement. In a study by Gleason and Koehler (1994), CART tends to underestimate wetness presence in 10 out of 14 stations.

P-M model is a physical model based on Penman-Monteith equation to estimate latent heat flux (LE) and utilises dew deposition, dew evaporation, and intercepted rain. The main advantage of P-M model is that it
A. Ghobakhlo et al., Estimation of Leaf Wetness Duration Using Adaptive Neuro-Fuzzy Inference Systems

does not require air temperature to be measured at crop level. Like other physical model, P-M model uses net radiation as an input, this often became a problem as net radiation is seldom available in standard weather stations. Thus, usually an estimated net radiation is employed using other available weather parameters. In a study by Sentelhas et al. (2006), P-M model overestimated wetness by 1.33 hours in average within three locations. P-M model also showed low spatial variability under diverse climatic conditions.

SWEB model is a model that is essentially a canopy water budget. This model consists of 4 different modules: surface water distribution, canopy water budget, energy balance, and a transfer coefficient calibrated to surface wetness. These four modules complement each other and lead to the first module, obtaining surface wetness by comparing its fraction index of canopy wet surface area (\(W_{sw}\)) and wetness threshold (\(W_{th}\)). When the index number is higher than the threshold, surface wetness is present. Otherwise, it is stated as dry. In SWEB, net radiation is only used to calculate condensation at night. SWEB was developed with grape canopies in mind. It treats the whole grape canopy as a single large leaf. In a study by Magarey et al., SWEB performed with the highest accuracy in the relatively less humid site, with a mean absolute error of up to 0.7 h/day (2006).

FLS model incorporates physical and empirical approaches. This model is essentially an empirical model that complies with energy balance principles. FLS detects wetness by calculating the relationship between latent heat flux, wind speed, and net radiation. In research by Kim et al. (2004) FLS showed a smaller error rate and superior accuracy compared to the CART model. The main challenge in creating an FLS model is in determining rules and the membership function of each variable; the user needs to understand the characteristics of each variable, and often both number of rules and fuzzy subsets are chosen arbitrarily. This uncertainty may cause a suboptimal model for the particular dataset.

3. ANFIS MODEL

ANFIS is essentially a hybrid model which combines two intelligent models, an Artificial Neural Network (ANN) and a Fuzzy Inference System (FIS) (Alves, et al., 2011). The FIS consists of a fuzzy logic method, fuzzy IF-THEN rules, and fuzzy reasoning. The main feature of this model is that inferences based on rules and known facts are used to make decisions based on human-like reasoning (e.g. wet, most likely wet, and dry). Kim et al. (2004) used an FIS to predict leaf wetness; this specific application of FIS to LWD modelling was referred to as FLS. A drawback of the FIS model is that rules and membership functions were either created arbitrarily or defined based on a user’s understanding of input variable characteristics. The ANN in ANFIS handles the mapping of input to output space through layered neurons that are connected by synaptic junctions. With ANN mapping, membership functions and rules are always data-specific. ANFIS integrates the advantages of both intelligent systems to optimally model dynamic and complex problems.

The model is an adaptive neuro-fuzzy inference system since it utilises an adaptive network that can compute gradient vectors systematically. Figure 1 shows the architecture of the ANFIS network. The neural network classifies and finds patterns within an input parameter through the layers, and maps it into the desired output. Each layer handles different tasks in the network as described below.

Layer 1 is the input layer; this is where the inputs are ‘fuzzified’ or translated into linguistic labels, and the membership grade is generated for each label. Every node in this layer is adaptive, and the number of nodes depends on the number of inputs. After the input spaces are identified in layer 1, the degrees to which inputs satisfy the membership functions are found.

Layer 2, the implication layer, contains fuzzy rules and each node's firing strength of a rule. The firing strength is usually obtained as the product of membership grades generated by the previous layer. There are \(A^b\) number of rules in ANFIS, where ‘A’ is the number of membership functions of every input, and ‘b’ is the number of inputs. The number of nodes in this layer is reliant on the number of rules, and is fixed. The next layer node number will follow this layer’s number.

![Figure 1. The architecture of ANFIS network](image-url)
In layer 3, the normalisation layer, the ratio of each rule’s firing strength to the total of all rule firing strengths is calculated. Then in layer 4, the defuzzification layer, the output of each node is weighted and computed towards the overall output. The output of this layer of the model is a crisp value. In layer 5, the combining layer, the overall output from the previous layer is computed as the summation of every rule’s contribution. ANFIS model construction time depends on the number of inputs and membership functions; some inputs increase the complexity of computation, and also impair the transparency of the underlying input model. Therefore, it is necessary to do heuristic input selection before constructing the model (Jang, 1996).

### 4. EXPERIMENTS

Five weather stations were selected within New Zealand. These stations were located across the country’s North and South Islands in Pukekohe (PKE), Martinborough (MTB), Gisborne (HXT), Marlborough (RPU), and Hawke’s Bay (TRI). These stations are located at least 150 kilometres apart to ensure that each station has different micrometeorological characteristics. Days with rainfall measured ≥0.25 mm were defined as rain days, otherwise any dew-only days. Table 1 shows average wetness period per day in both rain days and dew-only days within the two-months data range used in this study. The TRI station is the wettest station with the highest wetness period in both rain days and dew-only days. Whereas the RPU station is the driest with 8.3 h/day on rainy days and 2.4 h/day wetness on dew-only days.

### 4.1. Data preparation

The data used herein was collected during summer 2012 from January to February. The data includes hourly measurements of air temperature, rainfall, relative humidity, and leaf wetness obtained from five stations maintained by the New Zealand Institute for Plant & Food Research. The wind speed data was obtained from the nearest National Institute of Water and Atmospheric Research (NIWA) stations. Each station contains approximately 1,440 data points for the period of two months. An electronic leaf wetness sensor (Model 237, Campbell, Logan, UT) was used to measure leaf wetness at each station. The sensor was connected to a Campbell CR-10 data logger, with outputs scaled from 0 (driest) to 100 (wettest). The data logger was set to scan the sensor every minute and logged every hour. Campbell model 237 leaf wetness sensors were used as a reference on numerous research (Pedro, 1981; Lau et al., 2000; Sentelhas et al., 2005) that showed the sensor has 15-30 minutes of response time compared to visual observation, confirming their accuracy in the field. When the mean sensor reading of the hour was greater than 50, wetness was considered present on the leaf, otherwise it was considered dry.

We split the data into three randomly selected subsets used for training, validation and testing of each model. The training set contained 70%, the validation set 10% and the testing set 20% of the entire data set. The process of resampling was conducted five times and each time a fresh model was created over the new sets. The approach of resampling the training and testing sets for each iteration was intended to avoid statistical bias that might arise due to a single sampling of the data.

### 4.2. Implementation

A single output ANFIS model, to predict LWD, was trained and simulated in Matlab® R2012B using the Neuro Fuzzy-Design Toolbox. In this study an initial FIS was generated using subtractive clustering method to avoid dimensionality problems (Wei et al., 2007). Subtractive clustering is a fast, one-pass algorithm to take input-output training data and generate a Sugeno-type fuzzy inference system that models the data behaviour. Subtractive clustering estimates the number of rules and membership function for the dataset. In this case, Gaussian Curve membership function (gaussmf) was chosen with numbers of membership function ranging from 2 to 4 for each variable, depending on the dataset. The number of rules was generated according to the number of the membership function selected. To train the model, we used a hybrid learning method that combines a gradient method and least square estimates, this method was chosen to reduce processing time and

### Table 1. Daily mean wetness period and total rain and no rain days on five stations for Jan-Feb 2012

<table>
<thead>
<tr>
<th>Stations</th>
<th>Rainfall records</th>
<th>LW records (h/day)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dew-only day</td>
<td>Rain day</td>
</tr>
<tr>
<td>RPU</td>
<td>47</td>
<td>13</td>
</tr>
<tr>
<td>HXT</td>
<td>32</td>
<td>28</td>
</tr>
<tr>
<td>MTB</td>
<td>43</td>
<td>17</td>
</tr>
<tr>
<td>PKE</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>TRI</td>
<td>40</td>
<td>20</td>
</tr>
</tbody>
</table>

*Rain day defined as days with rainfall ≥0.25 mm was measured;
to avoid local minima (Pabreja, 2011). All the other models evaluated in this study were implemented as reported in the literature and described in Section 2 of this paper.

### 4.2.1. Input Selection for ANFIS

Input variables included in the selection were temperature, relative humidity, rainfall and wind speed. Furthermore, three derived variables: Vapour Pressure Deficit (VPD), para-Net Radiation (pRn), and the partial pressure of water vapour in the air ($e_a$) were included. These variables were required by the FLS model (Kim et al., 2004). In order to find the number of possible input combinations, a binomial coefficient formula was used: $C_b^a$, where $a$ = number of inputs to be tested, and $b$ = desired final number of inputs.

An ANFIS was trained for one epoch for any given input combination and the model with the lowest error rate was chosen as the optimal input combination. The selected model was then fine-tuned with more epochs and a more extensive dataset to find the most suitable model. This approach assumes that the ANFIS with the lowest error rate on one epoch has a higher chance to have a lower error rate during further training (Jang, 1996). This may not always be a correct assumption, but it is pragmatic. The best input combination with the lowest error rate was found to be wind speed, VPD, and pRn.

### 5. RESULTS

Five randomly resampled subsets of data were used to generate models for data obtained from each station as described in Section 4.1. All models were validated with the actual leaf wetness sensor readings and their performances were compared. For each model Mean Error (ME), Mean Absolute Error (MAE) and Estimation Accuracy (EA) were calculated. ME was computed by averaging the differences between measured and estimated LWD for 24 hours period. ME determines the tendency of a model to overestimate or underestimate LWD. MAE was calculated by averaging the absolute values of hourly errors. MAE determines overall accuracy of the model. The EA represents the degree of closeness of estimated and measured LWD in percentage, calculated according to equation 1, where $N$ is the total number of data points:

$$EA = \left(1 - \frac{\sum |\text{Actual} - \text{Estimated}|}{N}\right) \times 100$$

<table>
<thead>
<tr>
<th>Stations</th>
<th>ANFIS</th>
<th>FLS</th>
<th>CART</th>
<th>SWEB</th>
<th>NHRH&gt;90</th>
<th>PM</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPU</td>
<td>82.28</td>
<td>69.47</td>
<td>69.65</td>
<td>85.61</td>
<td>77.89</td>
<td>76.67</td>
</tr>
<tr>
<td>HXT</td>
<td>94.76</td>
<td>70.48</td>
<td>76.43</td>
<td>85.00</td>
<td>91.45</td>
<td>91.43</td>
</tr>
<tr>
<td>MTB</td>
<td>79.55</td>
<td>65.45</td>
<td>80.00</td>
<td>82.27</td>
<td>88.64</td>
<td>86.82</td>
</tr>
<tr>
<td>PKE</td>
<td>83.61</td>
<td>70.83</td>
<td>84.44</td>
<td>84.17</td>
<td>81.39</td>
<td>82.50</td>
</tr>
<tr>
<td>TRI</td>
<td>90.00</td>
<td>82.56</td>
<td>68.59</td>
<td>84.23</td>
<td>80.38</td>
<td>89.10</td>
</tr>
<tr>
<td>Overall</td>
<td>86.04</td>
<td>71.76</td>
<td>75.82</td>
<td>84.26</td>
<td>83.95</td>
<td>85.30</td>
</tr>
</tbody>
</table>

**Figure 2.** LWD models ME (a) and MAE (b) comparison on 5 different stations within Jan – Feb 2012
Inspecting the results from Table 2, ANFIS has the highest overall EA for all five stations, and scored the highest EA for the HXT and TRI stations. SWEB model exhibited the highest EA for the RPU station, while CART model was the most accurate on PKE station followed by SWEB which was slightly higher than ANFIS. The NHRH≥90% model had the highest EA followed by the P-M and SWEB models for the MTB station.

Figure 2(a) presents the ME comparison of each model on five stations. SWEB had similar overestimation and underestimation, which leads to only 0.01 h/day overall overestimation. Both ANFIS and PM are scored not far from 0 on most stations. CART and NHRH≥90% models tends to overestimate leaf wetness occurrence. Figure 2(b) shows the model’s MAE comparison in five different stations. ANFIS model had the smallest MAE in 4 out of 5 stations (overall 0.14 h/day). In contrast, FLS and NHRH≥90% have the largest MAE of 0.28, and 0.24 h/day respectively. Table 3 exhibits the overall ME, MAE, and EA for all models across the five stations. The overall MAE of ANFIS model was lower than other models. Overall EA for ANFIS was higher than the rest. SWEB had ME closer to 0 than ANFIS, while ANFIS and P-M were underestimating by 0.06 h/day.

6. DISCUSSION AND CONCLUSION

This study proposed the ANFIS model as a means of estimating LWD and compared its performance with five other models. The results presented in Section 5 show that ANFIS had higher estimation accuracy over FLS in all stations. ANFIS estimation was also comparable with that of physical models SWEB and P-M which incorporate a rainfall variable directly. These physical models performed comparatively well in this study. In all stations, the SWEB and P-M models both under-estimated and over-estimated wetness presence (Figure 2(a)), suggesting location-based errors occur in both models. In contrast, both empirical models (CART and NHRH>90%) constantly showed false positives across stations. The FLS model, with predetermined rules and membership functions from Kim et al. (2004), did not perform well in this study. The EA of ANFIS are higher over FLS in all five stations suggests that the learning ability of the ANFIS model is an advantage (see Table 2).

ANFIS eliminates the need to decide upon an arbitrary number of membership functions and fuzzy rules, and fits the model to the data set. In theory, ANFIS is more complicated than FLS, but in application it simplified the modelling process and resulted in higher estimation accuracy. The adaptive role of the ANFIS model’s neural network also offers an advantage over linear models: the non-linear representation of ANFIS supports greater adaptability of the model. Thus, simplicity and adaptability are the advantages of the ANFIS model. Although the overall results for ANFIS was showing to be more accurate than the rest of the models, it did not perform well on RPU and MTB stations. Both stations had the lowest number of rain days for the given period (see Table 1) which resulted in an imbalanced and small training set for ANFIS contributing to a lower accuracy. Three stations that ANFIS performed with the highest accuracy was in HXT, TRI, and PKE. These stations had more than 20 rain days out of 60 with adequate amount of rainfall. This suggests that ANFIS model’s input dataset needs to be much larger to enable better generalisation and more accurate estimation.

Further research is needed using a larger, more spatially and temporally diverse dataset for model training and to allow a detailed analysis of the timing and causes of leaf wetness (i.e. dew or rainfall).

ACKNOWLEDGMENTS

The authors would like to thank Robert Beresford and Gareth Hill from The New Zealand Institute for Plant & Food Research for their contribution of weather and leaf wetness data used in this study.

REFERENCES


A. Gobakhshlo et al., Estimation of Leaf Wetness Duration Using Adaptive Neuro-Fuzzy Inference Systems


