

Chapter 6

Volatile Organic Compounds (VOCs) for Noninvasive Plant Diagnostics

Alexander A. Aksenov,¹ Ana V. Guaman Novillo,²
Sindhuja Sankaran,³ Alexander G. Fung,¹ Alberto Pasamontes,¹
Frederico Martinelli,⁴ William H. K. Cheung,¹ Reza Ehsani,³
Abhaya M. Dandekar,⁵ and Cristina E. Davis^{*,1}

¹Mechanical and Aerospace Engineering, University of California, Davis,
One Shields Avenue, Davis, California 95616, U.S.A.

²Signal and Information Processing for Sensing Systems,
Institute for Bioengineering of Catalunya,
Carrer Baldori Reixac, 4, 08028, Barcelona, Spain

³Citrus Research and Education Center, University of Florida,
700 Experiment Station Road, Lake Alfred, Florida 33850, U.S.A.

⁴Department of Agricultural and Forest Sciences, University of Palermo,
Viale delle scienze, 90128, Palermo, Italy

⁵Plant Sciences, University of California, Davis, One Shields Avenue,
Davis, California 95616, U.S.A.

*E-mail: cedavis@ucdavis.edu.

Plant systems are frequently important agriculture commodity crops, and it can be critical to track emerging pathogen infections or nutrient deficiencies that can limit or prevent food production. While some methods have been developed for in field monitoring, they are frequently invasive and not rapid. Volatile organic compounds (VOCs) are produced by all plant systems, and present a possible noninvasive window through which we can monitor plant health. Here we review different methods of VOC sampling and detection and elucidate potential of VOC monitoring for precision agriculture.

Introduction

The volatile organic compounds (VOCs) emitted by a plant are a significant and important part of the plant's life cycle, and are reflective of the physiological status of the plant. The emission of VOCs may occur as a side effect of ongoing metabolic processes or certain chemicals could be emitted for a specific purpose, such as defense or signaling. For example, when plants are feed upon by insects or herbivores, their direct defensive response is frequently to release VOCs (1). These emitted volatiles are often referred to as induced VOCs (IVOCs), and these compounds can be released from the surface of the plant's leaves, fruits, and roots. It has been reported that the IVOCs play an important role in plant-to-plant communication (2–4), herbivore defense (1, 5, 6), and has been shown to aid resistance to biotic (biological) stress. This change in a plant's VOC profile is not necessarily only induced under biotic stresses, but can be caused by abiotic (non-biological) factors as well. There are thousands of emitted compounds, and the composition of these released VOC mixtures may relate important biochemical information of the underlying metabolic processes within the plant system (7–9). The expressed VOCs can be collected and measured using analytical methods to provide a momentary snapshot of the plant's health status at any given time. Therefore, measuring and interpreting the VOCs is an attractive avenue as a noninvasive and rapid way of monitoring the physiological processes in plants, including: flowering (10), ripening (11), maturation (12) and stress (13, 14). Plant VOCs can be sampled *in situ* from whole plants, fruits, and leaves, or directly from detached plant tissues (15).

Due to the extreme complexity of VOC patterns emitted by plants, it is challenging to directly correlate VOC production with a specific cause. Four main aspects of this analysis need to be comprehensively addressed in order to discern specific changes in VOC patterns and attribute them to possible causes. First, the alterations in the biological state of the plant need to be recognized and attributed to some cause. Multiple biotic and abiotic causes of VOC alteration may be occurring at the same time. Nutrition status, presence of a pathogen, infestation, diurnal cycle, geographic location, plant varietal, and other factors can all have an effect on VOC production. Furthermore, various parts of the plant may be producing VOCs at different rates. For example, young flush can be expected to be more metabolically active than older foliage. While some of the factors may be not significant, others will greatly affect VOC production. Certain factors affect VOC production differently in conjunction with other conditions, such as when nutrient stressed plants are exposed to a pathogen. It is critical that the study design take all of these aspects into account, so that causality of the VOC production can be attributed to the correct factor responsible for the compounds being released.

VOC sampling is a second important aspect of noninvasive diagnostics that needs to be addressed for practical field use. Biogenic variations that are indicative of certain changes in the plant may be subtle, and the use of improper sampling techniques can introduce large error. It is also possible for a sampling method itself to bias the VOC collection towards certain compounds, which may or may not be

important in that particular plant system. Either of these cases could ultimately result in correlations of interest being obscured due to sampling variation or error.

The third aspect important to the measurements is the selection of an appropriate analytical chemistry technique employed in the VOC study. It is essential to choose an appropriate chemical detection method that enables the identification of sampled chemical differences. Ideally, detection and quantification of all of the chemicals in a very complex mixture is sought. However, due to technological limitations as well as cost and time of analysis considerations, a method that provides only limited information might be chosen. For example, some devices such as the electronic nose (“E-nose”) do not identify the actual chemicals released by the plants. These devices only output a “fingerprint” pattern of the sensor response. This can be problematic if confounding chemicals are later present in unknown sample, because the sensor has rather limited chemical information. This highlights the importance of using methods that unambiguously identify the biomarker VOCs of interest.

Finally, the information provided by a sensor needs to be processed, interpreted and correlated to a specific biological status, or sets of conditions that are being studied. Often a very sophisticated data analysis is required to extricate complex correlations, especially if the signal is “noisy” or the biological variability is substantial. Many signal processing methods and algorithms have been developed to attempt this for VOC research in plants, animals and humans. In plant VOC research and biological studies, it is essential to address all the above aspects for a robust study. Failure to appropriately address any of the factors may lead to negative testing results and/or overlooked cause-effect relations. Despite the obvious challenges, great progress has been made to date. In the following sections, we will consider each of the four key factors needed for VOC detection success, discuss specific issues, and recommend future advances needed for state-of-the-art plant VOC monitoring technologies.

Existing Visual Strategies of In-Field Plant Health Assessment

There are several methods of assessing in field plant health. Visual inspection by the human eye is the most common method, while an alternative strategy is to use sensing technologies. Visual inspection and assessment of overall plant health status or specific issues in the field is called “scouting.” Scouting is the most commonly used procedure for monitoring overall crop health. Scouting is done on foot, while riding on all-terrain vehicles, while standing on raised platforms mounted on a vehicle, or by a combination of these methods (16). Scouts identify plant stress symptoms as either abiotic or biotic. Abiotic stress symptoms include nutrient deficiencies, water stress, freeze damage, and lightning damage. Biotic stress symptoms can include bacterial diseases, viral diseases, fungal diseases, and blight. Scouts use a guide prepared by plant pathologists or field managers to identify symptoms specific to the crop being inspected. Symptoms may be located on leaves, branches, flowers, fruits or roots. Often pathogen stress can mimic nutritional stress, or result in nutritional stress. For example, the Huanglongbing disease in citrus could be easily confused with zinc deficiency as both cause leaf

blotching. The training of scouts determines their efficiency, and over time scouts develop experience in identifying plant stress symptoms.

In specialty crops, especially tree fruits, global positioning systems (GPS) are used to register the position of stressed trees for additional monitoring and disease control (17). Scouting is usually performed to search for incidence of diseases rather than abiotic stress conditions. Typically, the crops are well maintained with respect to nutrients and water to achieve maximum productivity. The biotic stress conditions are more unpredictable and can have greater economic impact. Scouting for plant health usually accompanies scouting for other reasons, such as presence of insects/vectors, weeds, immature fruit fall, and others. The economics of field scouting rely on the crop of interest, location of field sites, total acreage to be monitored, availability of scouts, and their experience. Sometimes, scouts are paid more based on their level of expertise. Scouting is prone to human error caused by fatigue or effects of environmental factors on human sensing system. It can also be a complex, time-consuming and expensive process, depending on the crop acreage scouted. Alternatively, minimally-invasive sensing technologies can offer a less expensive and more accurate means to rapidly determine plant stress.

Recent advancements in noninvasive sensor technologies have led to new methods of sensing plant health. The best type of sensor is one that is accurate, rapid, specific and sensitive, and will detect the early onset of plant stress symptoms (18). Optical sensors are probably the most utilized technique for monitoring plant health. In the field of optical sensing, several spectroscopic and imaging techniques are being investigated to create a practical tool for a large-scale, real-time, in field assessment of plant health. Examples of optical techniques include: fluorescence imaging (19–22), multispectral or hyperspectral imaging (23–26), infrared spectroscopy (27, 28), fluorescence spectroscopy (29–32), visible/multiband spectroscopy (33–36), and nuclear magnetic resonance spectroscopy (37).

Optical sensors accurately detect specific health conditions because different regions of the electromagnetic (EM) spectrum are sensitive to certain physiological changes that occur in plants. For example, the visible region of the EM spectrum is sensitive to chlorophyll and other pigments. Unhealthy canopies reflect less radiation, while healthy canopies reflect higher radiation in the near-infrared region due to the scattering of light caused by the leaf structures. Similarly, the mid-infrared region can be used to identify the chemical signature based on the molecular vibrations of different leaf components, such as carbohydrates. Thermal infrared and terahertz frequencies work well in detecting water-related changes in plant canopies.

The desired spectral features, field-of-view and data resolution can be achieved by selecting the appropriate sensors and sensor platform that mechanically support the sensors and carry the sensors in the field. The sensor system can be ground-based or aerial in nature. Ground-based sensors can be handheld or mounted on different types of mechanical support platforms (Figure 1a-c). The selection of a suitable platform depends on the purpose of the sensor application, crop characteristics, and crop distribution. Another sensor support system is to mount the sensors on tractor components that are used for regular agricultural operations.



Figure 1. Sensor platforms. (a) Sensors mounted in front of a utility vehicle, (b) sensors mounted on a retractable mast connected to an agricultural vehicle, (c) high-lift platforms attached to a vehicle, and (d) unmanned aerial vehicle (not to scale). [Source: Reza Ehsani]

Several different types of aerial methods are used for monitoring crop health. These include use of satellites, fixed-wing planes, helicopters and unmanned aerial vehicles (UAVs). The major benefit of UAV-based systems is the ability to acquire high-resolution aerial images by adjusting the flying altitude (23). Different types of UAVs, also referred to as unmanned aerial systems, have been described in literature (38–40).

Some of the commonly used sensing techniques such as visible-near infrared spectroscopy, fluorescence spectroscopy and thermal imaging can be used to remotely sense plant health in field conditions (41). There are a few commercial handheld sensors available that can be utilized for agricultural applications such as the Green Seeker RT200, SPAD-502 Leaf Chlorophyll Meter, Multiplex®, Dualax®, and Crop Circle ACS-430 Active Crop Canopy Sensor. These sensors are based on visible-near infrared or fluorescence spectroscopy. The major advantage of visible-near infrared spectroscopic sensors is that they can be adapted for detection of specific stress conditions. Fluorescence spectroscopy is suitable for identifying the overall stress conditions based on fluorescence from chlorophyll and other leaf structures.

The aerial sensor platforms are often integrated with visible-near infrared multispectral imaging and thermal imaging systems, although fluorescence imaging remains challenging. The thermal imaging systems are most applicable

for water-stress detection, although they have been applied for other stress conditions. One of the major challenges in the application of optical sensor technology is the need for proper calibration and corrections. However, the technology is constantly improving, and presently several sensors have auto-calibration or active detection. In auto-calibrated sensors, the sensors are integrated with light-sensitive sensors to calibrate the instruments for existing light conditions. Another way to compensate for varying sunlight intensity is by using a light source with the sensors. Each sensing technique has its unique advantages and limitations. For example, although mid-infrared spectroscopy offers the benefit of identifying chemical signature related to plant physiological changes, the method requires some form of sample preparation. One method to overcome challenges associated with each individual sensor could be the application of sensor fusion approach. Multiple sensors systems can be integrated to complement each other, so that one method can eliminate the limitations of the other; thus creating a robust ideal sensor system with enhanced capabilities. Another critical step in plant health monitoring is the development of pattern recognition algorithms for proper analysis and interpretation, and selecting the right sensor system along with development of robust statistical models will allow early in field detection of plant stress.

Nucleic Acid-Related Methods of Plant Disease Detection

Analytic biochemistry assays are a very important diagnostic tool in plant pathology. Enzyme-linked immunosorbent assays (ELISA) with polyclonal and monoclonal antibodies have been used for plant disease detection in vector and plant tissues (42). However, the lack of suitable commercial ELISA kits, the low titers and the non-homogeneous presence of systemic (vascular) pathogens in plant tissues strongly limits their application on a large scale. DNA-based techniques have been developed for the detection of a variety of pathogens in insects and plants (1) using DNA hybridization methods. However for woody plants, the reliability of these methods is lower when the pathogen titers are extremely low and non-homogeneously distributed. Restriction fragment length polymorphism (RFLP) analysis of the amplicons has also been employed using appropriate restriction enzymes to detect subgroup relationships, such as for the detection of *Flavescence dorée* (43). Tremendous improvements have been made using polymerase chain reaction (PCR)-based techniques (44), allowing the development of universal PCR assays for the detection of many pathogen types, or specific pathogens associated with different diseases (45). Another current diagnosis method relies on real time PCR (RT-PCR) detection of a putative pathogen candidate, and this is now commonly used to confirm diagnosis made with less specific and sensitive methods (such as symptom visualization and ELISA). Different PCR strategies such as conventional PCR, nested PCR, and semi-quantitative PCR have been developed, although the qRT-PCR analysis (SYBR Green real time PCR and Taqman assays) are more sensitive and have been used for both qualitative results and for quantifying pathogen load. The

analysis is easy, rapid (from DNA extraction to qRT-PCR results can take a minimum of 2-3 hours), cheap (few dollars per sample), extremely sensible and with low false positive results.

However some disadvantages exist. First, qRT-PCR methods rarely identify pathogens during the early asymptomatic stage of the disease. Although PCR methods have tremendously helped diagnosis confirming visual symptomatic observations, they are not reliable at asymptomatic stage. Yet, early identification of asymptomatic plant and trees is desirable to detect primary infection sources and thus control secondary spread by infected plant removal. As occurs for other vascular-limited pathogens (46), long-distance, asymptomatic primary spread of plant disease are caused by pathogens with very low titers, undetectable with current real-time PCR technologies. PCR-based detection methods usually allow reliable and sensitive detection of the putative pathogen only when symptoms are present (47, 48). At the asymptomatic stage, different PCR methods may also often yield contrasting results. This is due to the technology's detection limits (down to 4.6×10^2 L/g) and the presence of PCR inhibitors, as shown for the detection of *Candidatus liberibacter asiaticus*. A second substantial issue is related to the high mutation rates of all pathogens (especially viruses). Real time assays frequently cannot detect emerging or mutated pathogen strains, and new primers need to be continuously designed for them. A third important issue is when qRT-PCR is used to detect a systemic pathogen having non-homogenous distribution in plant organs. Sampling will greatly affect diagnostic results, and this exponentially increases the number of samples required for proper detection. This will substantially increase costs for reagents and labor, and can also lead to long turnaround times for results. Sampling different part of plant (such as different points in the foliage) can have effect on results as pathogens can be unevenly distributed, be not present in all parts of a plant or concentrated higher in some parts compared to others. Sampling errors can not be avoided in such cases. A forth disadvantage is the need for expensive lab infrastructure and operation costs due to the high costs for equipment and the need of highly-skilled personnel. Finally, every DNA-based methods would detect the pathogen without giving any information on its living status and the progression of the disease. On contrast, in vivo methods such as the analysis of plant volatile emissions are closely linked with disease status. An alternative microarray technology (lateral-flow microarray) has been used to directly detect pathogens in the orchard, although the range of sensitivity is lower and improvements in this technology are still needed.

Rapid and specific induction of messenger and small RNAs is also a potential early diagnostic biomarker for disease. This complexity can now be analyzed to an unprecedented depth using new DNA sequencing methods which reveal very rare mRNA, splice variants, allelic variants, and single-nucleotide polymorphism (SNPs). Analysis of the deep transcriptome using network theory will help define gene regulatory networks and identify key disease-specific biomarkers usable for in field sensor devices that detect not only transcripts, but also volatiles. In any biological system there is a complex relationship between genes, RNA, proteins, metabolites, and emitted volatile organic compounds (VOCs) that we expect network analysis to reveal. Changes to individual elements may trigger

an expression of others or alter signal transduction pathways, forming a complex interdependent network. Comparing interaction networks of healthy and diseased plants will identify pathways that allow identification of underlying early disease biomarkers. In the short term, pathogen biomarkers can assist plant removal decisions for infected trees. In the medium-term, this can help define pathways to target for therapeutic treatments. All plants organs (leaves, seedlings, bud, flower, fruit) are likely early biological warnings of pathogen infection, and all are a source of the biomarkers that can be monitored in order to detect pathogen at the earliest possible infection stage to limit disease spread.

Volatile Organic Compound (VOC) Sampling Methodologies

A number of different VOC sampling methodologies exist, and the most suitable method for a specific experiment should be chosen according to the biological system under investigation. In all of the existing methodologies, most sampling methods aim to minimize the number of artifacts from the environment. These methods also try to obtain the “best” VOC profiles that reflect the naturally occurring VOC distributions of the plant. These methodologies can be divided into two broad categories, namely active and passive sampling. In both cases, plant VOCs are typically transferred onto a chemically stable sorbent trap in the field, and then later returned to a laboratory for chemical analysis. Newer miniature analytical chemistry methods are allowing this to be done directly in the field, and they are still under development.

In active sampling, air is forced over the plant to an instrument sampling manifold or sorbent trap by a pump. Passive sampling relies on the diffusive motion of the VOC molecules as they come into contact with a sorbent sampling surface and are then trapped by their chemical interaction with that sorbent material. Further differences among the methodologies include the specific design of the enclosure to route the VOCs from the plant to a trap or chemical analysis tool, and types of sorbent phase that are employed. The choice of sorbent is often dictated by the chemical properties of the analyte of interest, such as polarity or molecular size. Detailed descriptions of sorbent options and information for guidance of specific sorbent selection have been previously published (49, 50).

With either active or passive sampling, an enclosure is typically required to support the sorbent and to isolate the VOCs originating from the plant or specific part of the plant, such as roots or leaves, from the surrounding environment. The enclosure is normally made out of an inert material to minimize the impact of the enclosure material on the sampled VOC composition. Therefore, materials with very low VOC off-gassing such as glass or Teflon are used to avoid sample contamination. These materials also offer minimal absorptivity which reduces the changes to the local environment. Other properties of the selected materials can also be important as they affect the sampling environment in other ways. For example, Stewart-Jones and Poppy found that during dynamic sampling in direct sunlight with glass and polyester enclosures led to an increase in the internal temperature of the enclosure by an average of 7.5 °C and 5.2 °C, respectively (51). Such temperature increases will affect plant metabolism and alter the released

VOCs. Furthermore, both of these materials were found to release and adsorb some of the VOCs produced by plants (51). Commercially available enclosure devices have been successfully used to isolate parts of the plant while maintaining control over temperature and photosynthetically active radiation (PAR) during active sampling (52).

In addition to the pre-concentration of VOCs onto sorbents, they can be directly collected by pumping the air off the plant into a canister or bag. The sorbent or VOC-containing vessel can then be stored, and VOC analysis may be performed at a later time given the proper storage conditions of low humidity and low temperature. Pre-concentration onto a sorbent is not always necessary, as certain analytical instruments can either receive direct injections of a VOC air sample or a steady stream of sample. However, it is often necessary to increase VOC concentration levels or to interface the analyte source with an instrument operating in a pulse mode. In these cases, the preconcentration/sorption step is essential, and there are numerous types of sorbent available (53).

Certain sorbents are commonly used in almost all VOC analysis. These sorbents are capable of collecting a broad range of organic molecules in order to achieve the most comprehensive collection of the volatiles emitted by plants, as opposed to sorbents designed to retain certain specific analytes and discriminate against many other compounds that may comprise certain chemical backgrounds. Since the majority of the VOCs emitted by plants, such as terpenes/terpenoids, esters, long-chain aldehydes and hydrocarbons are relatively non-polar, the sorbents that most efficiently collect non-polar compounds are often used. For example, PDMS polymer sorbent was used to monitor the emissions from in vitro mechanically and herbivore-wounded plants as well as diurnal differences in volatiles emissions (54).

Tenax® TA is a commonly employed sorbent for biological VOC sampling. This material is a porous polymer which is primarily used for active sampling. Tenax® TA is often pre-weighed and placed in a silicone-coated thermal desorption tube, where the sorbent held in place with glass wool. Additional sorbent phases, such as Carboxen® D, may be used in order to enable collection of a wider range of VOCs. Volatiles collected on Tenax® TA are then released with thermal desorption. Tenax® TA has been successfully applied in the characterization the VOC response of birch trees to larvae and the resulting bird predation (55), VOC plant response to water deficit stress (56), and VOC emission from trees with varying resistances to aphids (57).

In addition to using sorbent as pre-formed pellets of powder, it can also be coated or formed onto another surface. A very popular version of this approach is the solid phase microextraction (SPME) method, which was first developed in the mid-1990s (58, 59). SPME sampling is primarily used with static headspace sampling, although in some cases it was also used in dynamic sampling devices (60). The SPME unit consists of a fiber tip coated in a sorbent that can be retracted into a narrow case to protect the tip and minimize adsorption of extraneous chemicals during transportation and storage. Different sorbent phases and thicknesses can be applied to enable better selectivity towards different classes of analytes such as nonpolar and polar compounds. SPME have been extensively used in biological research. For example, SPME were used to study

the differences in volatiles emitted from flowers and flower organs from different sexes and cultivars of carob trees (61), and the differences in VOC emissions from certain species of trees in the Sierra Nevada mountains (62).

Another example of a sorbent-coated device is the stir bar sorptive extraction (SBSE) sampler. This device is comprised of a magnetic stir bar coated in PDMS (63). One key advantage of SBSE over SPME is a greater amount of sorbent coating, leading to an increased capacity to collect VOCs. The stir bar has approximately 24–100 μL sorptive capacity compared to SPME, which only has 0.5 μL capacity (64). While SBSE was primarily designed for aqueous solutions extraction, it has also been used for static VOC sampling with thermal desorption. SBSE has been used to study VOC emissions from various biological surfaces (65) as well as plant volatiles specifically: plant VOCs induced and released during wounding by herbivorous arthropods (66), volatile triggers in plant defense response (67), and the relationships between wine aroma and grape development (68).

An important consideration in VOC sampling studies is the abundances of sampled chemicals. In certain cases, when target or matrix compounds are present in large quantities, it is possible to overload the sampling device which will result in the deterioration of data quality. Such chemical overloading may present a problem for both *in situ* VOC measurements and VOC preconcentration on sorbents with laboratory-based analysis. Possible sources of high abundance volatiles are certain plants (e.g. citrus) that produce volatiles in relatively high amounts, plant parts such as blossoming flowers, or extraneous environmental sources. In addition, physical damage to the plant may lead to extensive production of chemicals such as “green leaf volatiles”, e.g. *cis*-3-hexenal (leaf aldehyde) and methyl salicylate. The contamination and high chemical background will lead to difficulties in data analysis and interpretation. Optimization of sampling time and selection of appropriate sorbents that allows discriminating against the prevalent background compound(s) are, therefore, necessary. In addition, thorough sorbent conditioning and baking of surfaces that were exposed to the sample are essential in order to diminish “memory” effects and reduce the consequences of overexposure.

Analytical Instrumentation for Plant VOC Detection and Biomarker Chemical Characterization

A number of analytical techniques can be used to analyze global VOC fingerprints from various plant systems (69). Gas chromatography mass spectrometry (GC/MS) is the most common analytical technique used to characterize plant VOCs (70). It is the method of choice for such studies due to the high amount of chemical information it provides. With further advances such as GCxGC (71), even the most complex VOC mixtures can be separated and individual chemicals can be identified. Certain types of MS instrumentation are particularly suitable for plant VOC analysis. For example, a special ion source based on proton transfer process from hydronium ions can be directly connected to an analyzing system such as quadrupole mass analyzer. This technique, called

Proton-Transfer-Reaction Mass Spectrometry (PTR-MS) (72) can be used for real-time monitoring of chemical composition of a (usually gaseous) sample due to a very short response time (~ 100 msec) of the instrument. Indeed, PTR-MS is most commonly used in VOC monitoring in ambient air in environmental and biological research (73). Another example of a MS technique that is particularly suitable for VOC analysis is Selected Ion Flow Tube Mass Spectrometry (SIFT-MS) (74). This technique is based on the chemical ionization of trace volatile compounds by selected positive precursor ions such as H_3O^+ , NO^+ and O_2^+ in a flow tube connected to a mass spectrometer. SIFT-MS can be used for quantification of trace compounds in the gas phase and it has shown a great potential for use across a wide variety of fields, particularly in the life sciences, such as biology and medicine (75). Other techniques such as nuclear magnetic resonance (NMR) and liquid chromatography mass spectrometry (LC/MS) are commonly used to analyze the semi-volatile and higher molecular weight plant metabolites (76–80).

Both MS- and NMR-based methods are capable of providing the most comprehensive chemical information, but possess several drawbacks. First, all of these methods require time-consuming sample preparation and handling. Even though currently there are some available options for portable in field mass spectrometers (81), the MS analysis is almost exclusively done in laboratory conditions. Consequently, the analysis is very expensive, and requires complex and bulky equipment that needs to be operated by highly skilled personnel. This, in turn, makes the MS and NMR methods impractical for routine applications suitable for users outside communities of researchers in universities or designated facilities.

On the opposite side of the analytical information and metabolite content that a method measures, there is a family of lower-resolution VOC detection methods named the electronic nose or “E-nose” (82, 83). E-nose devices are comprised of an array of partially-selective chemical elements that are sensitive to various types of chemical compounds (similarly to odor receptors in biological olfactory systems) (84–88). E-nose devices are not capable of identifying and quantifying specific compounds, but instead provide the overall profile of a VOC mixture as a “fingerprint”. If the device is “tuned” to perform a specific discrimination task, it can distinguish among varying VOC signatures from different samples (89–92). Successful applications of E-nose to plant status monitoring were reported (93, 94). Compared to MS and other bench-top methods, the E-nose technology is rapid, inexpensive and easy-to-use as it does not require pre-concentration and sample preparation. Unfortunately, it has performance issues such as poor reproducibility, signal drift and difficulty in correlating E-nose responses with other analytical sensors outputs. This has limited their wide-spread use at the present time.

A family of ion mobility-based methods is positioned somewhere in between mass spectrometry and the E-nose methods, in terms of ease-of-use and information provided by the sensor output. An example of such technology is high Field Asymmetric waveform Ion Mobility Spectrometry (FAIMS), which is also called Differential Mobility Spectrometry (DMS). DMS is a gas phase separation and detection technique. It operates by exploiting the non-linear behavior of charged ions at rapidly alternating high and low electric fields to induce separation

and subsequent detection of chemical species (95). DMS operates at ambient pressure with routine sensitivity at the parts-per-billion (ppb) level. It has low power consumption, a small size and has the potential for further miniaturization (95, 96). At the same time, specific analytes can be identified by DMS if standards are run prior to field experiments, and quantitative measurements of chemical abundances are also feasible, sometimes down to even the parts-per-trillion (ppt) levels. Furthermore, hyphenation with gas chromatography (GC) provides additional chromatographic separation and further enhances the method's analytical capabilities. For example, GC/DMS has been applied to characterize and distinguish volatile compounds emitted from peel sections of normal healthy citrus fruit and those with citrus "puff" disorder (97) and trees affected by citrus greening (98). DMS is a relatively new analytical technique, but it already has been extensively applied to characterization of VOCs in complex systems, such as discrimination of proliferating bacterial samples (99).

Data Analysis and Signal Processing Approaches

Data analysis is one of the most important steps in determining chemical VOC biomarkers of significance in plant and biological systems. As discussed previously, the VOC mixtures emitted from plants are extremely complex and vary greatly in composition and abundance (100). Therefore, one of the main challenges is the high dimensionality of data sets (82, 101–105). It is also important to note approaches that may be used to numerically analyze the data from different types of analytical instrumentation in order to attribute biomarker chemical identities to specific biological processes, such as a disease or disorder.

Gutierrez-Osuna (82) describes the steps to be considered for the pattern analysis which could potentially be extended to many VOC analysis data types. The process consists in four main steps: signal preprocessing, normalization and dimensionality reduction, analysis, and validation.

Data analysis is divided in two main categories: supervised and unsupervised methods. Unsupervised methods refer to finding a hidden structure in unlabeled data while supervised methods relate to finding a mathematical transformation which correlates the data with their respective classes or categories. The supervised methods, in turn, are split into two groups: (i) explanatory, where the mathematical transformation performed by the algorithm is transparent; or (ii) inductive, where the aim is to find which metabolites are most important in separating different classes.

Signal Preprocessing

A preprocessing step is the application of different compression algorithms to obtain information regarding the change in the signal. As discussed previously, GC/MS and the E-nose are two of the most commonly employed analytical

chemistry techniques used to analyze plant volatiles, both in the laboratory and in the field. Each of these tools also produces different types of data outputs that are unique to the instrument used. Since signal preprocessing is specific to a given type of data structure, variations of traditional mathematical analysis are applied to the data from these instruments.

The signal from an E-nose sensor, either custom-built or commercially available, is often a time-varying response. The signal is comprised of a transient and steady state signal where the waveform changes according to presence of measured VOCs and their concentrations. The main signal preprocessing issues that need to be addressed are a baseline correction and compensation for sensor drift. Three types of baseline correction can be applied to E-nose signals: (i) difference correction, which is used to eliminate additive noise by subtracting the baseline; (ii) relative correction, which normalizes the signal by dividing by the baseline with the objective of removing multiplicative drift; and (iii) fractional correction, which is a combination between difference and relative corrections where a dimensionless and normalized response is obtained. There are some algorithms dedicated to extracting the information from the steady state response. However, in the last decade, the focus has shifted to evaluation of the transient response (*106–108*).

Preprocessing of GC/MS data is a critical procedure required to remove non-relevant information and to resolve overlapping peaks. In general, there are several main steps involved in GC/MS preprocessing (*109*): noise removal, peak deconvolution and identification, and chromatogram alignment. For GC/MS noise removal, different approaches have been developed, such as: median filtering (*110*); filtering based on a polynomial regression (*111*); wavelet transformation (*112, 113*); and the most popular, moving window filters (*114*).

After noise removal, a GC/MS signal is deconvolved to discern two or more coeluting (overlapping) components in the GC/MS data. The most popular software used for GC/MS deconvolution is called Automated Mass Spectral Deconvolution and Identification system (AMDIS) (*115*). AMDIS extracts qualitative and quantitative information such as peak height, an estimation of the total abundance of each deconvoluted compound, and identification of each compound using mass spectroscopic libraries such as NIST (National Institute of Standards and Technology Gaithersburg, MD, USA).

Normalization and Dimensionally Reduction

Normalization is the next step of VOC data analysis that should be applied after preprocessing and independently of the analytical technique used for data collection. Normalization procedures compensate for sample-to-sample variation caused by analyte concentration differences and/or ensure that the sensors output signal magnitudes are comparable within a data set. Normalization can frequently be carried out by correcting the signal using a single or multiple internal standards spiked into the sample (*116*), but this is not always feasible when analyzing ambient gas phase VOCs in field conditions.

In addition to normalizing the data, frequently dimensionality reduction techniques are applied (117) to remove irrelevant information (noise) thus enhancing the classification/regression performance. This also provides easier data interpretation, while improving computational and data storage costs. However, dimensionality reduction may lead to loss important features/information. There are especially salient concerns for VOC analysis, because we ultimately desire to know the identity of the biomarkers from our studies. To deal with this, there are two approaches that can be applied to VOC metabolite data: (i) feature extraction, and (ii) feature selection.

The aim of feature extraction is to extract important biological signal features via mathematical transformations of the original data vector. Principal Component Analysis (PCA) (118) is the most widely used technique for this process. It is an unsupervised method that is mostly used as a representation technique, and it projects the original data into a subspace along the directions of maximum variance. As this technique does not use prior information regarding data classes to generate a model, a supervised method needs to be used to validate discrimination of the data classes (e.g. health status). PCA is optimal under unimodal Gaussian assumptions or normally distributed systems. The new projected dimensions are called scores and loadings. The scores are a projection of each sample over the new subspace and the loadings represent the variables projected over the new subspace. Therefore, if any discrimination is found in the scores, the loadings will represent which variables in the data are responsible.

Linear Discriminant Analysis (LDA) or Fisher Discriminant Analysis (118) is also a supervised method which maximizes class separation in one dimension. LDA assumes unimodal Gaussian classes with different mean and class-covariance. However, LDA as dimension reduction technique requires that the number of samples must be at least 3 times greater than the number of features to avoid overfitting. Furthermore, LDA may fail if discriminating information is in the variance of the data but not in the mean. There are some variations of LDA such as non-parametric LDA (119), Orthonormal LDA (120), and Generalized LDA (121). Sammon's maps (122, 123), multilayer perceptrons (124), Kohonen self-organizing maps (125), Kernel PCA (126, 127), projection pursuit (128) and Independent Component Analysis (129) are used when the data are not normally distributed. Multivariate orthogonal projection methods (OP) aim to reduce the original data to a new subspace in which the new dimensions represent maximum variance related or unrelated to the information of interest. There are different techniques used to achieve this objective, and the final purpose of these algorithms is to improve the data and minimize the influence of non-desired variation in the data.

The main objective of *feature selection* (FS) is to find "the best" subset of VOC signal features maximizing prediction accuracy for a given plant data set. FSS evaluates the whole space of possibilities combining different features, either individually or by feature sets to determine the optimal feature combination. The simplest algorithm is an exhaustive search where all possibilities are evaluated, but this approach is time consuming and carries a high computational cost. Exponential, sequential and randomized search strategies are designed to explore the feature space more efficiently.

Data Interpretation and Analysis Methods

Once the complex VOC data sets from the plant systems have been simplified and normalized, various approaches can be applied to interpret the meaning of the data. Typically, these rely on having enough raw data from the biological experiments to divide it into two categories: a training data set to build models of the data, and a testing data set to understand accuracy of the models that we generate of the biological response to disease or disorder.

Classification Methods

Quadratic Classifiers (QC) (82, 118) are a simple way to separate measurements of two or more classes. K nearest neighbor classifiers (k-NNs) (130–132) are also powerful techniques for classifying VOC data sets based on the k closest examples in a training dataset. Even though k-NN generates highly nonlinear classifications with limited data, it is very sensitive to non-scaled data. Therefore, the selection of the k neighbors is crucial to avoid local boundaries and misinterpretation of the data. Multilayer Perceptron Classifiers (MLPs) (133–135) are the most popular example of artificial neural network (ANN). The Support Vector Machine (SVM) technique (136) has also been successfully used in different metabolomics fields. Although SVM is mainly used to solve classifying problems, recently their applications were extended to solving regression problems and density estimations. Partial Least Squares Discriminant Analysis (PLSDA) (137, 138) is especially powerful when dealing with high-dimensional data which are frequently obtained in VOC metabolite monitoring.

Regression Methods

Ordinary Least Squares (OLS) (139) is the simplest method to perform regression. OLS assumes that the dependent variable Y (such as the biomarker VOC analyte concentration) can be predicted from a linear combination of the independent variables X (the sensor response for a biological condition). It assumes that X and Y are mean-centered and the covariance matrix of Y is not singular. Partial Least Squares (140, 141) is the most used method in calibration problems due to its ability to handle collinear data and reduce the number of calibration observations. PLS maximizes the correlation between X and Y in sequential steps. PLS attempts to find new factors called latent variables which are linear combinations of X . The first latent variable is a projection of X along the eigenvector Y . Then the subsequent latent variables are defined by regression between X and the current PLS latent variable. This procedure is repeated until a criterion, such as a pre-set number of latent variables, is met.

Three main steps are involved in clustering algorithms (142, 143): (i) definition of a pattern proximity or distance metric measure between observations, typically Euclidean distance or Mahalanobis distance, (ii) optimization of clustering or grouping criterion, and (iii) data abstraction or/and data assessment.

Hierarchical Clustering methods (144–147) use a tree structure called a dendrogram which represents the biological states produced by hierarchical clustering. C-Means (148–150) is another one of the simplest methods to solve clustering problems. This algorithm attempts to classify a data set into C disjointed clusters (number C is defined prior). Self-Organizing Maps (SOM) (151, 152) are the most popular neural network models. SOM performs data clustering without knowing the class membership of the input data

Validation of Data Analysis

The validation procedure is one of the most important steps in the data analysis process used to attribute VOC biomarkers to specific biological states. The validation methodologies assess how well a statistical analysis of metabolite data (classification/prediction) will generalize to an independent data set. The simplest approach is frequently called the “hold out” strategy. The original raw data is separated into a validation set with 30% of the original raw data and an initial training set with the rest of the data from a study. This break down works well when the number of replicate biological samples is sufficiently greater than the number of variables in the biological question being studied.

One of the potential problems with this approach is the possibility of a non-uniform distribution of the data that can lead to misleading results. Certain variations of the “hold out” method are designed to overcome such problems. In the K-fold cross-validation method, the data are split into K partitions, with each partition used for both training and validation set. Another method is random subsampling cross-validation. This method is similar to K-fold cross-validation, but in this case each split is chosen randomly. When the number of folds is equal to number of samples, the method is called “leave-one-out.” This method is also valid when the number of samples is small. In these three cases, the performance of the model is calculated by the average over the K data partitions. The K value could be critical: when K is large, the variance across partitions will be large and the bias of the estimate will be small making the final prediction overly optimistic. On the other hand, if K is small, the bias will be large and the variance of the estimate will be small, therefore the final relationship we infer may be underestimated.

Concluding Remarks and Future Trends

In the present chapter, we have reviewed approaches to monitor and measure plant health via emitted volatile organic compounds (VOCs). Compared with traditional methods and visual inspection methods, these techniques

have great promise, although there is still much work to be done. Major advances in our understanding of plant physiology coupled with developments in analytical instrumentation and sampling techniques have continuously pushed the boundaries of these VOC-based diagnostic possibilities. While many examples of successful applications of plant VOC analysis in biological, ecological studies and agricultural applications are given, the full potential of VOC monitoring in plants has not yet been realized.

Ultimately, we envision VOC-based agriculture sensing that can be performed on plants in the field to interpret a chemical “signature” that reflects the biological/physiological status of the entire organism. This could include information on the presence of a pathogenic infection, changes in nutrition status, or any other relevant information about plant’s health status. In turn, this information can be used to guide decisions for actions such as plant removal, pesticide spray levels, watering, or administering fertilizer and nutrients. Apart from being reliable, fast, reasonably inexpensive and environmentally safe, such VOC monitoring also needs to be amenable for monitoring of large acreages. The output of VOC sensors needs to be easily interpretable by non-scientists which would allow growers to take advantage of technology directly, and not rely on scientific personnel.

At the present time, documenting the baseline variability of VOCs that emanate from plants is still a challenge. It is also important to understand how much diversity exists between commonly cultivated varieties of the same plant. Also, we still need to investigate how factors such as weather, soil quality, and geographic location will affect the VOC output. A part of the challenge in establishing such a baseline is a lack of standardization of sampling procedures for VOCs in the field. Concerted research efforts and standardization of sampling protocols by the community of dedicated plant VOC researcher may ultimately lead to further breakthroughs in our understanding of the plants’ “chemical language” of VOC production and signaling.

Also, great advances have been made in analytical instrumentation for complex VOC analysis. However, there is still a lack of instrumentation immediately suitable to assess and monitor the plant VOCs directly in field conditions. Current in field sampling and analysis approaches are still somewhat rudimentary and often have many manual steps. Consequently, laboratory-based analytical chemistry equipment is often heavily relied upon. A miniature, portable, easy-to-use VOC detection system that is highly sensitive and capable of quantifying the amount of specific VOCs that are released from plants is not yet widely available. At the same time, the application of advanced data analysis in conjunction with the constant stream of improvements in experimental design and more robust equipment will eventually lead to VOC analysis taking its rightful place as one of the most important tools at the disposal of growers around the world. Finally, understanding of VOC production in the larger framework of biochemical pathways will further advance the field. A host of applications beyond precision agriculture may also be possible. These applications may include, but are certainly not limited to: pest management, post-harvest monitoring, biodiversity and ecology studies. For example, an opportunity to find and study new and interesting endophytes in situ would be

fascinating from a basic science perspective (153). The full potential of plant VOC monitoring and the realm of future applications are still yet to be realized.

References

1. Kessler, A.; Baldwin, I. T. *Science* **2001**, *291*, 2141–2144.
2. Baldwin, I. T.; Halitschke, R.; Paschold, A.; von Dahl, C. C.; Preston, C. A. *Science* **2006**, *311*, 812–815.
3. Bezemer, T. M.; van Dam, N. M. *Trends Ecol. Evol.* **2005**, *20*, 617–624.
4. Rohloff, J.; Bones, A. M. *Phytochemistry* **2005**, *66*, 1941–1955.
5. Glendinning, J. I.; Foley, C.; Loncar, I.; Rai, M. J. *Comp. Physiol. A.* **2009**, *195*, 591–601.
6. Runyon, J. B.; Mescher, M. C.; De Moraes, C. M. *Science* **2006**, *313*, 1964–1967.
7. Dudareva, N.; Pichersky, E.; Gershenzon, J. *Plant Physiol.* **2004**, *135*, 1893–1902.
8. Sumner, L. W.; Mendes, P.; Dixon, R. A. *Phytochemistry* **2003**, *62*, 817–836.
9. Weckwerth, W. *Physiol. Plant.* **2008**, *132*, 176–189.
10. Muller, K.; Pelzing, M.; Gnauk, T.; Kappe, A.; Teichmann, U.; Spindler, G.; Haferkorn, S.; Jahn, Y.; Herrmann, H. *Chemosphere* **2002**, *49*, 1247–1256.
11. Herrmann, U.; Jonischkeit, T.; Bargon, J.; Hahn, U.; Li, Q. Y.; Schalley, C. A.; Vogel, E.; Vogtle, F. *Anal. Bioanal. Chem.* **2002**, *372*, 611–614.
12. Rapparini, F.; Baraldi, R.; Facini, O. *Phytochemistry* **2001**, *57*, 681–687.
13. Karl, T.; Guenther, A.; Turnipseed, A.; Patton, E. G.; Jardine, K. *Biogeosciences* **2008**, *5*, 1287–1294.
14. Loreto, F.; Barta, C.; Brilli, F.; Nogues, I. *Plant Cell Environ.* **2006**, *29*, 1820–1828.
15. Tholl, D.; Boland, W.; Hansel, A.; Loreto, F.; Rose, U. S. R.; Schnitzler, J. P. *Plant J.* **2006**, *45*, 540–560.
16. Futch, S.; Weingarten, S.; Irey, M. *Proceedings of the Florida State Horticultural Society*; 2009; Vol. 122; pp 152–157.
17. Ehsani, R.; Buchanan, S.; Salyani, M. Agricultural and Biological Engineering Department, Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, University of Florida, 2009.
18. Lopez, M.; Bertolini, E.; Olmos, A.; Caruso, P.; Gorris, M.; Llop, P.; Penyalver, R.; Cambra, M. *Int. Microbiol.* **2003**, *6*, 233–243.
19. Bravo, C.; Moshou, D.; Oberti, R.; West, J.; McCartney, A.; Bodria, L.; Ramon, H. *Agric. Eng. Int.: The CIGR J. Sci. Res. Dev.* **J2004**, *6*.
20. Moshou, D.; Bravo, C.; Oberti, R.; West, J.; Bodria, L.; McCartney, A.; Ramon, H. *Real-Time Imaging* **2005**, *11*, 75–83.
21. Chaerle, L.; Lenk, S.; Hagenbeek, D.; Buschmann, C.; Van der Straeten, D. *J. Plant. Physiol.* **2007**, *164*, 253–262.
22. Wetterich, C.; Kumar, R.; Sankaran, S.; Belasque, J.; Ehsani, R.; Marcassa, L. *J. Spectrosc.* **2013**.
23. Garcia-Ruiz, F.; Sankaran, S.; Maja, J. M.; Lee, W. S.; Rasmussen, J.; Ehsani, R. *Comput. Electron. Agr.* **2013**, *91*, 106–115.

24. Moshou, D.; Bravo, C.; West, J.; Wahlen, T.; McCartney, A.; Ramon, H. *Comput. Electron. Agric.* **2004**, *44*, 173–188.
25. Shafri, H. Z. M.; Hamdan, N. *Am. J. Appl. Sci.* **2009**, *6*, 1031–1035.
26. Qin, J.; Burks, T.; Ritenour, M.; Bonn, W. *J. Food Eng.* **2009**, *93*, 183–191.
27. Spinelli, F.; Noferini, M.; Costa, G. *Acta Hort.* **2006**, *704*, 87–90.
28. Purcell, D.; O'Shea, M.; Johnson, R.; Kokot, S. *Appl. Spectrosc.* **2009**, *63*, 450–457.
29. Sankaran, S.; Ehsani, R. T. *Trans. Asabe* **2012**, *55*, 313–320.
30. Marcassa, L.; Gasparoto, M.; Belasque, J.; Lins, E.; Nunes, F.; Bagnato, V. *Laser Phys.* **2006**, *16*, 884–888.
31. Belasque, J.; Gasparoto, M.; Marcassa, L. *Appl. Optics* **2008**, *47*, 1922–1926.
32. Lins, E.; Belasque, J.; Marcassa, L. *Precis. Agric.* **2009**, *10*, 319–330.
33. Sankaran, S.; Mishra, A.; Maja, J.; Ehsani, R. *Comput. Electron Agric.* **2011**, *77*, 127–134.
34. Yang, C.; Cheng, C.; Chen, R. *Crop Sci.* **2007**, *47*, 329–335.
35. Delalieux, S.; van Aardt, J.; Keulemans, W.; Schrevens, E.; Coppin, P. *Eur. J. Agron.* **2007**, *27*, 130–143.
36. Chen, B.; Wang, K.; Li, S.; Wang, J.; Bai, J.; Xiao, C.; Lai, J. In *IFIP International Federation for Information Processing, Computer and Computing Technologies in Agriculture*; Springer: Boston, 2008; Vol. 259; pp 1169–1180.
37. Choi, Y.; Tapias, E.; Kim, H.; Lefeber, A.; Erkelens, C.; Verhoeven, J.; Brzin, J.; Zel, J.; Verpoorte, R. *Plant Physiol.* **2004**, *135*, 2398–2410.
38. Watts, A.; Ambrosia, V.; Hinkley, E. *Remote Sens.* **2012**, *4*, 1671–1692.
39. Zhang, C.; Kovacs, J. *Precis. Agric.* **2012**, *13*, 693–712.
40. Primicerio, J.; Di Gennaro, S.; Fiorillo, E.; Genesio, L.; Lugato, E.; Matese, A.; Vaccari, F. *Precis. Agric.* **2012**, *13*, 517–523.
41. Sankaran, S.; Mishra, A.; Ehsani, R.; Davis, C. *Comput. Electron Agric.* **2010**, *72*, 1–13.
42. Caudwell, A.; Kuszala, C. *Res. Microbiol.* **1992**, *143*, 791–806.
43. Bianco, P. A.; Casati, P.; Belli, G. *Dev. Plant Pathol.* **1997**, *11*, 179–182.
44. Prince, J. P.; Davis, R. E.; Wolf, T. K.; Lee, I. M.; Mogen, B. D.; Dally, E. L.; Bertaccini, A.; Credi, R.; Barba, M. *Phytopathology* **1993**, *83*, 1130–1137.
45. Teixeira, D. d. C.; Danet, J. L.; Eveillard, S.; Martins, E. C.; Cintra, d. J. J. W.; Yamamoto, P. T.; Lopes, S. A.; Bassanezi, R. B.; Ayres, A. J.; Saillard, C.; Bove, J. M. *Mol. Cell. Probes* **2005**, *19*, 173–179.
46. Dandekar, A. M.; Martinelli, F.; Davis, C. E.; Bhushan, A.; Zhao, W.; Fiehn, O.; Skogerson, K.; Wohlgemuth, G.; D'Souza, R.; Roy, S.; Reagan, R. L.; Lin, D.; Cary, R. B.; Pardington, P.; Gupta, G. *Crit. Rev. Immunol.* **2010**, *30*, 277–289.
47. Jagoueix, S.; Bove, J. M.; Garnier, M. *Mol. Cell. Probes* **1996**, *10*, 43–50.
48. do, C. T. D.; Luc, D. J.; Eveillard, S.; Cristina, M. E.; de, J. J. W. C.; Takao, Y. P.; Aparecido, L. S.; Beozzo, B. R.; Juliano, A. A.; Saillard, C.; Bove, J. M. *Mol. Cell. Probes* **2005**, *19*, 173–179.
49. Woolfenden, E. J. *Chromatogr. A* **2010**, *1217*, 2674–2684.
50. Tholl, D.; Wihelm, B.; Hansel, A.; Loreto, F. R.; Ursula, S. R.; Schnitzler, J.-P. *Plant J.* **2006**, *45*, 540–560.

51. Stewart-Jones, A.; Poppy, G. M. *J. Chem. Ecol.* **2006**, *32*, 845–864.
52. Laothawornkitkul, J.; Moore, J. P.; Taylor, J. E.; Possell, M.; Gibson, T. D.; Hewitt, C.; Nicholas, P.; Nigel, D. *Environ. Sci. Technol.* **2008**, *42*, 8433–8439.
53. Woolfenden, E. *J. Chromatogr., A* **2010**, *1217*, 2685–2694.
54. Vercammen, J.; Pham-Tuan, H.; Sandra, P. *J. Chromatogr., A* **2001**, *930*, 39–51.
55. Mantyla, E.; Alessio, G.A.; Blande, J.D.; Heijari, J.; Holopainen, J.K.; Laaksonen, T.; Piirtola, P.; Klemola, T. *PLoS ONE* **2008**, *3*, e2832.
56. Ormeno, E.; Mevy, J.P.; Vila, B.; Bousquet-Melou, A.; Greff, S.; Fernandez, B.C. *Chemosphere* **2007**, *67*, 276–284.
57. Staudt, M.; Jackson, B.; El-Aouni, H.; Buatois, B.; Lacroze, J.P.; Poessel, J.L.; Sauge, M.H. *Tree Physiol.* **2010**, *30*, 1320–1334.
58. Pawliszyn, J. *Solid Phase Microextraction – Theory and Practice*; Wiley-VCH: New York, 1997.
59. Zhang, Z.; Yang, M.J.; Pawliszyn, J. *Anal. Chem.* **1994**, *66*, 844–553.
60. Fabio Augusto, J. K.; Pawliszyn, J. *Anal. Chem.* **2001**, *73*, 481–486.
61. Custodia, L.; Serra, H.; Nogueira, J. M. F.; Goncalves, S.; Romano, A. *J. Chem. Ecol.* **2006**, *32*, 929–942.
62. Bouvier-Brown, N. C.; Holzinger, R.; Palitzsch, K.; Goldstein, A. H. *Atmos. Environ.* **2009**, *43*, 389–401.
63. Baltussen, E.; Sandra, P.; David, F.; Cramers, C. *J. Microcolumn Sep.* **1999**, *11*, 737–747.
64. Soini, H. A.; Bruce, K.E.; Wiesler, D.; David, F.; Sandra, P.; Novotny, M. *J. Chem. Ecol.* **2005**, *31*, 377–392.
65. Soini, H. A.; Bruce, K. E.; Klouckova, I.; Brereton, R. G.; Penn, D. J.; Novotny, M. V. *Anal. Chem.* **2006**, *78*, 7161–7168.
66. Shiojiri, K.; Ozawa, R.; Kugimiya, S.; Uefune, M.; van Wijk, M.; Sabelis, M. W.; Takabayashi, J. *PLoS ONE* **2010**, *5*, e12161.
67. Shiojiri, K.; Ozawa, R.; Matsui, K.; Sabelis, M.W.; Takabayashi, J. *Nat. Sci. Rep.* **2012**, *2*, 1–5.
68. *Flavor Chemistry of Wine and Other Alcoholic Beverages*; Qian, M. Shellhammer, T., Eds.; ACS Symposium Series 1104; American Chemical Society: Washington, DC, 2012; pp i–v.
69. Goff, S. A.; Klee, H. J. *Science* **2006**, *311*, 815–819.
70. Lytovchenko, A.; Beleggia, R.; Schauer, N.; Isaacson, T.; Leuendorf, J.; Hellmann, H.; Rose, J.; Fernie, A. *Plant Methods* **2009**, *5*, 4.
71. Stoll, D. R.; Wang, X. L.; Carr, P. W. *Anal. Chem.* **2008**, *80*, 268–278.
72. Hansel, A.; Jordan, A.; Holzinger, R.; Prazeller, P.; Vogel, W.; Lindinger, W. *Int. J. Mass Spectrom. Ion Processes* **1995**, *149/150*, 609–619.
73. Biasioli, F.; Yeretzian, C.; Gasperi, F.; Mark, T. D. *TrAC, Trends Anal. Chem.* **2011**, *30*, 968–977.
74. Spanel, P.; Rolfe, P.; Rajan, B.; Smith, D. *Ann. Occup. Hyg.* **1996**, *40*, 615–626.
75. Spanel, P.; Smith, D. *Eur. J. Mass Spectrom.* **2007**, *13*, 77–82.
76. Wolfender, J.-L.; Ndjoko, K.; Hostettmann, K. *J. Chromatogr., A* **2003**, *1000*, 437–455.

77. Paolini, J.; Tomi, P.; Bernardini, A.-F.; Bradesi, P.; Casanova, J.; Kaloustian *J. Nat. Prod. Res.* **2008**, *22*, 1276–1284.
78. Dafoe, N. J.; Zamani, A.; Ekramoddoullah, A. K. M.; Lippert, D.; Bohlmann, J.; Constabel, C. P. *J. Proteome Res.* **2009**, *8*, 2341–2350.
79. Lee, J.-E.; Hong, Y.-S.; Lee, C.-H. *J. Agric. Food Chem.* **2009**, *57*, 4810–4817.
80. Prestes, R. A.; Colnago, L. A.; Forato, L. A.; Carrilho, E.; Bassanezi, R. B.; Wulff, N. A. *Mol. Plant Pathol.* **2009**, *10*, 51–57.
81. Ouyang, Z.; Cooks, R. G. *Annu. Rev. Anal. Chem.* **2009**, *2*, 187–214.
82. Gutierrez-Osuna, R. *IEEE Sens. J.* **2002**, *2*, 189–202.
83. Gardner, J. W.; Bartlett, P. N. *Sens. Actuators, B* **1994**, *18*, 211–220.
84. Fernandez, M. J.; Fontecha, J. L.; Sayago, I.; Aleixandre, M.; Lozano, J.; Gutierrez, J.; Gracia, I.; Cane, C.; Horrillo, M. D. *Sens. Actuators, B* **2007**, *127*, 277–283.
85. Patel, S. V.; Mlsna, T. E.; Fruhberger, B.; Klaassen, E.; Cemalovic, S.; Baselt, D. R. *Sens. Actuators, B* **2003**, *96*, 541–553.
86. Srivastava, A. K. *Sens. Actuators, B* **2003**, *96*, 24–37.
87. Taurino, A. M.; Distanto, C.; Siciliano, P.; Vasanelli, L. *Sens. Actuators, B* **2003**, *93*, 117–125.
88. Wolfrum, E. J.; Meglen, R. M.; Peterson, D.; Sluiter, J. *Sens. Actuators, B* **2006**, *115*, 322–329.
89. Gibson, T. D.; Prosser, O.; Hulbert, J. N.; Marshall, R. W.; Corcoran, P.; Lowery, P.; Ruck-Keene, E. A.; Heron, S. *Sens. Actuators, B* **1997**, *44*, 413–422.
90. Lebrun, M.; Plotto, A.; Goodner, K.; Ducamp, M. N.; Baldwin, E. *Postharvest Biol. Technol.* **2008**, *48*, 122–131.
91. Zhang, H. M.; Wang, J. *J. Stored Prod. Res.* **2007**, *43*, 489–495.
92. Bastos, A. C.; Magan, N. *Sens. Actuators, B* **2006**, *116*, 151–155.
93. Laothawornkitkul, J.; Moore, J. P.; Taylor, J. E.; Possell, M.; Gibson, T. D.; Hewitt, C. N.; Paul, N. D. *Envir. Sci. Technol.* **2008**, *42*, 8433–8439.
94. Baratto, C.; Faglia, G.; Pardo, M.; Vezzoli, M.; Boarino, L.; Maffei, M.; Bossi, S.; Sberveglieri, G. *Sens. Actuators, B* **2005**, *108*, 278–284.
95. Kolakowski, B. M.; Mester, Z. *Analyst* **2007**, *132*, 842–864.
96. Eiceman, G. A.; Karpas, Z. *Ion Mobility Spectrometry*, 2nd ed.; CRC Press: Boca Raton, FL, 2005.
97. Zhao, W.; Sankaran, S.; Ibanez, A. M.; Dandekar, A. M.; Davis, C. E. *Anal. Chim. Acta.* **2009**, *647*, 46–53.
98. Aksenov, A. A.; Cheung, W.; Zhao, W.; Bardaweel, H.; Martinelli, F.; Fiehn, O.; Dandekar, A. M.; Davis, C. E. *Citrograph* **2012**, *54*–56.
99. Shnayderman, M.; Mansfield, B.; Yip, P.; Clark, H. A.; Krebs, M. D.; Cohen, S. J.; Zeskind, J. E.; Ryan, E. T.; Dorkin, H. L.; Callahan, M. V.; Stair, T. O.; Gelfand, J. A.; Gill, C. J.; Hitt, B.; Davis, C. E. *Anal. Chem.* **2005**, *77*, 5930–5937.
100. Zhang, Z. M.; Li, G. K. *Microchem. J.* **2010**, *95*, 127–139.
101. Marco, S.; Gutierrez-Galvez, A. *IEEE Sens. J.* **2012**, *12*, 3189–3214.
102. Broadhurst, D. I.; Kell, D. B. *Metabolomics* **2006**, *2*, 171–196.
103. Trygg, J.; Holmes, E.; Lundstedt, T. *J. Proteome Res.* **2007**, *6*, 469–479.

104. Goodacre, R.; Vaidyanathan, S.; Dunn, W. B.; Harrigan, G. G.; Kell, D. B. *Trends Biotechnol.* **2004**, *22*, 245–252.
105. Hendriks, M. M. W. B.; van Eeuwijk, F. A.; Jellema, R. H.; Westerhuis, J. A.; Reijmers, T. H.; Hoefsloot, H. C. J.; Smilde, A. K. *Trend Anal. Chem.* **2011**, *30*, 1685–1698.
106. Gutierrez-Osuna, R.; Hierlemann, A. *Annu. Rev. Anal. Chem.* **2010**, *3*, 255–276.
107. Hierlemann, A.; Gutierrez-Osuna, R. *Chem. Rev.* **2008**, *108*, 563–613.
108. Scott, S. M.; James, D.; Ali, Z. *Microchim. Acta* **2006**, *156*, 183–207.
109. Katajamaa, M.; Oresic, M. *J. Chromatogr., A* **2007**, *1158*, 318–328.
110. Hastings, C. A.; Norton, S. M.; Roy, S. *Rapid Commun. Mass. Spectrom.* **2002**, *16*, 462–467.
111. Savitzky, A.; Golay, M. J. E. *Anal. Chem.* **1964**, *36*, 1627&.
112. Coombes, K. R.; Tsavachidis, S.; Morris, J. S.; Baggerly, K. A.; Hung, M. C.; Kuerer, H. M. *Proteomics* **2005**, *5*, 4107–4117.
113. Chen, S.; Billings, S. A.; Luo, W. *Int. J. Control* **1989**, *50*, 1873–1896.
114. Villas-Bôas, S. G.; Roessner, U.; Hansen, M. A. E.; Smedsgaard, J.; Nielsen, J. In *Metabolome analysis: An Introduction*; John Wiley & Sons, Inc: Hoboken, NJ, 2006; pp 146–187.
115. Stein, S. E. *J. Am. Soc. Mass. Spectrom.* **1999**, *10*, 770–781.
116. Sysi-Aho, M.; Katajamaa, M.; Yetukuri, L.; Oresic, M. *BMC Bioinf.* **2007**, *8*, 93.
117. Raudys, S. J.; Jain, A. K. *IEEE Trans. Pattern Anal.* **1991**, *13*, 252–264.
118. Webb, A. R. *Statistical Pattern Recognition*; John Wiley & Sons: New York, 2002.
119. Li, Z.; Lin, D.; Tang, X. *IEEE Trans. Pattern Anal.* **2009**, *31*, 755–761.
120. Okada, T.; Tomita, S. *Pattern Recogn.* **1985**, *18*, 139–144.
121. Ji, S.; Ye, J. *IEEE Trans. Neural Networks* **2008**, *19*, 1768–1782.
122. Sammon, J. W. *IEEE T. Comput.* **1969**, *C 18*, 401.
123. Byun, H. G.; Persaud, K. C.; Khaffaf, S. M.; Hobbs, P. J.; Misselbrook, T. H. *Comput. Electron Agric.* **1997**, *17*, 233–247.
124. Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. *Nature* **1986**, *323*, 533–536.
125. Melssen, W. J.; Smits, J. R. M.; Buydens, L. M. C.; Kateman, G. *Chemometr. Intell. Lab. Lab.* **1994**, *23*, 267–291.
126. Wu, W.; Massart, D. L.; deJong, S. *Chemom. Intell. Lab. Syst.* **1997**, *36*, 165–172.
127. Scholkopf, B.; Smola, A.; Muller, K. R. *Neural Comput.* **1998**, *10*, 1299–1319.
128. Huber, P. J. *Ann. Stat.* **1985**, *13*, 435–475.
129. Comon, P. *Signal Process.* **1994**, *36*, 287–314.
130. Henley, W. E.; Hand, D. J. *Statistician* **1996**, *45*, 77–95.
131. Bezdek, J. C.; Chuah, S. K.; Leep, D. *Fuzzy Sets Syst.* **1986**, *18*, 237–256.
132. Cover, T. M.; Hart, P. E. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27.
133. Pal, S. K.; Mitra, S. *IEEE Trans. Neural Networks* **1992**, *3*, 683–697.
134. Ruck, D. W.; Rogers, S. K.; Kabrisky, M.; Oxley, M. E.; Suter, B. W. *IEEE Trans. Neural Networks* **1990**, *1*, B863–B868.
135. Gardner, M. W.; Dorling, S. R. *Atmos. Environ.* **1998**, *32*, 2627–2636.

136. Burges, C. J. C. *Data Min. Knowl. Discovery* **1998**, 2, 121–167.
137. Westerhuis, J. A.; Hoefsloot, H. C. J.; Smit, S.; Vis, D. J.; Smilde, A. K.; van Velzen, E. J. J.; van Duijnhoven, J. P. M.; van Dorsten, F. A. *Metabolomics* **2008**, 4, 81–89.
138. Barker, M.; Rayens, W. *J. Chemom.* **2003**, 17, 166–173.
139. Duda, R. O.; Hart, P. E.; Stork, D. G. *Pattern Classification*, 2nd ed.; John Wiley & Sons, Inc.: New York, 2000.
140. Geladi, P.; Kowalski, B. R. *Anal. Chim. Acta.* **1986**, 185, 1–17.
141. Haaland, D. M.; Thomas, E. V. *Anal. Chem.* **1988**, 60, 1193–1202.
142. Dubes, R.; Jain, A. K. *Pattern Recognit.* **1976**, 8, 247–260.
143. Dubes, R.; Jain, A. K. *Pattern Recognit.* **1979**, 11, 235–254.
144. Johnson, S. C. *Psychometrika* **1967**, 32, 241–254.
145. Guess, M. J.; Wilson, S. B. J. *Clin. Neurophysiol.* **2002**, 19, 144–151.
146. Olson, C. F. *Parallel Comput.* **1995**, 21, 1313–1325.
147. Murtagh, F. *Comput. J.* **1983**, 26, 354–359.
148. Bobrowski, L.; Bezdek, J. C. *IEEE Trans. Syst. Man Cybern.* **1991**, 21, 545–554.
149. Yu, J. *IEEE Trans. Pattern Anal.* **2005**, 27, 1197–1211.
150. Hartigan, J. A.; Wong, M. A. *Appl. Stat.* **1979**, 28, 100–108.
151. Flexer, A. *Intell. Data Anal.* **2001**, 5, 373–384.
152. Kohonen, T. P. *Proc. IEEE* **1990**, 78, 1464–1480.
153. Strobel, G. A. *Crit. Rev. Biotechnol.* **2002**, 22, 315–333.