Thermochimica Acta 668 (2018) 132–141

Contents lists available at ScienceDirect

Thermochimica Acta

journal homepage: www.elsevier.com/locate/tca

Identification of volatiles from heated tobacco biomass using direct thermogravimetric analysis—Mass spectrometry and target factor analysis

Ashley Davies[a,*], James Thomas Grant Nicol[b], Chuan Liu[a], John Tetteh[c], Kevin McAdam[a]

[a] GR&D, British American Tobacco, Regents Park Road, Southampton SO15 8TL, United Kingdom
[b] JTN Consulting Limited, 272 Bath Street, Glasgow, Scotland, G2 4JR, United Kingdom
[c] DiKnow Ltd., 84 Rushdean Road, Rochester, Kent ME2 2QB, United Kingdom

ARTICLE INFO

Keywords:
Thermogravimetric analysis (TGA)
Mass spectrometry (MS)
Multivariate analysis
Heated biomass
Volatiles
Target factor analysis (TFA)

ABSTRACT

Target Factor Analysis (TFA) was used to identify full spectral profiles in the presence of complex overlapping signals in mass spectrometry data. This allowed interpretation of direct thermogravimetric analysis-mass spectrometry (TGA-MS) data from two varieties of tobacco biomass, heated below 350 °C. Previous TGA-MS approaches reported in biomass research use Selected Ion Monitoring (SIM) to monitor specific ions in the evolved gases. This cannot distinguish between isobaric mass fragments, and is challenged by the complex mixture of volatiles evolved from tobacco biomass. The TFA approach instead uses the complete reference mass spectra of the target compounds. Eighteen mass spectral references were used as target compounds to test their qualitative presence, based on correlations between the target spectra and the predicted spectra from the TFA process. Both simulated and experimental data sets were used to evaluate this approach.

Empirical statistical analysis recommended 7 or 8 principal components for our data, but investigating predictive capabilities beyond these levels revealed correlation trends and enhanced analytical insights. Based on this multivariate analytical strategy, which we call a “deep search”, it was found that there was a distinction between identifiable and nonidentifiable spectral groupings for the 18 target references chosen. The non-identifiable targets persistently scored below 0.55 correlation coeffecient (R), even when the search was based on two to three times the empirically recommended number of factors.

The results show that components present in a complex data matrix, generated from TGA-MS experiments on complex biomass materials could be identified by full spectral matching using TFA instead of SIM.

1. Introduction

Most chemical systems consist of mixtures of components, and the primary analytical goal is to identify the components and compare or quantify them. Many theoretical and technical publications on the subject of analytical chemistry have been published [1,2]. Most traditional analytical methods for volatile mixture analysis involve the use of Gas Chromatography (GC) for mixture separation, followed by detection using an appropriate instrument.

Chromatography is not ideal for some research applications, due to the time constraints for separation of components prior to detection. Therefore, when real-time or time-resolved phenomena are of primary interest, chromatography is limiting. Systems that are fast changing and require detailed monitoring are best studied using direct real-time analysis, normally based on spectroscopic detection [3–5].

Thermogravimetric analysis (TGA) is a well-established technique for studying various thermal processes in a wide range of materials. TGA was used over half a century ago to calculate activation energies for thermal events in polymers [6,7]. In 1965, Ozawa published a paper titled “A new method of analysing thermogravimetric data” [8]. In this paper, the thermal stability of calcium oxalate and nylon 6 were quickly and effectively determined compared to previous methods. Many TGA applications have since been published [9,10]. TGA as a stand-alone technique is limited to the determination of the thermal processes that give rise to weight loss and endothermic or exothermic events. The linking of TGA systems to fast detection techniques such as Mass Spectrometry (MS) and Infrared (IR) spectroscopy provides further insight into the physicochemical processes by monitoring the evolved gases [11]. Where necessary, residues of the sample, after thermal decomposition, can be analysed with techniques such as X-ray and electron microscopy [12].

---

[*] Corresponding author at: British American Tobacco Ltd, Regent’s Park Road, Southampton, SO15 8TL, United Kingdom.
E-mail address: ashley_davies1@bat.com (A. Davies).

https://doi.org/10.1016/j.tca.2018.08.007
Received 19 February 2018; Received in revised form 22 July 2018; Accepted 7 August 2018

Most TGA-GC–MS and TGA-MS systems have the ability to capture full mass spectral data however the majority of studies reported in the literature rely on Selected Ion Monitoring (SIM) [10,13] for identification. Without GC, SIM is unable to efficiently distinguish between interfering fragment ions if they are present. Consequently, TGA-MS data utilizing SIM must be viewed with caution [11,13]. This univariate approach is therefore of limited use when complex volatile mixtures are monitored, due to the high probability of interference from other compounds with isobaric ions. Research shows that even with Multiple Ion Monitoring (MIM), the probability of false positives for identification of peaks can be a major problem [14,15]. Fortunately, time and temperature resolved factor TGA systems with full spectral detection capabilities generate datasets suitable for multivariate analysis. The generated multivariate data can be mathematically decomposed to determine the number of main constituents in the evolved gas. This capability enables real-time *in situ* monitoring of systems which are otherwise very challenging to study.

Multivariate Curve Resolution (MCR) analysis based on Alternating Least Squares (ALS) has been used in chemical mixture analysis [18,19]. In 1987, Windig *et al.* used self-modelling curve resolution to model pyrolysis MS data [20]. The MCR-ALS technique is used to resolve overlapping spectral signals by self-modelling techniques based on factor analysis (or principal component analysis). It provides qualitative and semi-quantitative solutions for chemical mixture systems. Such chemometric techniques have found applications in different areas. Recent reviews have been reported covering applications in the pharmaceutical sciences [21], oil and gas explorations [22–24], food [25,26], toxicology [27] and environmental sciences [28,29]. In all of the above reports, a major step in the analytical process is to determine the optimum number of factors to enable the completion of the modelling process. Target Factor Analysis (TFA) is a curve resolution technique capable of testing the presence or absence of a target vector in a factor analysable data matrix [30,31]. Recently, Rostami et al. used Enhanced Target Factor Analysis (ETFA) on GC–MS data from red wine. They presented an approach where areas of feasible solutions in the TFA process could be identified [32]. When biomass materials are subjected to thermal degradation, characteristic volatile components are produced. These can be trapped and analysed by GC–MS but details of time resolved processes are lose and trapping can influence the chemical profile [33]. Direct TGA-MS however, provides a rich but complex full spectral data set with the clear advantage of gaining a detailed understanding of thermally induced processes in both time and temperature domains [34]. Such data sets are suitable for TFA [29].

Due to the complex nature of full spectra TGA-MS data, the use of selected ions to summarise and simplify the analytical process has been widely adopted in the literature. When it is heated, tobacco biomass can evolve thousands of compounds, with relative concentrations spanning many orders of magnitude [35]. There are a number of characteristic chemical subgroups, often leading to common fragment ions under electron impact ionisation, hence the assignment of individual compounds by SIM is precarious. Ahamad et al. used TGA-MS to monitor specific ions resulting from the pyrolysis of tobacco powder [36]. They were able to assign tentative identities to a number of compounds, including benzene, phenol, cresol, nitrosamines, butadiene and acrolein, all of which are cancer-causing chemicals. In similar work, Fischer et al. combined TGA with single photon ionisation-time of flight mass spectrometry (SPI-TOFMS) to investigate the evolved gases from tobacco pyrolysis [37]. The lack of fragmentation from SPI meant that Fischer's group could only assign tentative identities.

The application of full spectral analysis through chemometrics has the possibility of tackling research questions relating to specific volatiles of interest, based on full mass spectra profiles. In this paper, two types of biomass were examined at temperatures below 350 °C. Until recently, low temperature tobacco pyrolysis studies were rare [38,39]. The emergence of heated tobacco products as a product category has shown potential for greatly reducing the number of harmful components, in comparison to cigarettes. Recently, 22 different analytical methods were used to monitor 126 analytes in such a product [38]. This was a lengthy process which involved a range of targeted analytical methods. Fast analytical methods are desirable to enable screening of time/temperature dependent phenomena during development. Important insight can be gained from following the evolution of volatiles from biomass materials that are subjected to TGA-MS analysis.

Factor analysis has been used previously in biomass research, with application to Thermogravimetric analysis – Fourier-Transform Infrared spectroscopy (TGA-FTIR). Barontini *et al.* used TFA to target specific compounds formed during the pyrolysis of tobacco, and Coburn et al. used the technique to study the pyrolysis of tobacco saccharide ingredients [9,40]. In both cases, a reference spectral library of target compounds was used to test the presence or absence of the target compounds in the evolved gases and characterise their evolution profiles over time.

The extension of factor analysis to mass spectrometry is an important development in biomass research, because of the sensitivity and broad applicability of MS. Statheropoulos *et al.* used factor analysis and a contour variance diagram (ContVarDia) to interpret TGA-MS data from a synthetic polymer. They resolved spectra for individual thermal events, and interpreted these in terms of single compounds, or mixtures. We believe that this approach is not suitable for complex biomass, due to the number of co-evolving species. We report a novel approach which uses TFA to identify a set of target compounds in the volatile gases evolved from heated tobacco biomass, using a library of reference mass spectra to query the experimental data.

A critical step in the application of factor analysis is to determine the optimum number of factors to enable accurate modelling of the data [17,41]. This is an important subject of interest [16,42], and a critical step in building predictive models [43]. Over-fitting or under-fitting refers to the use of either too many or too few factors in the model, and relates to the trade-off between sensitivity and robustness. As a result, efforts are devoted to the determination of the optimum number of factors [44–48] to model a particular system. In contrast to the traditional selection approach [16,43,49], we used a strategy of monitoring the correlation between the known target spectra and the predicted spectra from the experimental data sets, as a function of the number of factors used in the model. Various spectral matching criteria have been reported and included in mass spectral search tools such as the National Institute of Science and Technology (NIST) software [50]. In this paper, the search strategy used involved inspection of the rate of change of correlation coefficient as a function of number of factors used. We call this a "deep search" strategy. Both simulated and experimental data were used to demonstrate this TFA approach. An overview of the TFA process is provided in Section 2.4.

2. Experimentation

The experimentation involved the analysis of simulated and laboratory generated data using two tobacco biomass varieties, Burley and Virginia. Burley tobacco is naturally low in sugar, but high in alkaloids, and Virginia tobacco is high in sugar, but has a slightly lower alkaloid content. An extended publication on tobacco varieties gives further details of the chemical differences between tobacco types [51]. The sample preparation, data simulation and TGA data generation and analysis are outlined below.

2.1. Sample preparation

The Burley and Virginia tobacco biomass were cryo-ground to approximately 5 μm particle size in a Retsch CryoMill [52]. Tobacco samples were conditioned for a minimum of 48 h at 22 ± 1 °C and 60 ± 2% relative humidity prior to TGA-MS analysis.

Eighteen different compounds were used as mass spectral target references based on previous headspace solid-phase microextraction