SYSTEMATIC STUDY OF WEATHER VARIABLES FOR RAINFALL DETECTION

Shilpa Manandhar[†1], *Soumyabrata Dev*[†2], *Yee Hui Lee*[1], *Stefan Winkler*[3] *and Yu Song Meng*[4]

[1] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
[2] ADAPT Centre, School of Computer Science and Statistics, Trinity College Dublin, Ireland
[3] Advanced Digital Sciences Center (ADSC), University of Illinois at Urbana-Champaign, Singapore
[4] National Metrology Centre, Agency for Science, Technology and Research (A*STAR), Singapore

ABSTRACT

Numerous weather parameters affect the occurrence and amount of rainfall. Therefore, it is important to study these parameters and their interdependency. In this paper, different weather and time-related variables – relative humidity, solar radiation, temperature, dew point, day-of-year, and time-of-day are analyzed systematically using Principal Component Analysis (PCA). We found that four principal components explain a cumulative variance of 85%. The first two principal components are applied to distinguish rain and no-rain scenarios as well. We conclude that all 7 variables have similar contribution towards rainfall detection.

Index Terms— Precipitation, weather sensor, remote sensing, PWV

1. INTRODUCTION

Precipitation is influenced by different factors to varying degree. Recently, because of the growing proliferation of geostationary satellites, there has been renewed interest in using GPS signal delays to better understand this phenomenon. Yao et al. [1] and Benevides et al. [2] studied GPS delays to predict rainfall. In our recent paper [3], we have also discussed the importance of GPS-derived precipitable water vapor (PWV) for understanding rainfall.

However, the classification of rain and no-rain scenarios using only GPS PWV values may not result in good accuracy. In [4], we have illustrated the usefulness of different weather variables in rainfall detection. For example, during rainy conditions, temperature generally approaches the dew point; relative humidity reaches nearly 100%, and precipitable water vapor increases. Solar radiation usually decreases prior to rain and during rain because of the presence of clouds.

In this paper, we first provide a systematic analysis of various weather parameters and use them in detecting precipita-

tion. The source code of all simulations is available online.[1]

2. WEATHER VARIABLES

2.1. Surface Weather Parameters

Surface weather parameters recorded by a weather station (Davis Instruments 7440 Weather Vantage Pro II) with tipping bucket rain gauge are used in this study. The weather station is located at Nanyang Technological University (NTU), (1.3°N, 103.68°E). The weather station measurements used for this paper include Surface Temperature $°C$, Relative Humidity (RH) (%), Dew Point $°C$, Solar Irradiance (W/m^2), and Rainfall Rate (mm/hr). All weather parameters are recorded at 1-minute intervals.

2.2. GPS Derived Water Vapor Content

In addition to the various surface weather parameters, precipitable water vapor (PWV) values derived from GPS signal delays are used as an additional important feature for classification. This section provides a brief overview of the derivation of PWV values from GPS signal delays.

PWV values (in mm) are calculated using the zenith wet delay (ZWD), δL_w^o, incurred by the GPS signals as follows:

$$\text{PWV} = \frac{PI \cdot \delta L_w^o}{\rho_l}, \tag{1}$$

where ρ_l is the density of liquid water (1000 kg/m^3), and *PI* is the dimensionless factor determined by [5]:

$$PI = [-\text{sgn}(L_a) \cdot 1.7 \cdot 10^{-5}|L_a|^{h_{fac}} - 0.0001] \cdot$$
$$\cos\frac{2\pi(DoY - 28)}{365.25} + 0.165 - 1.7 \cdot 10^{-5}|L_a|^{1.65} + f, \tag{2}$$

where L_a is the latitude, *DoY* is day-of-year, h_{fac} is either 1.48 or 1.25 for stations in the Northern or Southern hemisphere, respectively. $f = -2.38 \cdot 10^{-6}H$, where H is the station altitude above sea level, can be ignored for $H < 1000m$.

---

[1] https://github.com/Soumyabrata/weather-features

For this paper, the ZWD values for an IGS GPS station located at NTU (station ID: NTUS) are processed using GIPSY OASIS software and recommended scripts [6]. PWV values are then calculated for NTUS using Eqs. 1-2, with $L_a = 1.34$, $h_{fac} = 1.48$, $H = 78m$. The calculated PWV values have a temporal resolution of 5 minutes.

2.3. Seasonal and Diurnal Features

Singapore has tropical weather conditions, with scattered to abundant rainfall throughout the year. It has two distinct monsoon seasons, and witnesses uniform temperature throughout the year. Since seasons play an important role in rainfall, and rainfall occurrences in tropical regions also show clear diurnal characteristics. Thus, in addition to the weather sensor and GPS PWV data, seasonal and diurnal variables are included in the analysis the form of Day-of-Year (DoY) and Time-of-Day (ToD).

In this paper, all of the above mentioned variables and the rain data for the year 2010 are used as our experimental dataset. Since GPS PWV values have a resolution of 5 minutes, all surface weather parameters are also sampled at 5-minute intervals.

3. INTERDEPENDENCY OF VARIABLES

Based on the database, a 7-dimensional feature vector comprising relative humidity (RH), solar radiation (SR), temperature (T), dew point (DPT), precipitable water vapor (PWV), day of the year (DoY) and time of the day (ToD) is formulated. The degree of cross-correlation amongst these variables is calculated and shown in Fig. 1. It is important to analyze such correlation trends before any classification task is performed, because if two features are perfectly correlated, then the second feature does not contribute any additional information (it is already determined by the other).
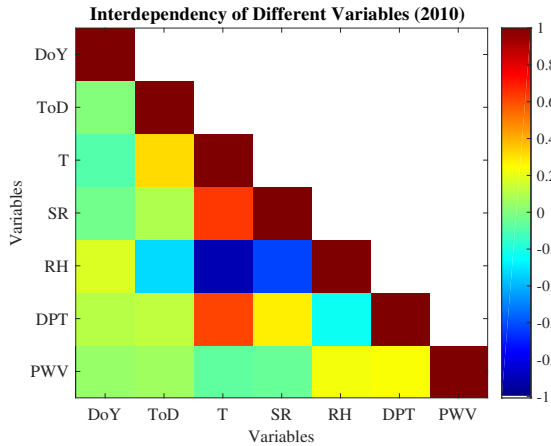

Fig. 1: Correlation between various features (best viewed in color).

Various observations can be made from the off-diagonal elements. T has a good correlation with RH, SR, and DPT. The correlation between T and RH is the highest with a correlation coefficient of -0.9, which indicates that the two variables are highly negatively correlated. T & SR and T & DPT have positive correlation coefficients of 0.6. RH and SR are negatively correlated with a correlation coefficient of -0.6. Solar radiation is lowest in the night, and relative humidity at night on a tropical island like Singapore is generally very high. Therefore a high negative correlation is observed between the two variables, dependent on ToD.

It is also interesting to note that the PWV does not show strong correlation with any of the other variables. However, it has a small positive correlation with RH and DPT. Different seasons and locations have a direct impact on the behavior and correlation of these variables. For temperate regions, a higher degree of correlation is observed between RH and PWV [7], whereas for tropical regions (viz. Singapore), the weather parameters exhibit a very different behavior.

4. PRINCIPAL COMPONENT ANALYSIS

We use Principal Component Analysis (PCA) to investigate the underlying structure of the 7 weather variables. It is also used to determine the individual contributions of these variables to rainfall.

Consider a variable matrix $\mathbf{X}$ of dimension $m \times n$, where m is the number of variables (7 in this case) and n is the number of observations. The individual variables v_{1-7} are extracted from $\mathbf{X}$ and reshaped into the column vectors $\widetilde{\mathbf{v}}_j \in \mathbb{R}^{mn \times 1}$ where $j = 1, 2, .., 7$. The $\widetilde{\mathbf{v}}_j$ obtained from $\mathbf{X}$ are stacked alongside, to form the matrix $\hat{\mathbf{X}} \in \mathbb{R}^{mn \times 7}$:

$$\hat{\mathbf{X}} = [\widetilde{\mathbf{v}}_1, \widetilde{\mathbf{v}}_2, \widetilde{\mathbf{v}}_3, \widetilde{\mathbf{v}}_4, \widetilde{\mathbf{v}}_5, \widetilde{\mathbf{v}}_6, \widetilde{\mathbf{v}}_7]. \quad (3)$$

$\hat{\mathbf{X}}$ is then normalized with respect to the mean $\bar{v}_j$ and standard deviation σ_{v_j} of the individual variables of $\mathbf{X}$, and is represented by $\ddot{\mathbf{X}}$:

$$\ddot{\mathbf{X}} = \left[\frac{\widetilde{\mathbf{v}_1} - \bar{v_1}}{\sigma_{v_1}}, \frac{\widetilde{\mathbf{v}_2} - \bar{v_2}}{\sigma_{v_2}}, .., \frac{\widetilde{\mathbf{v}_j} - \bar{v_j}}{\sigma_{v_j}}, .., \frac{\widetilde{\mathbf{v}_7} - \bar{v_7}}{\sigma_{v_7}} \right]. \quad (4)$$

Subsequently, the covariance matrix of $\ddot{\mathbf{X}}$ is computed. The eigenvalue decomposition of the covariance matrix then yields the eigenvalues and the eigenvectors. The eigenvectors define the new orthogonal axes called principal components (PC). The eigenvalues explain the variance contributed by each of the 7 principal components (PC 1-7).

4.1. Variance Explained by Principal Components

In this section, the most important principal components are analyzed based on the variance explained by each. The percentage of total variance explained by each PC is calculated and shown in Fig. 2. About 40% of the variance is explained

by the first principal component (PC1). PC1-PC4 together explain a cumulative variance of 85%.
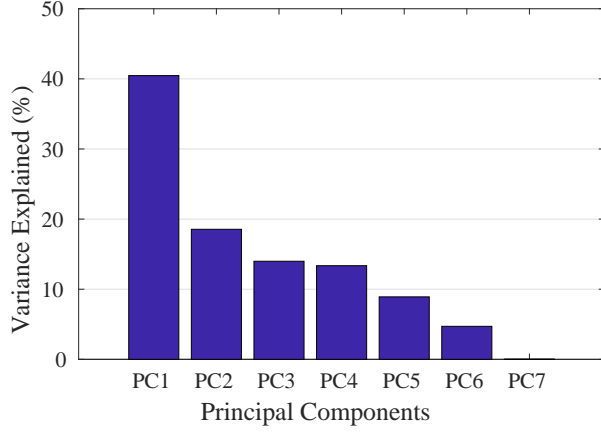

Fig. 2: Variance explained by different Principal Components.

4.2. Biplot Representation

Fig. 3 visualizes the distribution of the samples in the new principal component axes via a biplot representation. This biplot provides us interesting insights on our weather variables, and their contribution to the principal components. It shows the contribution of each of the weather variables to the two principal components, and the correlation amongst them. As an illustration, PWV and DoY are positively correlated with each other, whereas T and RH are negatively correlated to each other. The data points are represented as *red* points in the transformed vector subspace, comprising the two principal components.
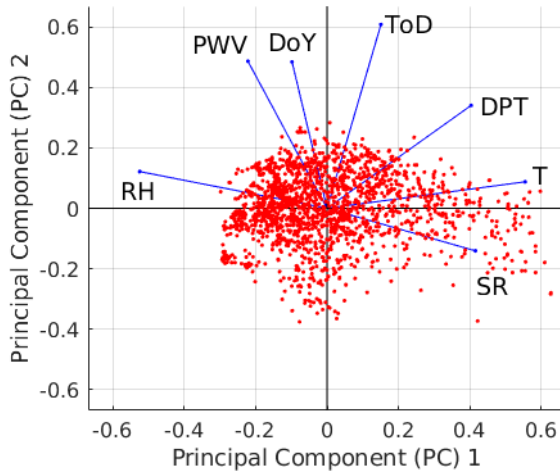

Fig. 3: Biplot representation of the weather variables. The red points indicate the different observations in our dataset, and the vectors represent the different weather variables.

4.3. Variable Loadings

Variable loadings are the contribution of different variables to an individual principal component. Table 1 lists the loading factors for the first four PCs, which together explain 85% of the cumulative variance (see above).

Table 1: Loading factors of the different variables for the first four principal components. Variables with above-average contribution are marked in bold.

Variables	PC1	PC2	PC3	PC4
T	**0.57**	0.03	0.03	-0.04
RH	**-0.53**	0.24	0.009	-0.05
SR	**0.44**	-0.06	0.22	-0.22
DPT	0.33	**0.54**	0.07	-0.02
ToD	0.23	0.08	**-0.47**	**0.79**
DoY	-0.06	**0.41**	**0.73**	**0.47**
PWV	-0.05	**0.68**	**-0.42**	-0.22

The sum of the squares of all loadings for an individual principal component equals unity. A positive (negative) loading indicates that the variable and the particular principal component are positively (negatively) correlated. A large (either positive or negative) loading factor indicates that a variable has a strong effect on that principal component. The threshold for loading is calculated by considering equal contribution from all variables to the principal component. In our case, the number of variables is 7, and therefore the loading threshold is $\sqrt{1/7} = 0.378$. Any variable that has a loading above this threshold can be be considered an important contributor to that principal component. Those variables are highlighted in Table 1.

All principal components are orthogonal to each other and hence uncorrelated. Therefore, each individual PC can be useful to explain a unique phenomenon. In our case, PC1 has strong positive loadings for T and SR and has a strong negative loading for RH. These variables depict changes with rainfall. When it rains, humidity goes higher, temperature and solar radiation both drop [4]. Thus, PC1 might be useful in classifying rainfall scenarios.

PC2 has strong positive loading for PWV, DoY and DPT. It might be useful in explaining the seasonal dependency of PWV and its correlation to DPT. PWV values are generally higher in rainy seasons and lower in non-rainy seasons. Thus, PC2 might also explain the seasonal dependency of rainfall.

The remaining principal components PC3 and PC4 have strong positive loadings for DoY and ToD respectively. Thus, these PCs are useful in explaining the temporal variations of different phenomenon. It can also be useful in explaining the nature of occurrence of frequent evening rain events in Singapore during the North-East Monsoon and early morning downpours during the Inter-Monsoon seasons.

4.4. Rainfall Detection

We now analyze the weather- and time-related variables considered in this paper for the purpose of rainfall detection. Fig. 4 shows the plot of samples against the first two principal components, similar to Figure 3. Sample points are distinguished based on the rainfall data recorded by the weather station in order to provide us an intuition on the degree of separability of *rain* and *no-rain* data points. As is apparent from the plot, the lower dimensional subspace created via PCA does indeed facilitate discrimination between *rain* and *no-rain* observations.
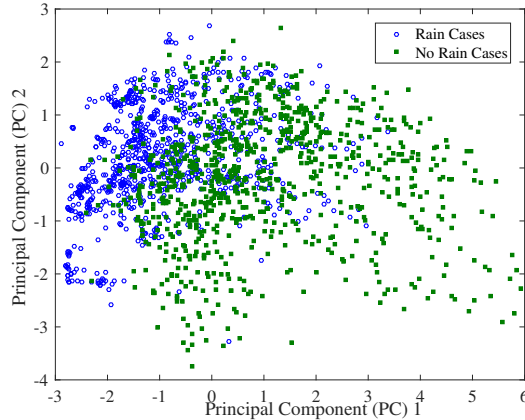

Fig. 4: Visualization of high-dimensional features in a lower dimensional subspace spanned by the first two principal components. The *rain* and *no rain* data points are represented as blue circles and green squares, respectively.

5. CONCLUSIONS & FUTURE WORK

We have conducted a systematic analysis of various weather parameters measured in the tropics. We analyzed the correlation coefficients between 7 different meteorological variables. We then used PCA and found that 4 principal components are needed to explain a cumulative variance of 85%. This means that the individual variables are relatively uncorrelated and have their own significance in describing different weather phenomena.

The subspace representation of data samples against PC1 and PC2 with the rain condition taken into consideration indicates the usefulness of PCA in separating rain and no-rain cases; it can therefore be used in machine learning classification algorithms to detect rain. However, as discussed, the other principal components also have significant contributions to the total variance explained, and might be useful in further improving the classification results.

In the future, we plan to include sky images [8] and image-based features for detecting precipitation [9]. We also aim to propose a machine-learning based framework [10] to predict precipitation in a multi-modal manner, using weather- and image-based features.

6. REFERENCES

[1] Y. Yao, L. Shan, and Q. Zhao, "Establishing a method of shortterm rainfall forecasting based on GNSS-derived PWV and its application," *Scientific Reports*, vol. 7, Sep 2017.

[2] P. Benevides, J. Catalao, and PMA. Miranda, "On the inclusion of GPS precipitable water vapour in the nowcasting of rainfall," *Natural Hazards and Earth System Sciences*, vol. 15, no. 12, pp. 2605–2616, Nov 2015.

[3] S. Manandhar, Y. H. Lee, and S. Dev, "GPS derived PWV for rainfall monitoring," in *Proc. International Geoscience and Remote Sensing Symposium (IGARSS)*, 2016.

[4] S. Manandhar, S. Dev, Y. H. Lee, Y. S. Meng, and S. Winkler, "A data-driven approach to detecting precipitation from meteorological sensor data," in *Proc. International Geoscience and Remote Sensing Symposium (IGARSS)*, 2018.

[5] S. Manandhar, Y. H. Lee, Y. S. Meng, and J. T. Ong, "A simplified model for the retrieval of precipitable water vapor from GPS signal," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6245–6253, Nov 2017.

[6] S. Desai, D. Kuang, and W. Bertiger, "GIPSY/OASIS (GIPSY) overview and under the hood," Tech. Rep., Dept. Geomatics, Univ. Newcastle upon Tyne, UK, Oct. 1996.

[7] W. P. Elliott and J. K. Angell, "Variations of cloudiness, precipitable water, and relative humidity over the United States 1973-1993," *Geophy. Res. Letters*, vol. 24, no. 1, pp. 41–44, Jan 1997.

[8] S. Dev, F. M. Savoy, Y. H. Lee, and S. Winkler, "Design of low-cost, compact and weather-proof whole sky imagers for High-Dynamic-Range captures," in *Proc. International Geoscience and Remote Sensing Symposium (IGARSS)*, 2015, pp. 5359–5362.

[9] S. Dev, S. Manandhar, Y. H. Lee, and S. Winkler, "Detecting rainfall onset using sky images," in *Proc. IEEE Region 10 Conference (TENCON)*, 2016.

[10] S. Dev, B. Wen, Y. H. Lee, and S. Winkler, "Ground-based image analysis: A tutorial on machine-learning techniques and applications," *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, no. 2, pp. 79–93, June 2016.