Applied Signal and Image Processing:
Multidisciplinary Advancements

Rami Qahwaji
University of Bradford, UK

Roger Green
University of Warwick, UK

Evor Hines
School of Engineering, University of Warwick, UK

Information Science
REFERENCE

Senior Editorial Director: Kristin Klinger
Director of Book Publications: Julia Mosemann
Editorial Director: Lindsay Johnston
Acquisitions Editor: Erika Carter
Development Editor: Joel Gamon
Production Coordinator: Jamie Snavely
Typesetters: Keith Glazewski & Natalie Pronio
Cover Design: Nick Newcomer

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Library of Congress Cataloging-in-Publication Data

Applied signal and image processing : multidisciplinary advancements / Rami Qahwaji, Roger Green, and Evor Hines, editors.
p. cm.
Includes bibliographical references and index.
Summary: "This book highlights the growing multidisciplinary nature of signal and image processing by focusing on emerging applications and recent advances in well-established fields, covering state-or-the-art applications in both signal and image processing, which include optical communication and sensing, wireless communication management, face recognition and facial imaging, solar imaging and feature detection, fractal analysis, and video processing"-- Provided by publisher.
ISBN 978-1-60960-477-6 (hardcover) -- ISBN 978-1-60960-478-3 (ebook) 1. Image processing. 2. Signal processing. 3. Image analysis. 4. Wireless communication systems. I. Qahwaji, Rami, 1972- II. Green, Roger, 1951- III. Hines, Evor, 1957-
TA1637.A69 2011
621.382'2--dc22
2011003448

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Chapter 6
Field Asymmetric Ion Mobility Spectrometry Based Plant Disease Detection:
Intelligent Systems Approach

Fu Zhang
School of Engineering, University of Warwick, UK

Evor Hines
School of Engineering, University of Warwick, UK

Reza Ghaffari
School of Engineering, University of Warwick, UK

Mark Leeson
School of Engineering, University of Warwick, UK

Daciana Iliescu
School of Engineering, University of Warwick, UK

Richard Napier
Warwick HRI, University of Warwick, UK

ABSTRACT

This chapter presents the initial studies on the detection of two common diseases and pests, the powdery mildew and spider mites, on greenhouse tomato plants by measuring the chemical volatiles emitted from the tomato plants as the disease develops using a Field Asymmetric Ion Mobility Spectrometry (FAIMS) device. The processing on the collected FAIMS measurements using PCA shows that clear increment patterns can be observed on all the experimental plants representing the gradual development of the diseases. Optimisation on the number of dispersion voltages to be used in the FAIMS device shows that reducing the number of dispersion voltages by a factor up to 10, preserves the key development patterns perfectly, though the amplitudes of the new patterns are reduced significantly.

INTRODUCTION

Ion Mobility Spectrometry (IMS) has become one of the most successful technologies widely used for detection and analysis of chemical substances (Eiceman & Karpas, 2005). This chapter introduces the potential application of a new generation of IMS, Field Asymmetric Ion Mobility Spectrometry (FAIMS), on greenhouse tomato

DOI: 10.4018/978-1-60960-477-6.ch006

plant health monitoring and the processing of the collected data sets. A continuous experiment was performed using three identical glass boxes simulating greenhouses and each glass box contains a tomato plant. One of the tomato plants was used as the health control and the others were infected with powdery mildew (*Oidium neolycopersici*) and two-spotted spider mites manually at the early stage of the experiment respectively. During the experiment, daily measurements were collected using a commercial FAIMS instrument. The post-processing on the collected data sets indicates that e-nose can be used as a tool to monitoring tomato plant disease. The data processing techniques include, Principal Component Analysis (PCA), Grey Incidence (GI), a key technique in Grey System Theory (GST), and Hierarchical Clustering (HC).

BACKGROUND

IMS Instrument

IMS generally refers to the principles, techniques and equipments designed to analysing gaseous chemical substances based on the transport of ions in electric fields. The foundational studies of IMS started in late 1940s followed by the development of practical IMS instruments in the 1970s (Eiceman & Karpas, 2005). IMS has been used as an inexpensive and powerful technique for the detection of many chemical compounds. For example, commercial IMS units are used at airports worldwide to detect explosives in carryon luggage for aviation security (Eiceman, 2002; Eiceman et al., 2004); tens of thousands of IMS units have been used by military on battlefield to determine chemical warfare agents (Eiceman, 2002); IMS can also be used to characterise drugs (Lawrence, 1989). Over the last decades, the interests of scientific researchers and engineers have been shifted from the conventional IMS to FAIMS, also known as Differential Ion Mobility Spectrometry (DIMS) or Nonlinear Ion Mobility Spectrometry (NIMS) (Eiceman & Karpas, 2005; Shvartsburg, 2009).

Conventional IMS

Conventional IMS, also called linear IMS as linear voltage gradients were used in the IMS units, was based on the determination of the velocities of the ions (Creaser et al., 2004). The basic components of a conventional IMS unit include ionization unit, drifting tube and detection plate. General ionization sources used in the ionization unit include corona discharge, photoionisaiton, electrospray ionisation, and radioactive source (Creaser et al., 2004). In the drifting tube, purified drift gas flows from the detection plate to the drafting tube entrance and a homogeneous electric field, which is generated by a series of charged rings of different voltages, is applied along the drifting tube to attract ions towards the detection plate. The electric field gradient can be alternated to allow the detection of both positive and negative ions generated in the ionization unit (Borsdorf &

Figure 1. Structure of a simplified drifting tube (Borsdorf & Eiceman, 2006)

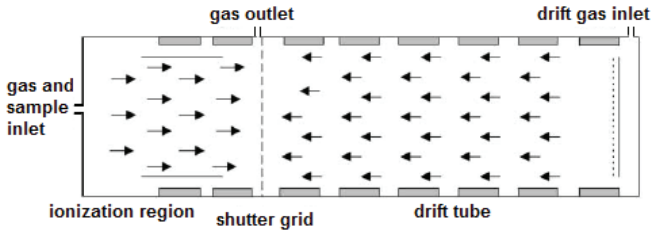
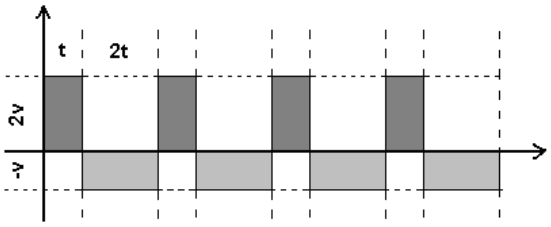


Figure 2. Asymmetric voltage waveform in FAIMS



Eiceman, 2006). Figure 1 illustrates the systematic structure of a simple drifting tube.

The sample molecules are taken into the functional unit by carrier gas and the molecules are ionised to carry positive or negative charges when passing through the ionization unit. The ionised sample molecules and the non-ionised molecules are separated at the entrance of the drifting tube by the drifting gas flow and the electric field applied in the tube. The ions move towards the detection plate along the electric field gradient at random paths and hit the detection plate at different times depending on their shapes, sizes and the strength of electric field gradient. When the ions hit the detection plate, weak current is generated; the strength of the current and the time of occurrence are recorded as the signature of the test sample. The current strength is proportional to the number of ions hit the detector.

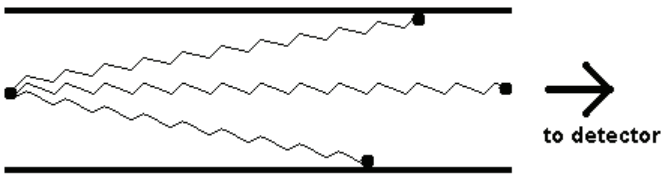
Field Asymmetric Ion Mobility Spectrometry (FAIMS)

A FAIMS unit consists of three major components, which are ionisation unit, filter and detector. Instead of using the electric field gradient generated by charged rings to attract ions towards the detection plate, DMS applies carrier gas to blow ions toward detection plate. A high-voltage asymmetric waveform termed Dispersion Voltage (DV) or Separation Voltage (SV) is applied to a set of parallel electrodes perpendicular to the detection plate. The voltages operate at high frequency asymmetric waveform and the time-voltage integrals above and below the time axis are equal. Figure 2 illustrates the asymmetric voltage waveform (Borsdorf & Eiceman, 2006; Guevremont, 2003; Krylov et al., 2007).

Ions are separable due to the different mobilities in different electric fields. As the voltage on the electrodes alternates, the ions oscillate along the electric field gradient depending on the applied voltage. The different mobilities in different electric field lead to displacement towards one of the electrodes. Ions with positive mobility dependence are displaced in a direction opposite to the ions with negative mobility dependence. Ions that hit the electrodes are neutralised and filtered out; ions reach the detection plate generate weak signals and the signals are recorded for processing. Figure 3 illustrates the possible movements of ions in the filter region. (Borsdorf & Eiceman, 2006)

Apart from the DV, a Compensation Voltage (CV) can be superimposed with the asymmetric voltage waveform to restore ions that are displaced from the centre of the gap between two electrodes. A certain CV is only efficient for ions of specific characteristics, such as charges, masses and shapes. A scan of various CVs and DVs can provide a complete measure of different ions.

Figure 3. Routes of ions in filter region



The OwlStone® LoneStar FAIMS device used in the experiment applies a comb-like ion filter, which carries the asymmetric voltages, to increase the number of filter regions and hence increases the sensitivity of the device. One FAIMS collection circle normally takes about 5 minutes and it consists of a full scan of positive ions and a full scan of negative ions. The dimensionalities of the positive scan and negative scan are identical, the compensation voltages were applied at 512 steps in the range [-6V, 6V] and the dispersion voltages were applied at various scales of a constant voltage in the range [1%, 100%]. Figure 4 presents a typical reading collected by the LoneStar FAIMS device.

Tomato Plant Diseases

Tomato plants suffer from many kinds of pests and diseases, which may cause severe damages to the tomato fruits and reduce the tomato yields. This work concentrates the initial studies on the development of a common tomato disease, powdery mildew, and a major pest, spider mite.

Powdery Mildew

Powdery mildew, caused by the fungus *Oidium neolycopersici*, is one of the most common diseased of tomato (both in greenhouse and fields) around the world. Powdery mildew can cause large damage on tomato production if not controlled properly, especially in the greenhouse tomatoes. The symptoms of powdery mildew are the white patches on the upper surface of leaves in early stages (few days after infection). The disease causes defoliation as the powdery spots develop into brown lesions. Severe infections can lead to premature senescence and significant reduction in fruit size, quality, and yield. Effective control of powdery mildew can be achieved by applying fungicides at early stages of the disease development. (Jacob et al., 2008; Jones et al., 2001)

Spider Mite

There are over 1500 species of spider mites around the world causing damages to hundreds of species of plants. They feed on plant cells and the symptoms are the typical small, yellowish, speckled feeding marks on leaves. Spider mites are tiny, usually less than 1mm. The most well known species of spider mite is *Tetranychus urticae*, also called glasshouse red spider mite or two-spotted spider mite. Spider mites are serious pests in tomato crops due to their great reproduction rate. They can cause severe damages to tomato plants and reduce the yields. If left uncontrolled, they can destroy a plant within a short period of time.

Figure 4. A typical FAIMS reading

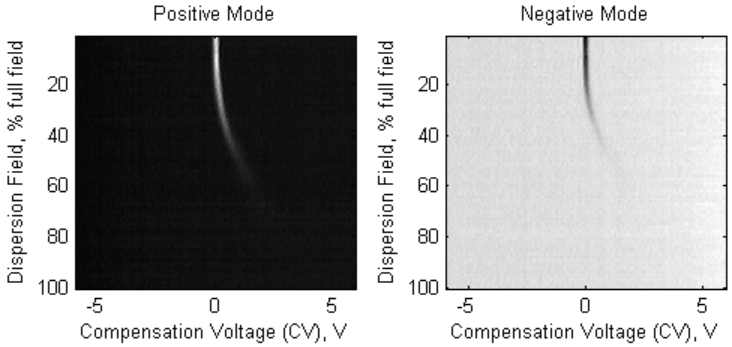
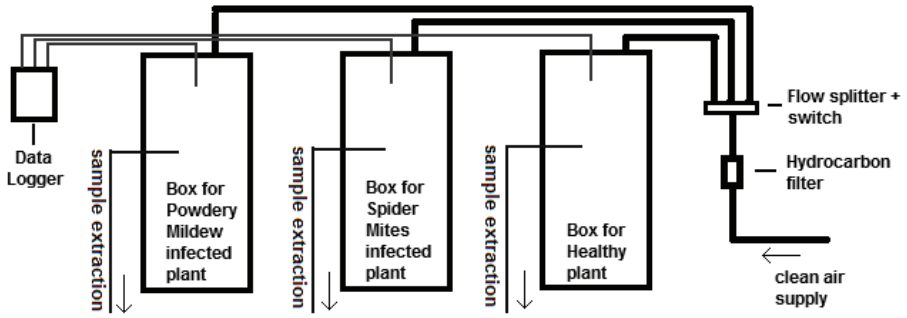


Figure 5. Systematic structure of the experiment



EXPERIMENT AT WARWICK

Initial studies on the early detection of tomato diseases and pests were carried out at the Engineering School of the University of Warwick in 2008. The objective of experiment was to study the development of powdery mildew and spider mites on tomato plants by measuring the chemical volatiles emitted by the infect plants using a FAIMS device. Three Espresso tomato plants were used in the experiment and the plants were provided by the Warwick Horticultural Research Institute. The plants were growing in a controlled greenhouse for 6 weeks since seedling and then moved into three glass boxes (150cm H, 50cm W, 50cm D), namely Box 1, Box 2 and Box 3, in a controlled chamber. The tomato plants in the glass boxes were under different treatments. Box 1 contains the plant which will be infected by powdery mildew; Box 2 contains the plant which will be infected by spider mites and Box 3 the plant which will be used as the health control. The glass boxes are specifically made for this experiment using glass panels and alloy frames to minimize the gas exchange with the external environment and the possibility of contamination. The Day and Night times in the chamber were set to be 16 hours and 8 hours respectively. Continuous ventilation was provided to the glass boxes by pressured gas at equal flow rates. The ventilation was turned off one hour before taking measurements and switched back on after the measurements. The plants were watered manually when necessary. Each glass box has been fitted with a humidity and temperature logging device which saves the current humidity and temperature inside the box every 5 minutes. During the experiment, repetitive experimental readings were taken daily during the Day hours in the chamber. Figure 5 illustrates the systematic structure of the experiment.

PROCESSING TECHNIQUES

As mentioned in the previous section, the data processing techniques presented in this work include Grey Incidence (GI), Hierarchical Clustering (HC) and Principal Component Analysis (PCA). PCA is used to extract the key information from a batch of collected FAIMS readings. GI and HC are used together to optimise the FAIMS data collection process by reducing the amount of data to be collected.

Principal Component Analysis (PCA)

PCA is a common mathematical technique which is used to reduce the dimensionality of a data set consisting of a large number of variables, while retaining the variation of the original data set as much as possible. PCA transforms a dataset into a new system of coordinates linearly with the possibility of reducing the number of dimensions. The first dimension of the new coordinates

system is called the first principal component, which retains the highest variations of the original data set; the second dimension retains the second highest variation, and so on (Jolliffe, 2002). The principal components are determined by following four steps:

Step 1: subtract mean value

$$X' = X - \overline{X} \quad (1)$$

where $\overline{X}$ is the mean of a data set X, and X' is the new data set.

Step 2: calculate covariance matrix

$$\mathrm{cov}(X) = \frac{XX^T}{N-1} \quad (2)$$

where X is the matrix consisting of data sets, X^T is the transpose of matrix X and N is the size of each data set.

Step 3: calculate eigenvectors and eigenvalues

$$Ax=\lambda Ix \quad (3)$$

$$(A-\lambda I)x=0 \quad (4)$$

$$\det(A-\lambda I)=o \quad (5)$$

where A is the square covariance matrix derived in step 2, I is the identity matrix, λ is an eigenvalue and x is the eigenvector.

Step 4: reorder the eigenvectors based on their associated eigenvalues, highest to lowest, which represent the explained variance of the eigenvectors.

The matrix of the ordered eigenvectors represents the original data transformed into the new coordinate system. The eigenvector of the largest eigenvalue is the first dimension (the first principal component) in the new coordinate system and it accounts for the most information (variance) of the dataset; the eigenvector with the second largest eigenvalue is the second dimension (the second principal component) of the new coordinate system; and so on. As the first few principal components usually account for the most information (variances) of the original dataset, they could be used to represent the original dataset. For visualisation purposes, two or three sets of principal components with the most significance are normally selected to express the original dataset graphically.

In this work, PCA is used to reduce the complexity of the FAIMS readings by extracting the first principal component from each measurement and visualise the development trend of the complete dataset collected during the experiment.

Grey Incidence (GI)

GI is one of the key techniques under the scope of Grey System Theory (GST). GST is a series of techniques that initially appeared in the 1980s including Grey Equations, Grey Incidence, Grey Systems Modelling, Grey Prediction, Grey Decisions and Grey Control. The theory states that the information available is always uncertain and limited due to noise. After over 2 decades of development, GST had been applied successfully in various scientific areas, including industry, agriculture, economics, etc. (Deng, 1989; Liu & Lin, 2005).

As mentioned previously, the LoneStar FAIMS was used to collect data under 100 different dispersion voltages and it has the ability to operate under customised dispersion voltages. In this work, GI analysis is used to find the closeness between two signals collected at different dispersion voltages, represented by a single GI value. Higher similarity (closeness) levels would generate higher GI values, and vice versa. An upper triangular GI matrix can be generated to represent the similarities of all

possible pairs of signals. In the case of this work, the size of the GI matrix is 100 by 100 holding the effective GI values of 4950 pairs of signals.

The GI values can be calculated by the following steps. Assume that there exist two signals, each of *n* elements.

$Y = (y(1), y(2), \ldots, y(n))$

$X = (x(1), x(2), \ldots, x(n))$

where *Y* and *X* are the signals collected at different dispersion voltages.

Step 1

The first step is to find the initial image of each sequence using

$Y' = Y/y(1) = (y'(1), y'(2), \ldots, y'(n)) = (y(1)/y(1), y(2)/y(1), \ldots, y(n)/y(1))$

$X' = X/x(1) = (x'(1), x'(2), \ldots, x'(n)) = (x(1)/x(1), x(2)/x(1), \ldots, x(n)/x(1))$

where *Y'* and *X'* are the image sequences or the initial sequences of the original signals, *Y* and *X*.

Step 2

Find the difference sequences between the target sequence and the influencing factors using

$\Delta = X' - Y' = (\Delta(1), \Delta(2), \ldots, \Delta(n)) = (x'(1) - y'(1), x'(2) - y'(2), \ldots, x'(n) - y'(n))$

Step 3

Compute the global maximum and minimum differences using

$M = max(max(\Delta(j)))$

$m = min(min(\Delta(j)))$

Step 4

Calculate the incidence coefficients using Equation (6)

$$\lambda(k) = \frac{m + \zeta M}{\Delta(k) + \zeta M} \tag{6}$$

where $\lambda(k)$ represents the incidence coefficients of the kth element and ζ is a user defined factor in the range [0, 1] and generally taken to be 0.5.

Step 5

Calculate the GI using equation (7)

$$r = \frac{1}{n}\sum_{k=1}^{n} \lambda(k) \tag{7}$$

Thus the GI between the signals *X* and *Y* can be derived as

$\gamma = (\gamma(1) + \gamma(2) + \ldots + \gamma(n))/n$

Hierarchical Clustering Analysis (HCA)

HCA is a powerful exploratory methods widely used in many disciplines. It consists of a set of statistical techniques which is capable of finding the underlying structure of objects and separating the objects into constituent groups. The grouping of objects is based on a multilevel hierarchical tree or dendrogram, which is constructed from the similarities or distances between the objects

(Almeida et al., 2007; Leung et al., 1999). In this work, HCA is performed using the built-in commands provided by the Statistics Toolbox in Matlab®. The HCA in Matlab® takes the following three steps:

Step 1: Find Dissimilarities (Distances) between Data Pairs

The distance of each data pair is represented by a numerical value and a distance matrix is used to hold the distances of all the data pairs. Similar data would generate lower distance. There exist many algorithms calculating distance under various metrics. Common distance metrics include Euclidean distance, city block distance (Manhattan distance) and so on. Different algorithms generate different distance values. For example, the distance between points (1, 1) and the origin is $\sqrt{2}$ under Euclidean metric and 2 under city block metric (MathWorks, 2009). Application of various metrics may lead to distinct HCA results. In this work, the distance between the data pair is calculated using the GI instead of the conventional metrics. We may call it the Grey metric. As the GI is a measure of the data closeness in the range [0, 1], the complement of GI, 1-GI, is a better representation of the distance and is used in the following procedures.

Step 2: Group Data into a Binary Hierarchical Cluster Tree

In this step, the distance matrix generated in Step 1 is used by the linkage function to construct the hierarchical cluster tree. The linkage is an iterative program, it links single objects to each other in groups gradually. The first connection is the data pair of the highest similarity. Once the first group is constructed, the similarities between the new group and the remaining data need to be recalculated. The resulting hierarchical tree is usually illustrated using a dendrogram where each linkage step is represented by a U-shaped connection line and the height of the connection line represents the distance between the two objects connected. (MathWorks, 2009)

Step 3: Determine Clusters

The hierarchical cluster tree generated in Step 2 contains the completed cluster division information. As the connection lines represent the distances between the connected nodes or sub-trees, by selecting the appropriate cut-off distance, the data can be divided into the desired number of groups. (MathWorks, 2009)

In this work, the HCA is used to group the signals collected under various dispersion voltages based on their similarities. The signals in the same group are believed to have similar patterns and the entire group can be represented by one of the signals within the group.

RESULTS AND DISCUSSIONS

The tomato plants were moved in the dedicated glass boxes three days before the infection. Since then, repetitive measurements were taken every day. On the day of infection, the tomato plant in Box 1were infected manually by shaking the powdery mildew source (a tomato branch severely infected) over the plant for several seconds; thirty spider mites were handpicked using a small brush from the spider mite source (a pea plant used to feed the spider mites) and placed on the tomato plant. During the experiment, powdery mildew spots became visible on several leaves about 5 days after infection. However, the 30 spider mites cannot be found after the day of infection and the signs of their survival and reproduction were not obvious either. The health control plant was healthy throughout the experiment.

Figure 6. Development of the powdery mildew infected plant

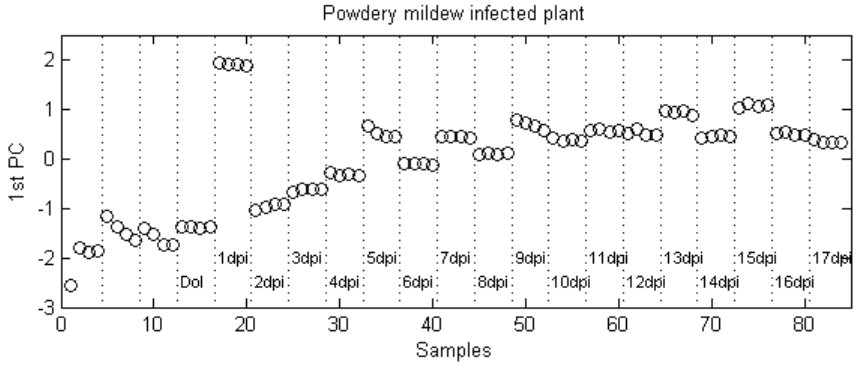
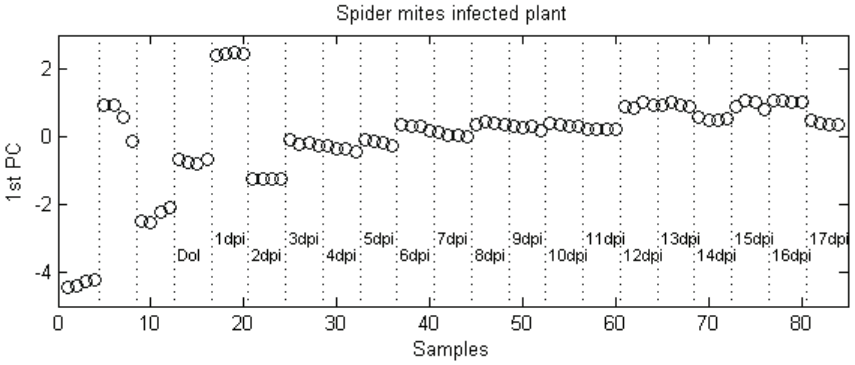


Figure 7. Development of the spider mites infected plant



Disease Development Patterns

To discover the development trends of the powdery mildew and spider mites on the tomato plants, 4 readings were selected from each set of the daily measurements and the PCA was used to extract the first principal components from the readings. The PCA results show that, on average, the first principal components carry about 51.43% variability; the second principal components carry only

Figure 8. Development of the health control plant

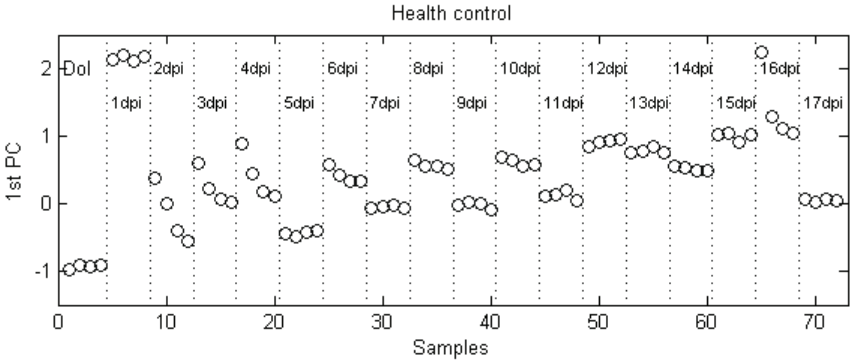
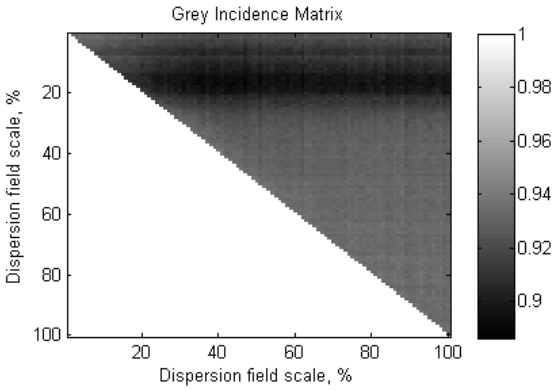


Figure 9. Grey incidence matrix



9.85% variability; the third principal component carry only 1.16% variability and so on. By placing all the first principal components together in the same graph, the development trend of the infection can be discovered. Figures 6, 7 and 8 illustrate such trends of the three tomato plants.

In Figure 6, the powdery mildew infected plant, a clear increment pattern can be observed. The pattern can be divided into three periods, which are the healthy period, the development period and the stable period. On and before the day of infection (DoI), the healthy period, the values of the 1st PCs are about the same. After the infection, the development period, the 1st PCs fluctuate and follow a gradual increasing pattern, expect the data collected on the 1st day post infection (DPI), which is invalid due to instrumental errors. The increasing pattern may due to the development of the disease; the condition of the plant was getting worse day by day. Since the 10th DPI, the stable period, the 1st PCs fluctuate around a constant value. In these days the condition of the plant is severe and the relevant volatiles in the glass box reached a certain density.

The data of the spider mites infection plant illustrated in Figure 7 can be divided into three periods as well. During the healthy period, the 1st PCs are not stable and show distinct differences. After the infection, a weak increasing pattern is noticeable until the 12th DPI. From the 12th DPI onwards, the 1st PCs show a flat pattern.

In Figure 8, the health control plant, the data points form a gradual increasing pattern. This may be due to the fact that as the tomato plant grows the leaves at the bottom of the stem turn yellow and fall off eventually. The leaves turned yellow may emit volatiles different from those emitted by the healthy green leaves.

Dispersion Voltages Optimisation

As mentioned in the previous section, the dispersion voltages applied by the FAIMS can be set manually and a full scale scan (100 different dispersion voltages) takes about 5 minutes. By reducing the number of dispersion voltages, the scanning speed of the FAIMS can be increased dramatically. The GI is used to calculate the

Figure 10. HCA dendrogram calculated using the GI distance matrix

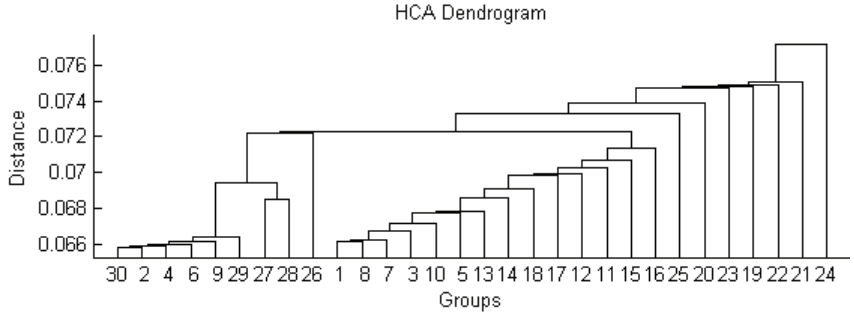
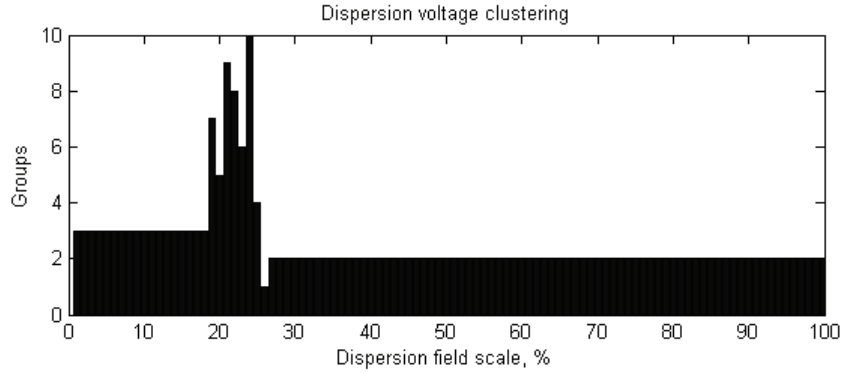


Figure 11. Clustering of dispersion fields



similarities between the signals collected under various dispersion voltages. Figure 9 illustrates the Gray Incidence matrix (similarity matrix) derived using all the data.

As the linkage procedure of HCA acquires the distance matrix rather than the similarity matrix, the complement of the GI matrix is used in the linkage procedure. Figure 10 illustrates the dendrogram generated based on the linkage.

The original data was collected under 100 different dispersion voltages. If we aim to reduce the number of dispersion voltages by a factor of 10, we need to divided the 100 different dispersion voltages into 10 (100/10) groups base on the dendrogram and find a representative dispersion voltage for each group. Figure 11 illustrates the resulting clusters.

The first element of each group can now be used as the representative of that group. By extracting the signals collected under the selected dispersion voltages and perform the PCA again, patterns similar to the ones obtained from the original can be produced. Figure 12 illustrates the disease development pattern of the powdery mildew infected plant generated using 10 dispersion voltages, instead of 100. The new development retains about 96% of the original pattern.

Table 1 summarises the performance of the HCA with the application of various numbers of dispersion voltages. The reduced number of dispersion voltages work well as it always retains over 90% of the similarities of the original development patterns.

CONCLUSION

This chapter presents the initial studies on the responses of greenhouse tomato plants to two kinds of diseases and pests, the powdery mildew

Figure 12. Disease development pattern generated using reduced number of dispersion voltages

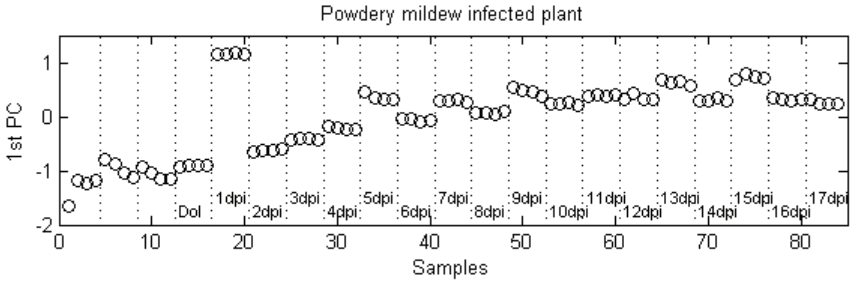


Table 1. Performance of the selection of various numbers of dispersion voltages

	Various numbers of remaining DVs				
	10 DVs	**20 DVs**	**30 DVs**	**40 DVs**	**50 DVs**
Powdery mildew infected plant	96.0%	92.7%	95.6%	96.9%	97.4%
Spider mites infected plant	95.7%	96.9%	96.4%	96.4%	96.9%
Health control	96.2%	93.5%	95.3%	95.7%	96.5%
Average performance	96.0%	94.4%	95.8%	96.3%	96.9%

and spider mites, by measuring the chemical volatiles emitted from the tomato plants using a FAIMS device. Post processing of the collected data shows that clear increment patterns can be observed using PCA. This may be due to the development of the diseases.

Further analysis on the collected data using GI and HCA shows good results. By grouping the signals collected under the full FAIMS scan (100 various dispersion voltages) into various numbers of groups and selecting one representative dispersion voltage from each group, the amount of measurement can be reduce effectively by various factors. Statistical results show that reducing the amount of dispersion voltages by a factor between 2 and 10 can always retain over 90% of the variation of the original development patterns.

REFREENCES

Almeida, J.A.S., Barbosa, L.M.S., Pais, A.A.C.C. & Formosinho, S.J. (2007). Improving hierarchical cluster analysis: A new method with outlier detection and automatic clustering. *Chemomotrics and Intelligent Laboratory Systems,* 87.

Borsdorf, H., & Eiceman, G. A. (2006). Ion mobility spectrometry: Principles and applications. *Applied Spectroscopy Reviews*, *41*, 323–375. doi:10.1080/05704920600663469

Creaser, C. S., Griffiths, J. R., Bramwell, C. J., Noreen, S., Hill, C. A., & Thomas, C. L. (2004). Ion mobility spectrometry: A review. Part 1: Structural analysis by mobility measurement. *Analyst (London)*, *129*, 984–994. doi:10.1039/b404531a

Deng, J. (1989). Introduction to Grey system theory. *Journal of Grey System*, *1*, 1–24.

Eiceman, G. A. (2002). Ion-mobility spectrometry as a fast monitor of chemical composition. *Trends in Analytical Chemistry*, *21*, 259–275. doi:10.1016/S0165-9936(02)00406-5

Eiceman, G. A., & Karpas, Z. (2005). *Ion mobility spectrometry* (2nd ed.). CRC Press. doi:10.1201/9781420038972

Eiceman, G. A., Krylov, E. V., Krylova, N. S., Nazarov, E. G., & Miller, R. A. (2004). Separation of ions from explosives in differential mobility spectrometry by vapor-modified drift gas. *Analytical Chemistry*, *76*, 4937–4944. doi:10.1021/ac035502k

Guevremont, R. (2003). High-field asymmetric waveform ion mobility spectrometry (FAIMS). *Canadian Journal of Analytical Cciences and Spectroscopy*, *49*, 105–113.

Jacob, D., David, D. R., Sztjenberg, A., & Elad, Y. (2008). Conditions for development of powdery mildew of tomato caused by Oidium neolycopersici. *Ecology and Epidemiology*, *98*, 270–281.

Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). Springer.

Jones, H., Whipps, J. M., & Gurr, S. J. (2001). The tomato powdery mildew fungus Oidium neolycopersici. *Molecular Plant Pathology*, *2*, 303–309. doi:10.1046/j.1464-6722.2001.00084.x

Krylov, E. V., Nazarov, E. G., & Miller, R. A. (2007). Differential mobility spectrometer: Model of operation. *International Journal of Mass Spectrometry*, *266*, 76–95. doi:10.1016/j.ijms.2007.07.003

Lawrence, A. H. (1989). Characterization of benzodiazepine drugs by ion mobility spectrometry. *Analytical Chemistry*, *61*, 343–349. doi:10.1021/ac00179a012

Leung, S. C., Fung, W. K., & Wong, K. H. (1999). The identification of credit card encoders by hierarchical cluster analysis of the jitters of magnetic stripes. *Science & Justice*, *34*, 231–238. doi:10.1016/S1355-0306(99)72054-X

Liu, S., & Lin, Y. (2005). *Grey information: Theory and practical applications*. Springer.

MathWorks T. (2009). *Statistics toolbox 7: User's guide*. The MathWorks.

Shvartsburg, A. A. (2009). *Differential ion mobility spectrometry: Nonlinear ion transport and fundamentals of FAIMS*. CRC Press.

KEY TERMS AND DEFINITIONS

Field Asymmetric Ion Mobility Spectrometry (FAIMS): Is a new technology for quick and accurate detection of a broad range of chemicals at low quantities (below part per billion) with high confidence. The FAIMS operates by applying high frequency asymmetric electric fields on the samples carried by appropriate carrier gas passing through the filtering channel. Ions are separable due to the different motilities under different electric fields. Ions that hit the electrodes in the filtering channel are neutralised and filtered out; those reach the detection plane generate weak signals, which are recorded as the fingerprint of the sample.

Grey Incidence (GI): Is one of the key techniques under the scope of Grey System Theory (GST). It is an alternative way to measure the distances or similarities between data sequences.

Hierarchical Clustering Analysis (HCA): Is a powerful exploratory methods widely used in cluster analysis. It consists of a set of statistical techniques capable of finding the underlying structure of objects and group together objects that are 'close' to one another.

Principal Component Analysis (PCA): A common mathematical technique developed in 1901. It can reduce the dimensionality of a data set by transforming it to a new coordinate system and retaining the dimensions of high variances. The first dimension of the new coordinate system holds the highest variation of the original data set and the second dimension retains the second highest variation, and so on.

Powdery Mildew: (*Oidium neolycopersici*)**:** A type of common fungal disease around the world that affects a wide range of plants, including grapes, onions, tomatoes, and so on. It appears as dusty gray or white coating on leaf surfaces and other parts. It can cause large damage on fruit production if not controlled properly.

Spider Mites: Are tiny pests, usually less than 1mm. There are over 1500 species of spider mites causing damages to hundreds of species of plants. The most well known species is *Tetranychus urticae*, also known as glasshouse red spider mite or two-spotted spider mite.