A Multivariat Dispersion Analysis and its Application on the Leaves Curl Identification

Khoirin Nisa, Celestin C. Kokonendji, Mona Arif Muda

ABSTRACT

Dispersion concerns with how spread out or dispersed a distribution is, the statistic that conveys this information is called "measure of dispersion". Here we present multivariate dispersion measures for multivariate data. A generalized form and its properties are introduced. An application on leaves curl identification from polarimetric near infra-red spectrum data is presented.

Keywords: G-dispersion matrix, generalized variance, NIR.

INTRODUCTION

In statistical analysis of multivariate data, along with measures of central tendency, measure of dispersion plays an important role in descriptive statistics and inferences. The dispersion of multivariate data usually measured by covariance matrix and generalized variance (i.e. the determinant of the covariance matrix).

Nowadays, most precision agriculture (PA) research is oriented to be able to remotely detect crops conditions in quasi-real time using remote sensing systems. Image data resulted from the system give information about the crops. In this paper we apply the generalized variance for leaves curl identification from polarimetric near infra-red spectrum data. If the leaves curl is existent on crops, it generally indicates some disease existent on the crops.

Table-1. Generalized variance of angle of polarization (AOP) of leaves.

Leaf#	pixels	Generalized variance
1	79994	872,3352
2	32439	1018,127
3	8538	320,7197
4	27692	738,8698
5	7080	1929,161
6	15095	1102,096
7	6172	885,2961
8	5107	1399,42
9	7685	788,8686
10	4822	853,7134

AOP image from the sample leaves showing the relationship between the levels of the leaves curl surface with the AOP generalized variance.

RESULT AND DISCUSSION

Below we introduce a definition of dispersion matrix of a distribution around a fixed point, then we also introduce the definition of its generalization.

Definition 1. Dispersion Matrix

Let $X \in R^K$ and P_θ be a unimodal distribution with the 2nd moment exists. Then the dispersion matrix of P_θ around fixed point $x_0 \in R^K$ is given by

$$\Sigma_{x_0} = \int_{\mathbb{R}^k} (\mathbf{x} - \mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)^\top P_\theta(d\mathbf{x}).$$

If $x_0 = \mu$ then we have $\Sigma_\mu = \int_{\mathbb{R}^k} (\mathbf{x} - \mu)(\mathbf{x} - \mu)^\top P_\theta(d\mathbf{x})$ which is equal the covariance matrix.

Definition 2. G-Dispersion Matrix

Let G be a symmetric definite positive matrix. A generalization of dispersion matrix of P_θ around fixed point $x_0 \in R^K$ is given by:

$$\Sigma_{G,x_0} = \int_{\mathbb{R}^k} (\mathbf{x} - \mathbf{x}_0)\mathbf{G}^{-1}(\mathbf{x} - \mathbf{x}_0)^\top P_\theta(d\mathbf{x}).$$

The determinant of the dispersion and G-dispersion matrices are called "generalized dispersion" and "generalized G-dispersion" respectively.

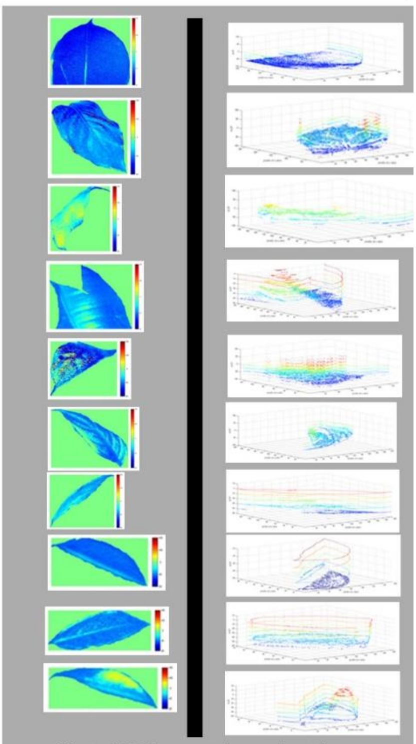

Fig.4. AOP from each numbered of the leaves

Application of multivariate dispersion on leaves curl identification

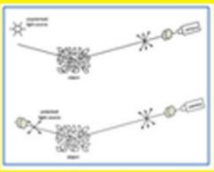

Figure-1. Imaging stokes polarimetric (top), and Meuller (bottom).

Figure 2. Leaves objects

Figure 3. Angle of polarization from the scene

Conclusions

In this paper we introduced a generalization of covariance matrix, i.e. dispersion matrix and G-dispersion matrix. Some properties of generalized G-dispersion are established (not presented in this poster). In the analysis of leaves polarimetric NIR spectrum data, it is shown that generalized variance can used to identify leaves curls.