

ACCEPTED MANUSCRIPT • OPEN ACCESS

Passive breath monitoring of livestock: Using factor analysis to deconvolve the cattle shed

To cite this article before publication: Ben Langford *et al* 2022 *J. Breath Res.* in press <https://doi.org/10.1088/1752-7163/ac4d08>

Manuscript version: Accepted Manuscript

Accepted Manuscript is “the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an ‘Accepted Manuscript’ watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors”

This Accepted Manuscript is © 2022 The Author(s). Published by IOP Publishing Ltd..

As the Version of Record of this article is going to be / has been published on a gold open access basis under a CC BY 3.0 licence, this Accepted Manuscript is available for reuse under a CC BY 3.0 licence immediately.

Everyone is permitted to use all or part of the original content in this article, provided that they adhere to all the terms of the licence <https://creativecommons.org/licenses/by/3.0>

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions may be required. All third party content is fully copyright protected and is not published on a gold open access basis under a CC BY licence, unless that is specifically stated in the figure caption in the Version of Record.

View the [article online](#) for updates and enhancements.

Passive breath monitoring of livestock: Using factor analysis to deconvolve the cattle shed

Ben Langford¹, James M. Cash^{1,2}, Georgia Beel^{1,3}, Chiara Di Marco¹, Carol-Anne Duthie⁴ Marie Haskell⁴, Gemma Miller⁴, Laura Nicoll⁴, S. Craig, Roberts⁵ and E. Nemitz¹

¹ UK Centre for Ecology & Hydrology, Penicuik, UK

² School of Chemistry, University of Edinburgh, Edinburgh, EH9 3FJ, UK

³ Department of Environment and Geography, University of York, Heslington, York, YO10 5NG

⁴ SRUC, West Mains Road, Edinburgh, EH9 3JG, UK

⁵ Faculty of Natural Sciences, University of Stirling FK9 4LA, UK

E-mail: benngf@ceh.ac.uk

Received xxxxxx

Accepted for publication xxxxxx

Published xxxxxx

Abstract

Respiratory and metabolic diseases in livestock cost the agriculture sector billions each year, with delayed diagnosis a key exacerbating factor. Previous studies have shown the potential for breath analysis to successfully identify incidence of disease in a range of livestock. However, these techniques typically involve animal handling, the use of nasal swabs or fixing a mask to individual animals to obtain a sample of breath. Using a cohort of 26 cattle as an example, we show how the breath of individual animals within a herd can be monitored using a passive sampling system, where no such handling is required. These benefits come at the cost of the desired breath samples unavoidably mixed with the complex cocktail of odours that are present within the cattle shed. Data were analysed using positive matrix factorisation (PMF) to identify and remove non-breath related sources of VOC. In total three breath factors were identified (endogenous-, non-endogenous breath and rumen) and seven factors related to other sources within and around the cattle shed (e.g. cattle feed, traffic, urine and faeces). Simulation of a respiratory disease within the herd showed that the abnormal change in breath composition were captured in the residuals of the 10 factor PMF solution, highlighting the importance of their inclusion as part of the breath fraction. Increasing the number of PMF factors to 17 saw the identification of a “diseased” factor, which coincided with the visits of the three “diseased” cattle to the breath monitor platform. This work highlights the important role that factor analysis techniques can play in analysing passive breath monitoring data.

Keywords: Volatile organic compounds, Positive matrix factorization, PTR-TOF, Breath, Cattle, Respiratory disease

1. Introduction

1.1 Background

The spread of disease among livestock is of major concern for farmers, with zoonotic diseases alone estimated to cost the global economy in excess of \$220 Billion over the last decade (1). The direct costs (e.g., veterinary treatment and management and carcass disposal) only account for a small fraction of this total (20 billion) with the often less obvious indirect costs accounting for the majority. These can include adoption of measures to control the spread of disease, loss of animal productivity and disruption of supply chains. Wider societal costs include the overuse of antimicrobial drugs, which contributes to growing resistance in bacterial populations and increased greenhouse gas emissions associated with the final product (meat or milk). Therefore, the health of livestock has wide ranging societal impacts affecting not only the economy but the environment and public health.

Many of these impacts can be reduced through timely diagnosis, which allows infected animals to be identified and isolated from the herd which both limits transmission and allows targeted anti-microbial therapies to be administered. In the UK, most farms rely on an unstructured assessment of clinical signs carried out by the producer to identify diseased animals (2). Yet, livestock are adept at concealing illness with clinical symptoms absent in many cases resulting in poor diagnosis rates by farmers (3). As a result, diagnosis and intervention typically occur late in the course of the disease often leading to poorer outcomes, both in terms of veterinary costs and the long-term performance, health and welfare of the individual (4). In cattle, as in other livestock, these challenges have led to the increasing practice of metaphylactic anti-microbial therapy, where the entire calf cohort is treated to reduce the overall pathogen burden in both clinical and sub-clinical cases (5). While this has become an important and effective industry tool for reducing the negative health and performance effects associated with respiratory disease, it is at odds with industry commitments to reduce antimicrobial usage amid the growing concerns of increasing antimicrobial resistance (6).

Several other approaches for early detection of disease in livestock have been investigated, including use of thermography and analysis of behavioural changes. Pyrexia often precedes clinical signs by as much as three days (7) and thermography is viewed by many as a promising diagnostic tool (4). Indeed, similar temperature-based metrics including thermometric rumen boluses (8, 9) and inner ear temperature probes (10) have also proved excellent at detecting pyrexia when applied to cattle and are thought to offer producers a vital early warning of clinical infection. However, pyrexia is not disease specific and is notably absent in many chronic or sub-clinical cases (7). Furthermore, thermal changes in animals can be expected due to environmental factors, time of day and have also been linked to stress response. As a result, thermal based approaches currently lack the sensitivity and specificity required for disease detection. Changes in behaviour have also been shown to accompany disease with changes in activity and feeding evident up to 3 days before clinical signs are shown (11). However, like pyrexia, these behavioural changes are not specific to the disease state and do not allow causative agents to be identified.

Controlled breath analysis has shown some promise in detecting a range of respiratory and metabolic diseases in cattle (12), sheep (13), pigs (14) and goats (15). For example, Peled et al. (12) were able to detect 100% of cattle naturally infected with *Mycobacterium bovis* based on an analysis of breath (with 21% false positives). Their approach used a respiratory mask fitted to the animal for two minutes while an adequate volume of air could be sampled. The use of the mask allows for full control of the sampling process, with inspired air first passed through two charcoal filters to remove background volatile organic compounds (VOCs). However, habituating the animals to the mask takes time and the acquisition of the sample requires trained personnel to manually handle each animal, which can cause stress.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Other studies have successfully identified four compounds that appear indicative of bovine respiratory disease (phenol, benzothiazole, *p*-cresol and 5-octadecanal) by comparing VOCs emitted from the nasal secretions of healthy cows and those infected with BRD (16). While nasal swabs can be collected relatively easily, they still involve a degree of handling that might otherwise be avoided.

More recently, Gierschner *et al.* (17) showed it possible to monitor the health status of an entire herd by sampling the ambient air within a cattle shed to identify the presence of animals infected with paratuberculosis. The concept of “crowd-based” sampling is attractive because it captures the entire herd’s volatilome without the need for handling of individual animals. However, identifying the specific individuals that may be in need of treatment is not possible.

Here we present a breath monitor platform (BMP) that can be used to obtain breath samples from individual animals, with no requirement for handling. In this study our focus is on cattle, but the BMP could be applied to a variety of livestock and for the detection numerous respiratory and metabolic diseases (assuming suitable breath biomarkers are present). This approach avoids the stress associated with animal handling by obtaining samples passively. Although this method does not allow as much control over breath capture as is possible with a respiratory mask, we show how this can be compensated for, using factor analysis methods to identify and remove the non-breath components of the sampled air. Finally, we simulate markers of disease to determine whether individual “diseased” animals can be identified directly using positive matrix factorization.

2. Method

2.1 Breath Monitoring Platform

Breath measurements were obtained using a modified Beef Monitor Platform (Ritchie Implements Ltd., Forfar, Scotland), a piece of farm equipment routinely used to track the performance of beef cattle for finishing and hereafter termed the Breath Monitor Platform. The platform, shown in Fig. 1(a), comprises a weigh station with integrated water trough. Its width is designed to allow only one animal access at a time. As an animal enters the platform to drink, its electronic ear tag is automatically read (TruTest, New Zealand) and logged together with its weight, allowing the farmer to track the animal’s performance over time (Fig. 1(b)). The breath monitor differs only by having an enclosed hood over the water trough which allows exhaled breath to accumulate while the animal drinks (Fig. 1(c)). The exhaled air is then sampled in real-time, in our case by an online mass spectrometer, with the collected data immediately associated with the ID read by the TruTest system (Fig. 1(d)). Once the cow exits the platform, integrated fans within the hood quickly flush the exhaled air in readiness for the next animal.

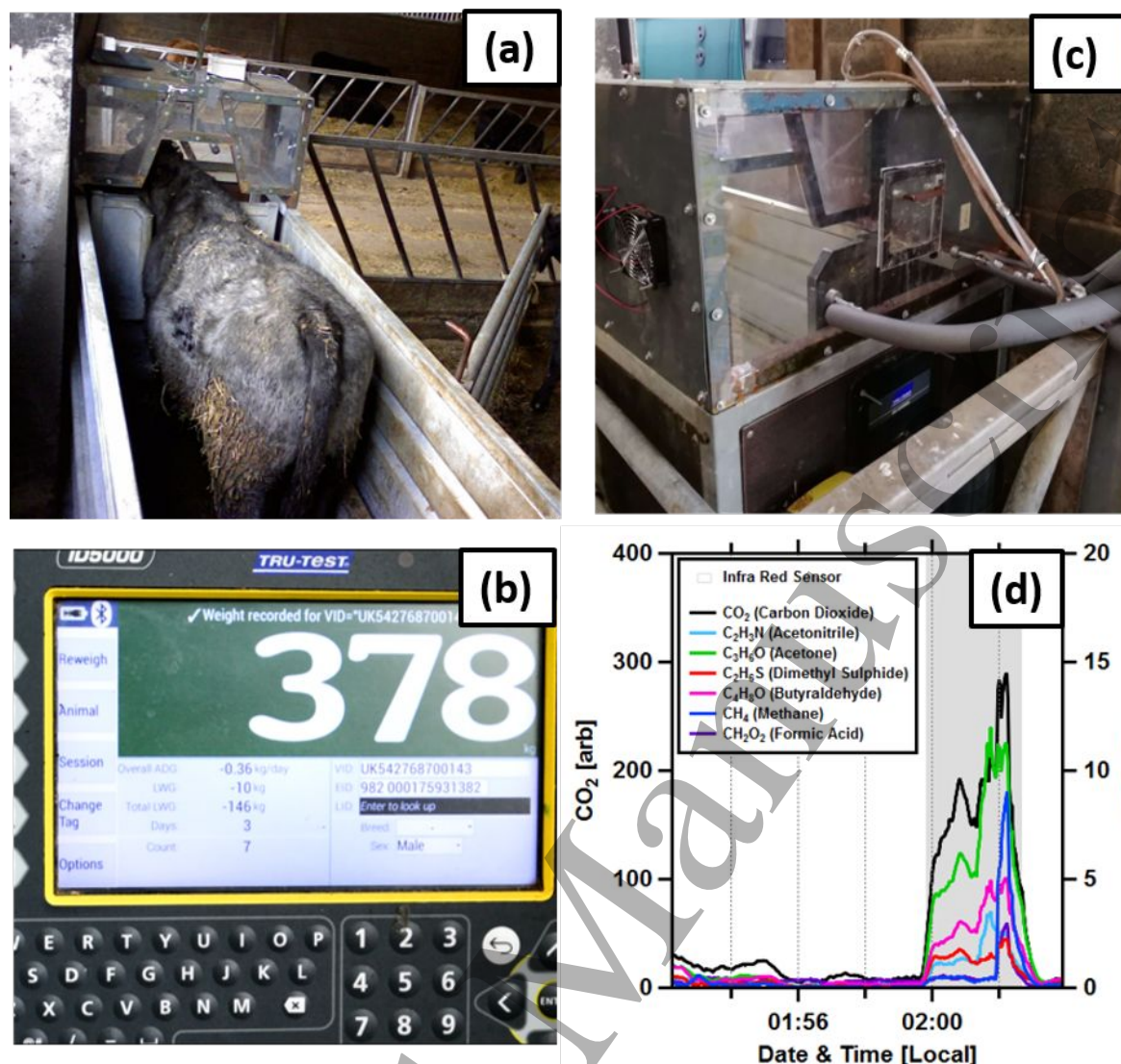


Figure 1. (a) Breath monitor platform, (b) TruTest system used to read cattle ID and log each visit to the BMP and (c) hood covering water trough to capture accumulated breath. Panel (d) shows an example time series of data collected by the BMP. The shaded grey area shows the signal from an infra-red sensor that indicated the presence of an animal within the hooded area of the BMP.

2.2 VOC measurements

Measurements of VOC concentrations were made using a proton transfer reaction – time-of-flight mass spectrometer with quadrupole ion guide (PTR-QiTOF). The instrument has been described in detail by Jordan *et al.* (18) and here we outline only those features pertinent to the experimental setup. The instrument was run in an H_3O^+ reagent ion mode with a drift tube pressure, voltage and temperature set to 3.2 mbar, 711 V and 80 °C respectively, yielding an E/N ratio 120 Td. Where E is the electric field and N is the number density of molecules within the drift tube.

The PTR-QiTOF was housed within a mobile laboratory which was parked adjacent to the cattle shed. Air from within the breath monitor hood was sampled along a ~10 m length of ¼" O.D. (I.D. 3.2 mm) perfluoroalkoxy (PFA) tubing at a rate of ~10 L min⁻¹. In order to limit the adsorption of compounds to the tube walls, the sample lines were wrapped with heating tape and insulated with pipe lagging to maintain a temperature of 60 °C. The PTR-QiTOF

subsampled from the main sample line at a rate of 300 ml min⁻¹ and measured mass-to-charge ratios between m/z 16 and m/z 200 with a one second time resolution.

The data were saved in hourly files which were analysed in Tofware (Tofwerks, Switzerland, version 3.2.2). Mass scale alignment was applied every 20 seconds using the NO⁺ peak (m/z 29.997) and two peaks associated with an internal diiodobenzene standard that is continually bled into the sample air stream ((C₆H₄I)H⁺, m/z 203.9431 and (C₆H₄I₂)H⁺, m/z 330.848). In total, 106 ions were detected that could be assigned a molecular formula within a tolerance of 50 ppm. Among the list were ammonia ((NH₃)H⁺), carbon dioxide ((CO₂)H⁺) and methane (CH₄)H⁺. Ammonia has a proton affinity close to that of water and can undergo back reactions in the drift tube making absolute quantification difficult without additional instruments (19). Carbon dioxide and methane have proton affinities lower than water and do not typically undergo proton transfer reactions. However, they are both present in such high concentrations in breath (ppm), that a small fraction undergo endothermic proton transfer reactions at the transition between the drift tube and time-of-flight chamber (20). This effect is enhanced in instruments with quadrupole ion guides due to the increased energies associated with the ion guide (Markus Mueller, personal communication). The CO₂ concentrations are a key indicator of breath and form a central part of the analysis presented below. However, because the proton transfer is endothermic and inefficient (21), it was not possible to provide accurate concentrations of these species based on the instrument transmission efficiency and no gas standards or ancillary measurements were available for direct calibration. Therefore, CO₂ and NH₃ concentrations are in units of normalised counts per second and all other compounds are presented as mixing ratios.

Normalised counts were calculated as

$$ncps = 1 \times 10^6 \left(\frac{RH^+}{M19 + M37} \right), \quad (1)$$

where RH^+ is the transmission corrected ion signal (cps) and $M19$ and $M37$ are the transmission corrected ion counts for the primary ions (H₃O⁺) and first water cluster (H₃O⁺ + H₂O⁺), respectively.

VOC measurements were analysed using positive matrix factorization (PMF) (22) to identify and remove non-breath components from the dataset. PMF is a receptor-only, bilinear model that, when applied to PTR-QiTOF data, assumes the total mass spectrum of a quantity measured over time represents the linear combination of a number of discrete sources or 'factors', (p) each with a distinct chemical signature (or mass spectral profile) that does not change over time, such that (22)

$$\mathbf{X} = \mathbf{GF} + \mathbf{E}. \quad (2)$$

In the case of VOCs measured within the breath monitor platform, \mathbf{X} comprises the breath sample of interest plus the cattle shed background. In practical terms \mathbf{X} is a two-dimensional $m \times n$ matrix, where rows (i) are the m individual one second mass spectra and the columns (j) are the counts per second of n individual ions. \mathbf{F} is two-dimensional $p \times n$ matrix that represents a number of mass spectral profiles, or factors, which best describe specific sources within the cattle shed, such as the breath sample, animal waste (e.g. faeces and urine), feed or farm traffic. The relative contribution of these factors to the total measured odour is given by the elements of an $m \times p$ matrix, \mathbf{G} . Accordingly, any change in the measured mass spectrum of the sample air (\mathbf{X}) is a consequence of the varying contributions of the individual factors over time plus some residual spectrum, \mathbf{E} ($m \times n$, with elements e_{ij}). Here, our objective is to use PMF to identify the non-breath components from the passive

sample so they may be removed from the analysis leaving only the desired breath contributions plus any residuals. Importantly, no *a priori* assumptions are made about the mass spectral profile of individual factors or their contribution to the measured signal. Rather, the number of factors (p) is decided by the user, with \mathbf{F} and \mathbf{G} calculated iteratively to minimize the sum of the squared residuals, Q , relative to their respective uncertainties as

$$Q = \sum_{i=1}^m \sum_{j=1}^n \left(\frac{e_{ij}}{\sigma_{ij}} \right)^2. \quad (3)$$

Here, σ_{ij} are the uncertainties associated with the individual elements of \mathbf{X} , which are typically calculated as the signal-to-noise ratio of the individual measurements (22) shown here as

$$\sigma = \alpha \sqrt{\frac{l}{t}}, \quad (4)$$

where l , is the ion signal, t , is the acquisition or dwell time in seconds and the term α is a factor applied to account for the fact that the signal of a signal ion is not constant, but rather part of a Gaussian distribution of pulse areas. Allan et al. (23) determined the standard deviation of this distribution to be 0.68 which when convolved with a Poisson distribution yields a value of 1.2.

Data collected from the breath monitor platform were here analysed using the PMF evaluation tool (PET) described in Ulbrich et al. (24) which is based upon the PMF2 algorithm that uses a weighted least squares approach. In this particular tool, the maximum size of a data matrix is restricted to 100,000 rows and, therefore, the measured data were first smoothed (3-point box smoothing) and subsequently resampled to three second data reducing the number of measurements. Consequently, the data matrix \mathbf{X} and error matrix had dimensions of $m = 70,000$ and $n = 106$.

3. Results

3.1 Standard Analysis

The breath monitor platform was trialled over a period of 2.5 days with a herd of 26 beef cattle (10 female, 16 male; 8 Beef Short Horn cross, 14 Aberdeen Angus cross and 4 Limousin cross) aged between 9 and 15 months. Typically, each animal visited the platform four times a day for an average of 3.5 minutes giving a total breath measurement time over the study of ~35 minutes per individual (see Table 1).

Table 1. Summary of the 26 cattle used in the study and time spent within the breath monitor platform (BMP) during the 2.5 days of measurements.

Cow #	Breed*	Sex	Number of BMP Visits [#]	Total time in BMP [min ⁻¹]	Average time spent in BMP per day [min ⁻¹]	Average duration of BMP visit [min ⁻¹]
1	BSX	M	7	28.4	11.4	4.1
2	BSX	M	6	14.0	5.6	2.3
3	AAX	M	12	42.4	16.9	3.5
4	AAX	M	10	45.6	18.2	4.6
5	AAX	M	12	36.0	14.4	3.0
6	AAX	M	14	58.5	23.4	4.2
7	AAX	M	16	52.2	20.9	3.3
8	AAX	M	15	42.0	16.8	2.8
9	BSX	M	14	51.0	20.4	3.6
10	AAX	F	11	28.4	11.3	2.6
11	LX	F	7	20.2	8.1	2.9
12	AAX	F	11	49.2	19.7	4.5
13	AAX	F	10	35.5	14.2	3.6
14	LX	F	10	46.5	18.6	4.6
15	AAX	F	12	43.1	17.2	3.6
16	AAX	F	10	36.4	14.5	3.6
17	AAX	F	7	27.5	11.0	3.9
18	LX	F	9	26.0	10.4	2.9
19	AAX	F	11	65.2	26.1	5.9
20	LX	M	12	51.5	20.6	4.3
21	AAX	M	9	25.9	10.3	2.9
22	BSX	M	13	19.6	7.8	1.5
23	BSX	M	11	36.2	14.5	3.3
24	BSX	M	9	38.1	15.2	4.2
25	BSX	M	15	42.8	17.1	2.9
26	BSX	M	15	56.4	22.5	3.8

* BSX = Beef Short Horn cross; AAX = Aberdeen Angus cross; LX = Limousin cross.

When an animal enters the breath monitor platform its breath and body odour accumulate in the headspace above the water trough. The PTR-QiTOF measures these volatiles together with the host of compounds present in the background air. The cattle shed comprises a particularly complex blend of VOCs with a number of strong local background sources that include the animals' waste products (faeces, urine), straw, feed, local farm traffic and the regional background. Removing these non-breath components from the periods where a cow was present within the BMP is not a trivial task, not least because the background odours can vary on a timescale shorter than the average time spent at the water trough. As an initial step, it is necessary to determine the periods when an animal was present in the BMP. An infrared sensor was used to indicate when an animal's head was above the water trough, but this was found to be unreliable due to debris falling onto the reflector and causing false readings. As an alternative, we chose to use the time series of CO₂ signal to indicate the

1
2
3 presence of an animal within the BMP based on a threshold value. A low frequency
4 background concentration was established for each compound by calculating a centred
5 running minimum which was subsequently subtracted from the measurements. The running
6 minimum was set to 25 minutes, an interval greater than the typical duration of any single visit
7 to the BMP. Figure S1 shows the time series of CO₂ and acetaldehyde measured by the BMP
8 with and without the low frequency background removed.
9

10
11 Removing the low frequency baseline works well for some compounds (e.g. CO₂ and
12 acetone), but it is less successful for others, that have a strong diurnal pattern (e.g. methanol
13 and acetaldehyde), for which further steps are required to isolate the breath sample. One
14 solution is to subtract a background that is calculated as a linear interpolation between the
15 period before and after the animal enters the breath monitor. Figure 3 shows how this works
16 in practice. Here, a CO₂ threshold of 135 ncps is used to indicate the presence of an animal
17 within the BMP. At around 18:05 the CO₂ concentration remains above the threshold despite
18 the infrared sensor (shaded grey areas) not indicating the presence of an animal within the
19 BMP. An analysis of webcam footage showed that in this case, the animal had remained on
20 the platform but had retracted its head from the hooded area. From this position, its breath
21 was still carried into the sample area and detected by the PTR-QiTOF. This was a fairly
22 common occurrence and was a second motivating factor in using a CO₂ threshold to alert of
23 an animal's presence, rather than the signal from the infrared sensor.
24
25

26
27 The interpolated background works well for compounds that have large breath
28 concentrations relative to the background (e.g. for CO₂ (Fig. 2a) and acetone (Fig. 2b)), but is
29 less effective for compounds like ammonia (Fig. 2f), that have a high background relative to
30 the breath component or, in the case of acetic acid (Fig. 2e), that have a background that
31 varies on a similar timescale to a typical visit to the water trough (e.g. 2 to 4 minutes). In these
32 situations, the concentrations measured from within the BMP may fall below that of the
33 interpolated background, resulting in a negative breath sample. While uptake of compounds
34 to the lungs of cattle is theoretically possible, acetic acid is a known component of bovine
35 breath and, therefore, in this instance the negative concentrations can be attributed to a
36 decrease in the background during the visit to the BMP. In such cases, the measured data
37 cannot be reliably attributed to the breath of the animal and, therefore, a different approach is
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

required.

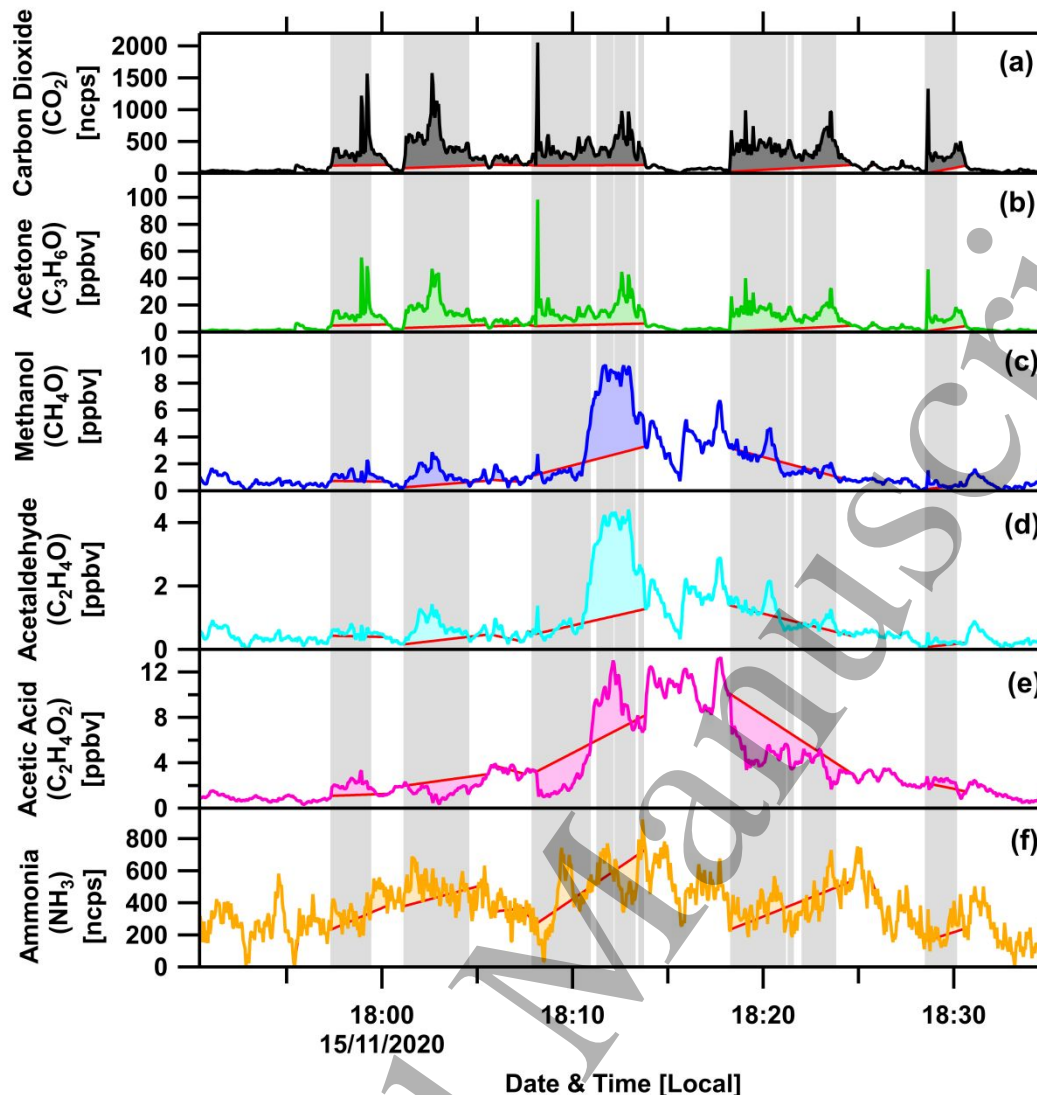


Figure 2. Example time series of CO₂ (a), acetone (b), methanol (c), acetaldehyde (d), acetic acid (d) and ammonia (f) concentrations after a running minima background has been subtracted. The solid red line represents an interpolated background between the period before and after the cow enters the BMP. The presence of an animal was determined on the basis of a CO₂ concentration > 135 ncps. The coloured shaded areas represent the portion of the signal that would be attributed to the animal within the BMP and the grey shaded areas indicate when the infrared sensor detected a cow within the BMP hood.

3.2 PMF Analysis

PMF analysis was applied to the set of VOCs measured from within the BMP to see if factors unrelated to the breath could be identified and removed. The analysis was run iteratively, gradually increasing the number of factors used in the solution until a minimum in the Q/Q_{exp} could be achieved (see figure S3 of the SI). The final solution selected had a total of 10 factors and the individual mass spectra (F) and their time series (G) are shown in figure 3 together with the residuals. A further reduction in the Q/Q_{exp} could be achieved in solutions with up to 14 factors. However, moving beyond 10 factors did not significantly reduce the residuals as

shown in figure S4 of the Supplementary Information. The robustness of the solution was assessed by varying both the fPeaks (rotations) and initialisation seeds used in the PMF algorithm (Paterro, 2005). The Q/Q_{exp} showed a clear minimum at the zero fPeak indicating that the solution had found the global minimum. The same 10 factors were identified for each initialisation seed and the attribution of variance to each of those factors showed minimal differences indicating the solution to be robust. Based on this analysis uncertainty estimates for each of the three breath factor profiles were established and are shown in Section 2.1 of the SI.

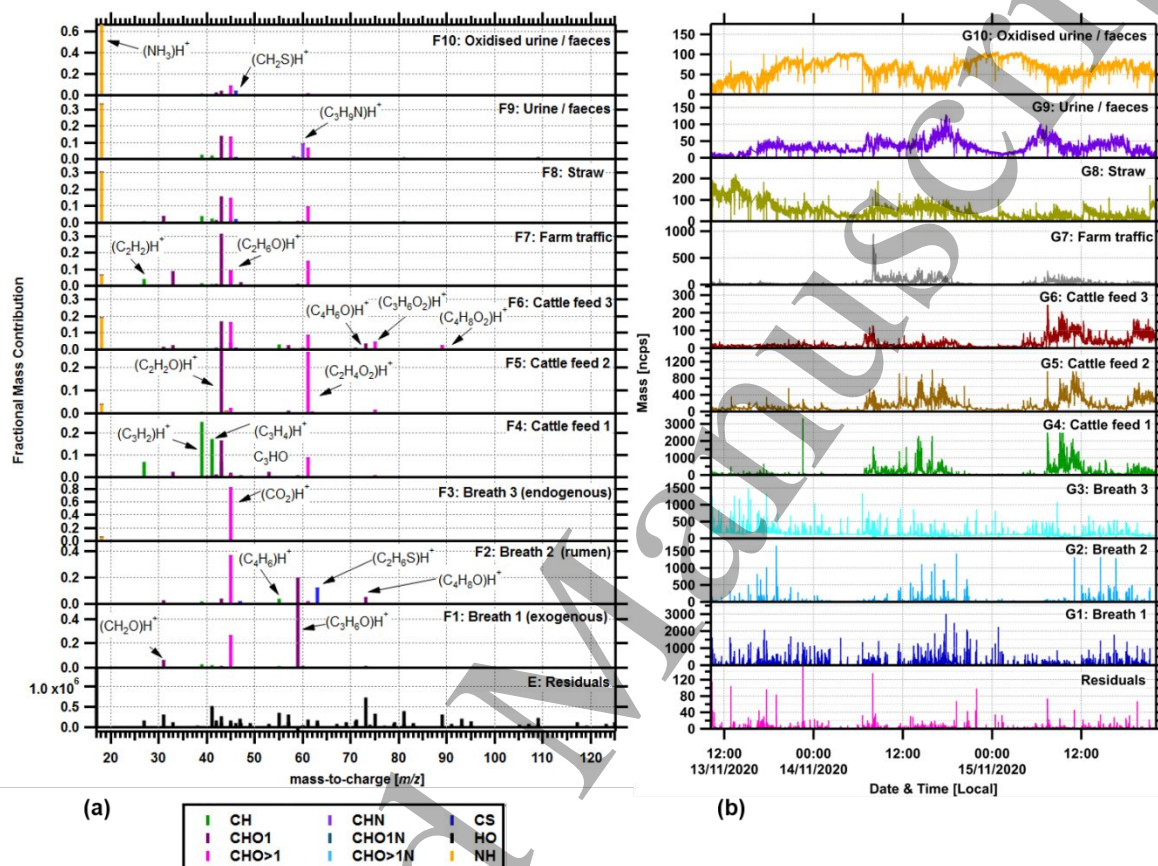


Figure 3. Panel (a) shows the 10 mass spectral profiles identified by the PMF analysis together with the residuals, colour-coded by chemical ‘families’. Panel (b) shows the time series of these factors over the 2.5 day measurement period. Panel (a) is shown again as Figure S5 of the Supplementary Information, with the scale adjusted to show the contributions of the minor peaks.

Each factor was assessed and named based on its temporal pattern and chemical composition to determine which sources within the cattle shed they might represent. The interpretation of the factor profiles was aided by headspace analysis of potential sources found within the cattle shed which included samples of straw, sawdust (used as bedding in adjacent barns), faeces, saliva and skin swabs (see SI Section S3). Quantification of the local sources in this way allows for a more meaningful interpretation of the data, especially in this context where there are no previous studies to compare to. However, this step is not strictly necessary, because following identification of the factors associated with the BMP, all remaining factors are ultimately removed

Three factors were associated with animals within the BMP and were identified on the basis of the short-term spikes in concentration that coincided with the periods where animals were

accessing the BMP. The identification of more than one breath factor is because the chemical composition of the breath is likely influenced by how recently an animal has fed or whether eructation occurred during the visit, which would release gases from the rumen with a different composition to alveolar breath. Breath factor one was characterised by elevated acetone, CO₂ and formaldehyde, whereas breath factor three was dominated by CO₂ and NH₃. In contrast, breath factor two had a large contribution from dimethyl sulphide and 2-Butanone, two compounds known to be emitted from the rumen during eructation (25). It is therefore likely that it represents emissions from the rumen, while breath factors one and three are associated with alveolar breath. The fact that breath factor one contains (C₃H₂)H⁺ and (C₃H₄)H⁺, two ions that are characteristic of the cattle feed, may indicate that factors two and three separate on the basis of how recently an animal has fed, with factor three representing endogenous breath and factor one contaminated by exogenous sources. The VOCs detected while an animal is within the BMP may also originate from its skin/fur as well as from extraneous sources, such as urine/faeces that may be present on the animal. These factors therefore may still have some artefacts, but in this instance, these could not be further separated by increasing the number of factors, likely because they are highly correlated in time which PMF cannot resolve.

The dominant odour within the cattle shed originated from the cattle feed which was a blend of silage, whole crop barley, brewer's grains, barley and molasses. The feed mixer dispensed the feed along the edge of the pens at around 08:00 hrs each day. The cattle typically eat what is in reach and any residual feed is swept closer to the pen, either later in the evening or around 07:00 hrs the following morning. Factors F4, F5, F6 and F7 each show peaks coinciding with the delivery of the feed, but only factor 4 has significant contributions from the (C₃H₂)H⁺ (C₃H₄)H⁺, (C₂H₂O)H⁺ and (C₂H₄O)H⁺ ions which were found to be characteristic of the feed when samples were analysed in the laboratory (see Section S3 of the Supplementary Information). These same markers ((C₃H₂)H⁺ and (C₃H₄)H⁺) were not particularly prevalent in factors F5 and F6, but the temporal profiles of these factors closely resembled that of the cattle feed factor, but with a reduced contribution to the overall mass. Our hypothesis is that factors F5 and F6 represent either different components of the feed mix that may dry out at different rates (or have different volatilities), or that they represent a more aged or oxidised version of the cattle feed factor. This last assertion is supported by the higher O:C ratios (~1.0) of these factors compared to that of factor F4 (O:C = 0.28).

Factor six showed the largest peak associated with the passing of the feed truck on the 14th of November. Our assumption is that this spike is associated with the exhaust fumes that briefly accumulated as the feed truck paused next to the BMP. As well as delivering the feed, there is a constant stream of farm vehicles around the cattle shed which all contribute to this factor as well as emissions from road traffic on the country road (A702) which runs to the east of the farm. The chemical profile of the traffic factor again has a significant contribution from the (C₂H₂O)H⁺ ion which could either be ethenone or a fragment of the acetic acid ion and has previously been found as a useful marker of traffic (26, 27). In addition this factor has much higher contributions of CO₂ and ammonia (for diesel vehicles with urea catalyst), two further known components of vehicle exhaust. Finally, factor six had the largest contributions of benzene, toluene and xylene (not visible at this scale) giving further evidence of this factor most likely being associated with vehicle emissions.

The final three factors were more difficult to identify with each having strong contributions from ammonia. The mass spectral profiles of factors F9 and F10 were very similar except for a much larger contribution of trimethylamine (TMA, (C₃H₉)H⁺) in factor F9. Trimethylamine is formed when the urea and trimethylamine N-oxide contained in urine are enzymatically processed to TMA and NH₃ by the microbes found in faeces (28). The time series of these two factors appear somewhat anti-correlated, which can be indicative of a single factor that has split. However, these two factors appear very robust, appearing early in

the solution (from 6 factors onwards), so factor splitting is thought unlikely in this case. Figure S6 shows that factor 10 has significantly more NH_3 , whereas factor 9 has more TMA. Our hypothesis is that factor F9 is associated with urine/faeces which generate emissions of TMA, particularly during the daytime when there is a fresh supply of faeces and urine and factor F10 represents a somewhat aged or oxidised version of the urine/faeces factor (Factor F10 O:C = 0.96; Factor F9 O:C = 0.78). Trimethylamine has a short atmospheric lifetime with respect to the primary atmospheric oxidant OH (7 to 10 hours), but is even more rapidly removed from the atmosphere due to condensation onto particles. Sintermann *et al.* calculated the typical condensation sink for trimethylamine within a cattle shed environment to be on the order of 30 to 1000 seconds (28). In contrast, the lifetime of NH_3 is slightly longer (hours to days with respect to OH) which may account for the distinct differences in the diurnal patterns of these two factors. Furthermore, NH_3 has other strong local sources, with emissions carried on the breath and from the skin/fur of the cattle as well as from traffic and the straw bedding material.

Finally, factor F8 was attributed to emission from the straw bedding used inside the cattle shed. Laboratory measurements of the bedding show it to have characteristic peaks of ethenone, CO_2 and acetic acid as well as contributions from NH_3 and the monoterpene fragment $(\text{C}_6\text{H}_8)\text{H}^+$. Factor F8 has a similar mass spectral profile to that of the straw sample tested and its time series showed a gradual decrease over time which is consistent with a decrease in source strength, as might be expected from straw.

Figure 4 (a) shows the total ion signal measured by the PTR-QiTOF together with the stacked contributions from individual factors identified using PMF. Panel 4 (b) shows the same time series but with the non-breath components removed plus the residuals. The remaining peaks, clearly coincide with periods where an animal accessed the water trough (highlighted in grey). The contribution of the breath factors is not zero between visits for two reasons. Firstly, exhaled breath makes up a portion of the background air within the cattle shed and secondly, PMF has a non-negative constraint so the average of a factor can never be exactly zero.

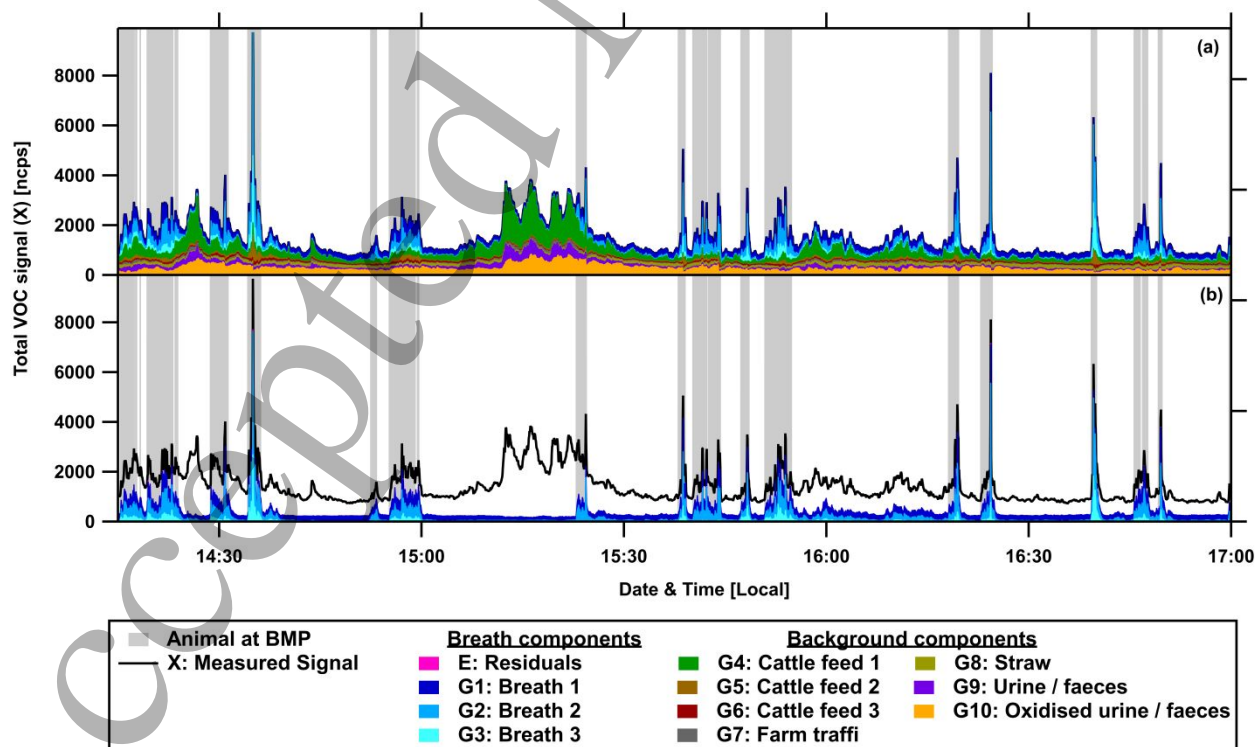


Figure 4. Panel (a) shows an example time series of the total ion signal measured by the PTR-QiTOF (thick black line) as well as the contribution from the 10 factors identified by PMF. Panel (b) shows the measured total ion signal together with just the three breath components and residuals. The shaded grey areas highlight the periods where animals visited the breath monitor platform.

After removing the non-breath components, the remaining data underwent the same analysis as the raw data described in Section 3.1, with a low frequency baseline first removed using a running minimum and subsequent subtraction of an interpolated background. Figure 5 shows a linear interpolation between the periods before and after the animal has entered the BMP is now more successful having first removed the non-breath components identified using PMF analysis. In contrast to Fig. 2, there are no negative contributions during the interpolated periods (shaded areas).

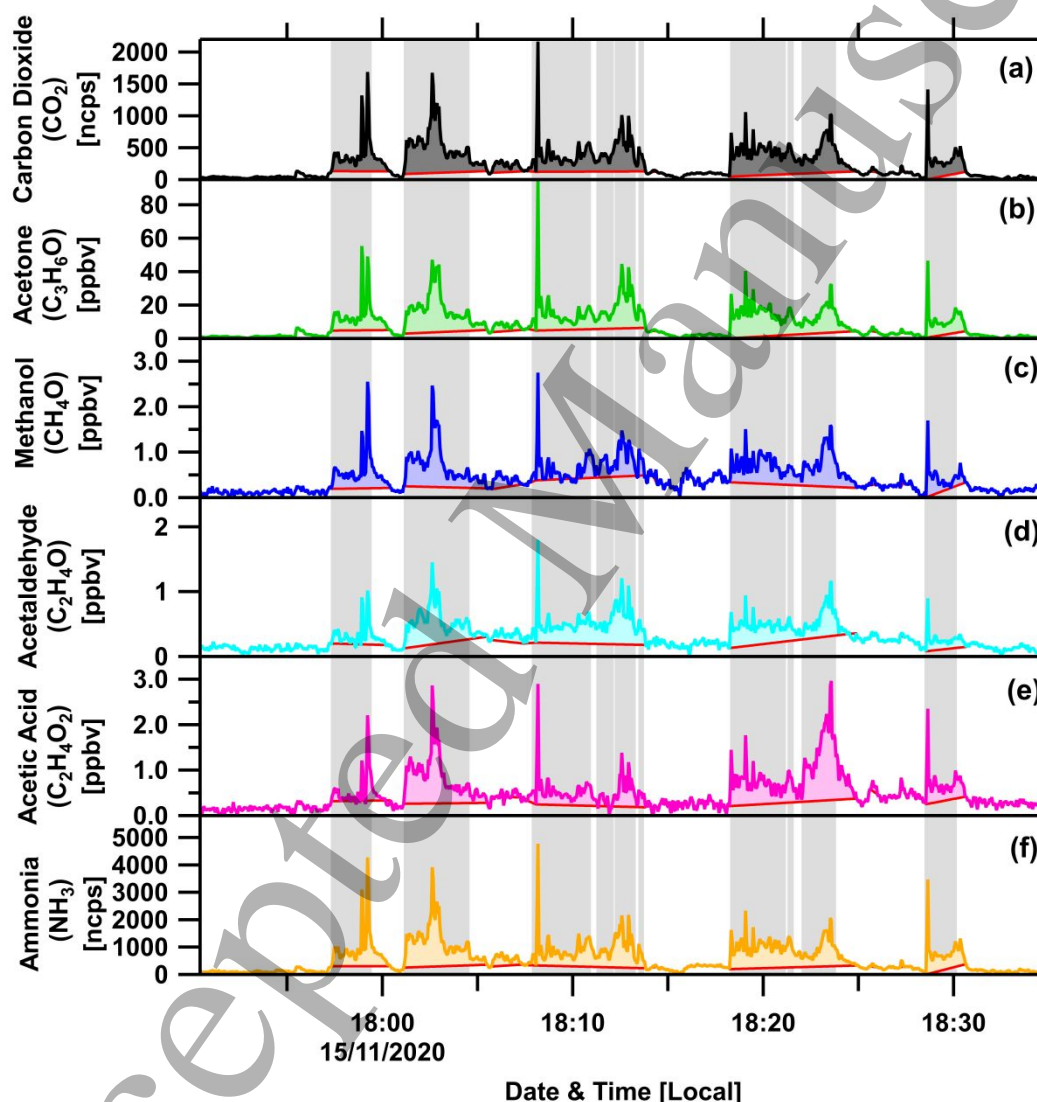


Figure 5 Example time series of CO₂ (a), acetone (b), methanol (c), acetaldehyde (d), acetic acid (d) and ammonia (f) concentrations after a running minima background has been subtracted and the contribution from non-breath sources have been identified and removed by PMF. The solid red line represents an interpolated background between the period before and after the cow enters the BMP. The presence of the cow was determined on the basis of a CO₂ concentration > 135 ncps. The coloured shaded areas

1
2
3 **represent the portion of the signal that would be attributed to the animal within the BMP**
4 **and the grey shaded areas indicate the periods where an animal visited the breath**
5 **monitor platform.**
6

7 *3.3 Variability of breath samples*

9 Processed BMP measurements were combined with the ear tag data recorded by the
10 TruTest system to give the average composition of passive air samples acquired for each
11 animal over the 2.5 day measurement period. Figure 6 compares the concentrations of CO₂,
12 methanol, acetone, acetaldehyde, acetic acid and ammonia between animals when the data
13 is processed using the standard analysis, where each compound has had (i) a running
14 minimum baseline removed and (ii) the subtraction of a linear interpolated background. In
15 addition, the contribution from the 10 individual factors to the measured total are included
16 together with both positive and negative residuals. Here, positive residuals represent the
17 shortfall between the sum of the factors (**GF**) and the measured concentration (**X**) and
18 negative residuals represent periods where the PMF solution is greater than the measured
19 concentration.
20
21

22 Combining the standard analysis with the output from the PMF analysis allows any variation
23 between passive breath samples to be attributed to either a genuine difference in breath
24 composition or the result of contributions from local sources that could not be entirely removed
25 using the standard analysis. Adopting the PMF approach allows for the removal of the non-
26 breath components and, as can be seen in Figure 6, focusing on the breath components only
27 (e.g. Breath 1 + Breath 2 + Breath 3 + residuals (both positive and negative)) removes much
28 of the within-herd sample variability, which, for several of the ions shown, was strongly
29 influenced by background odours. Section S4 of the Supplementary Information gives a more
30 detailed summary of the average concentrations of ammonia, methanol, acetaldehyde,
31 acetone and acetic acid for each of the 26 animals based on (a) the standard analysis and (b)
32 an analysis where the non-breath components are first identified and removed before applying
33 the standard background removal steps.
34

35 The magnitude of the breath sample is in part related to how long each animal spends within
36 the BMP, the exact head location relative to the sample inlet and the dispersion/dilution of the
37 exhaled breath during the visit. These effects were accounted for by normalising by the total
38 CO₂ concentration. For CO₂ and acetone, which were the two largest components of breath,
39 the fractional contribution of different factors to the measured total is almost entirely made up
40 from the three breath factors. CO₂ is dominated by factor 3 (endogenous breath) and acetone
41 is mostly attributable to factor 1 (exogenous breath), with both having small contributions from
42 factor 2 (rumen). In contrast, methanol, acetaldehyde, acetic acid and ammonia have much
43 larger contributions from non-breath sources, and it is these contributions that account for the
44 majority of the variation in total concentrations between cattle. This is a stark demonstration
45 of the importance of using a factor analysis like PMF to identify and remove the non-breath
46 components that can remain even after careful subtraction of the cattle shed background.
47
48
49
50
51
52
53
54
55
56
57
58
59
60

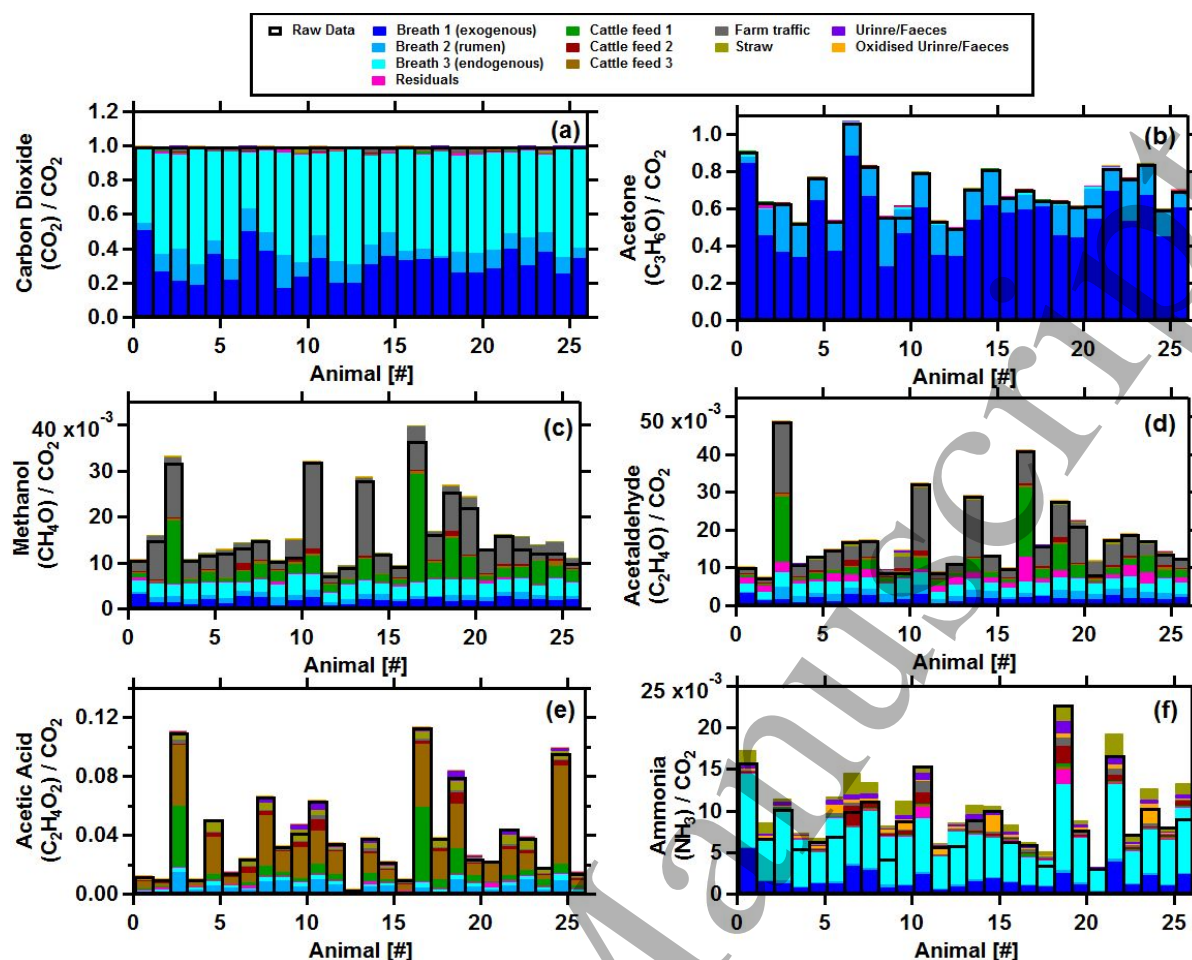


Figure 6. Stacked bars comparing the average concentrations of CO₂, acetone (b), methanol (c), acetaldehyde (d), acetic acid (e) and ammonia (f) measured from within the BMP for the 26 individual animals. Each total has had a running minima baseline and interpolated background subtracted before being normalised by CO₂. The relative contribution from the 10 factors identified by PMF is shown as stacked bars, with the breath component highlighted in shades of blue. The pink bars represent the positive residuals which are the portion of the measured VOC that could not be explained by the PMF algorithm. The negative residuals, where the PMF solutions is greater than the measured raw data, can be seen where the coloured bars exceed the solid black bars (raw data). VOC and CO₂ concentrations were in ncps and therefore ratios are arbitrary.

4. Discussion

4.1 Deconvolution of the cattle shed background

In order to identify changes in the health status of an animal, a robust measurement system is required to ensure observed variations in breath samples reflect metabolic processes rather than changes in environmental conditions. Figure 6 has demonstrated that the passive collection of breath from within the BMP cannot make this distinction for all compounds, and is particularly ineffective for compounds with a large background concentration that shows short term (minutes) variation. Figure 7 gives a broader assessment, ranking the 50 most

abundant ions detected in the breath of the cohort of cattle. Divergence between the breath components, identified by PMF, (closed blue circles) and the total measured signal (open black circles) give an indication of how much of the non-breath components remain following basic steps of background removal (e.g. running minimum baseline and interpolated background subtraction). Averaged over the cohort of 26 cattle, the contribution of background sources to the breath measurement can be up to a factor of two and many times larger for individual visits to the BMP.

This clearly demonstrates that breath samples collected passively must be carefully processed to tease apart the non-breath components that typically dominate the cattle shed air space. Failure to do so could mean the differences in observed concentrations between visits may reflect the changing environmental conditions rather than true variation in breath composition. The examples shown deliberately focus on compounds where background interferences are large, but Figure 7 demonstrates that for many of the breath compounds detected using the BMP, background interferences are negligible following basic processing of the data (e.g. removal of baseline and linear interpolated background). For example, dimethyl sulphide, $(C_2H_6S)H^+$, a compound primarily emitted from the rumen and with very few other sources within the cattle shed, shows no difference between the breath component and measured total. With this in mind, the requirement of PMF analysis to deconvolve the passive breath measurements is dependent on the markers of disease and whether they are unique to breath or emitted from other sources in the local environment.

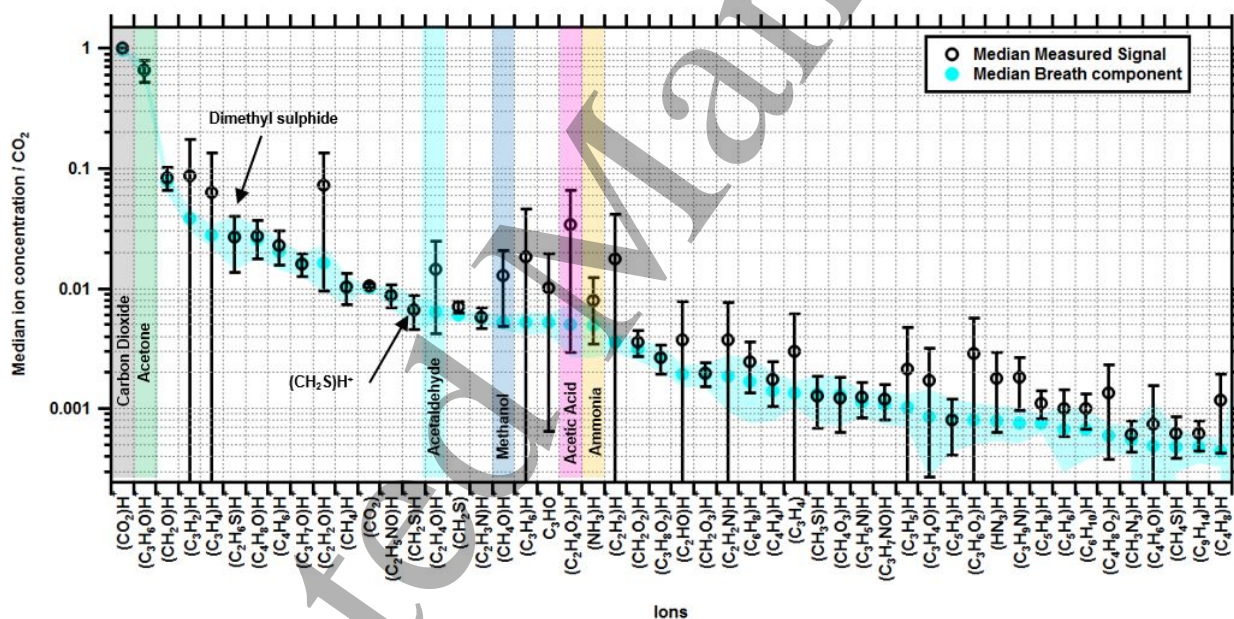


Figure 7. The median normalised breath concentration of the 50 most abundant ions identified in the breath of the 26 cattle ($\pm 1\sigma$). Blue circles represent the fraction of the measured signal that was attributed to components of the breath by PMF analysis and black circles represent the median measured signal following the removal of a running minima baseline and interpolated background. Note the logarithmic scale. VOC and CO_2 concentrations were in ncps and therefore the ratios can be considered arbitrary.

4.2 Potential for disease detection

Having established the suitability of PMF to remove background interferences from breath measurements collected using the BMP, we now explore whether this approach allows for variation in cattle breath composition, as might be associated with respiratory illness, to be reliably detected. In this pilot study, all animals were in a healthy condition and free from

1
2
3 respiratory disease, hence none of the identified factors was directly associated with a
4 particular individual or subset of the cattle. We therefore, chose three of the 26 cattle at random
5 (#4, #8 and #25) and altered the mass spectral profile associated with their breath by
6 upregulating methanol, acetaldehyde and acetic acid, three of the example compounds shown
7 in Figures 2, 5 and 6. In practice this involved four steps: (i) isolating breath factor 3
8 (endogenous breath), (ii) removing the baseline from this factor (running minima and linear
9 interpolation of the background) (iv) multiplying the marker compounds (methanol,
10 acetaldehyde and acetic acid) by a given scaling factor for the periods where cows #4, #8 and
11 #25 visited the BMP and (v) adding this modified time series to the raw data prior to analysis
12 by PMF.
13
14

15
16 Previous studies have found statistical differences in the concentrations of breath
17 compounds between healthy animals and those infected with *mycobacterium bovis*. For
18 example, Elis et al. (29), identified 14 marker compounds, with the majority showing an
19 upregulation ranging between 10% for 1-1-dimethyl-2-(1-methylethyl) cyclopropane and a
20 factor of 18 for 1,1-diethoxyethane. The downregulated compounds showed a more modest
21 change ranging between - 4% for 2-ethyl-1-hexanol and -43% for 3-heptanone. Overall,
22 upregulated compounds increased by an average factor of 3.9 and downregulated compounds
23 by a factor of 0.78. Using these results as a guide, we upregulated methanol, acetaldehyde
24 and acetic acid each by a factor of 3.9. The modified data were then re-analysed using PMF
25 and the results are compared with the unmodified data in Figure 8. The PMF analysis was
26 very consistent with the unmodified version, with a 10 factor solution still found to be the
27 optimum solution. Rather than a marked increase in the modified breath factor 3, the additional
28 “diseased” signal can be clearly seen within the residuals of the three compounds. Yet, the
29 solution is not perfect. The residuals do not increase by an amount consistent with a factor of
30 3.9 increase in breath factor 3, in part because PMF works by trying to minimize the residuals.
31 Instead, some of the additional signal appears to increase the magnitude of the traffic factor.
32 This result highlights the limitations of the PMF approach by demonstrating that the attribution
33 of mass to specific factors represents one of many possible solutions and, therefore, this
34 approach is perhaps less suitable when absolute values of breath composition are required.
35
36

37 This simple exercise has highlighted the importance of including the residuals together
38 with any identified breath components from the PMF analysis. This is because a single breath
39 factor taken in isolation represents the average composition of the herd and therefore, does
40 not capture individual variations which might be indicative of disease. Divergences from the
41 “average breath” are instead captured in the residuals, the portion of the measured data that
42 cannot be described by any one factor. Therefore, it is only the combination of the identified
43 breath factor(s) with the residuals that allows for variation between individual breath
44 measurements to be captured. For those individual variations to be captured as a specific
45 factor, the PMF solution would need to increase the number of additional factors included
46 within the solution. This concept is explored further below.
47
48
49
50
51
52
53
54
55
56
57
58
59
60

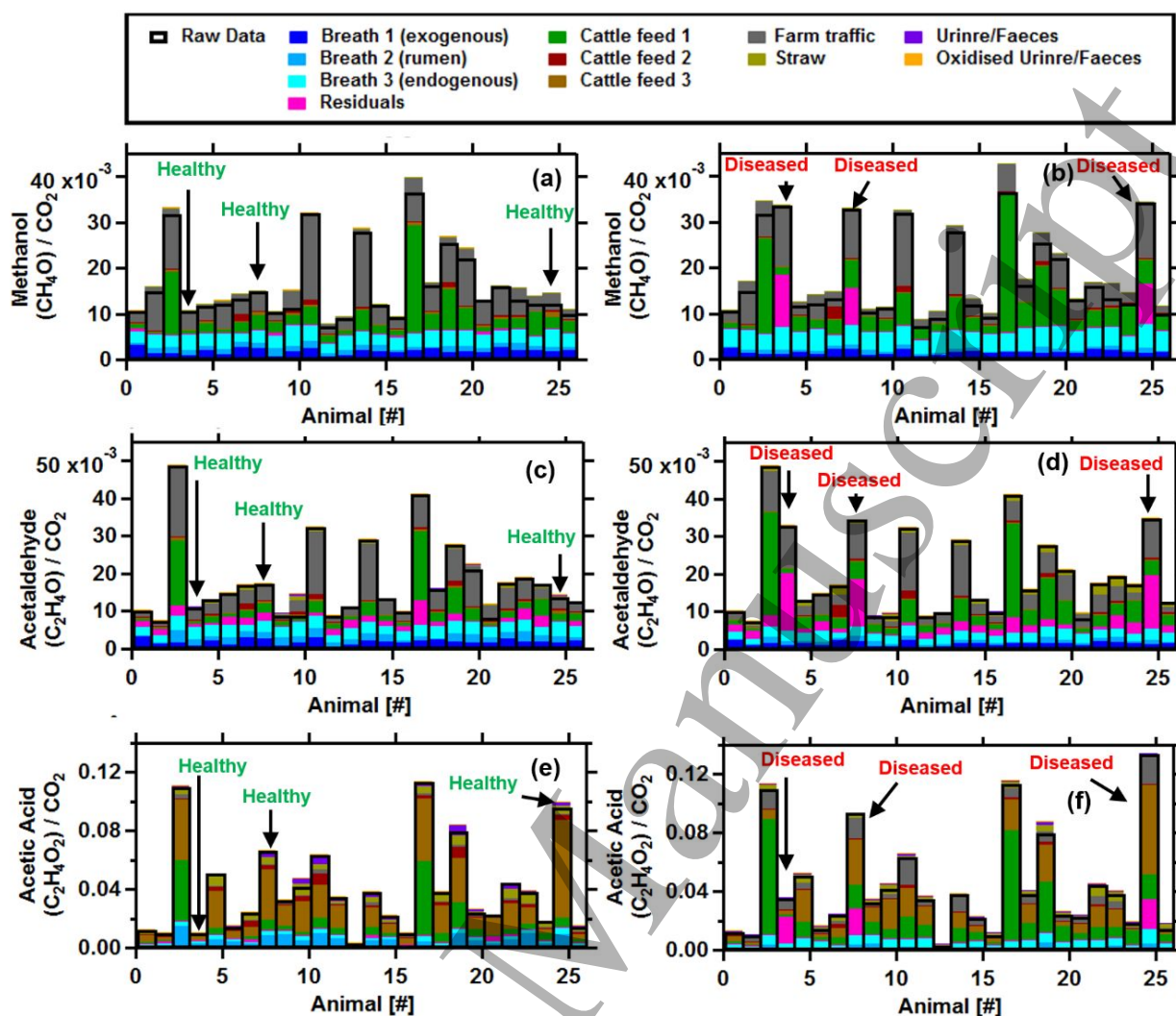


Figure 8. Stacked bars comparing the median concentrations of methanol (a), acetaldehyde (c) and acetic acid (e) measured from within the BMP for the 26 individual animals. Each total has had a running minima baseline and interpolated background subtracted before being normalised by CO_2 . The relative contribution from the 10 factors identified by PMF is shown as stacked bars, with the breath component highlighted in shades of blue. The pink bars represent the positive residuals which are the portion of the measured VOC that could not be explained by the PMF algorithm. The negative residuals, where the PMF solutions is greater than the measured raw data, can be seen where the coloured bars exceed the solid black bars (raw data). Panel's b, d and f show the same result for methanol, acetaldehyde and acetic acid, respectively, where animals #4, #8 and #25 were modified to simulate disease. For these cattle, breath factor 3 was increased by a factor of 3.9 before re-running the PMF algorithm. VOC and CO_2 concentrations were in ncps and therefore ratios are arbitrary.

4.3 Potential for PMF to directly detect diseased animals

Despite the addition of a simulated respiratory illness, the PMF analysis did not yield a specific factor that could be associated with the visits of the three animals with simulated disease to the BMP, even when extending the solution to a maximum of 20 factors. This is potentially because the endogenous breath component of the three compounds chosen was

1
2
3 relatively minor compared to the cattle shed background, therefore, a ~4 x increase in the
4 breath component is actually very small relative to the total measured signal.

5 A second PMF analysis was performed with the addition of a fourth marker compound.
6 The additional ion, $(\text{CH}_2\text{S})\text{H}^+$, was chosen because, although less abundant than the other
7 markers, endogenous breath represented a much larger fraction of its total compared to that
8 of the other markers and $(\text{CH}_2\text{S})\text{H}^+$ was almost exclusively emitted on the breath of cattle (see
9 Figure 7 and Figure S14 on the SI). Figure 9 (a) shows the time series of a factor identified by
10 PMF where concentrations clearly increase during visits from each of the diseased animals
11 (shaded red area) but not in the case of the healthy animals (shaded grey area). This factor
12 was only identified once the solution was increased to 16 factors. The mass spectral profile of
13 this factor is shown in Figure S15. Rather than being dominated by contributions from the four
14 marker compounds, the three most abundant ions were ammonia, CO_2 , and $(\text{C}_2\text{H}_2\text{O})\text{H}^+$
15 (ketone or ethenone). Acetic acid, methanol, acetaldehyde and $(\text{CH}_2\text{S})\text{H}^+$ were the next most
16 abundant ions, respectively. This discrepancy between the modified ions and those found in
17 the diseased factor profile again highlights that while the PMF analysis was able to produce a
18 solution that could identify the periods when a “diseased cow” visited the BMP, it was unable
19 to perfectly replicate the exact suite of ions that were modified.

20 An additional PMF analysis was conducted using three marker compounds, this time
21 replacing acetaldehyde with $(\text{CH}_2\text{S})\text{H}^+$. With this combination, the PMF analysis detected a
22 very similar “diseased animal” factor but the solution had to be increased to 17 factors before
23 it appeared. With slightly less information to work with the solution was not as robust, with
24 some instances of false positives, where spikes in the “diseased animal” factor were present
25 during the visits of healthy animals to the BMP. The time series is shown in Figure 9 (b).

26 Further PMF solutions were carried out iteratively to determine the lower limit of
27 detection for the “diseased animal” factor. For these compounds and this set of measurements
28 this limit was found to be a 3.5 multiplication of the endogenous breath fraction, but this will
29 undoubtedly vary depending on the number of markers, their magnitude and the relative
30 contribution of the background to the measured marker total. Similarly, here we have adjusted
31 each marker equally, whereas in reality different markers would be up or down regulated with
32 varying magnitude and may change as the disease progresses. In addition, the exact degree
33 of up-regulation of each marker will differ between diseased animals, whereas here it was kept
34 constant across the three diseased cattle. It is possible that in a real world example the
35 analysis would pick out several factors reflecting the range of expressions of the disease on
36 emissions with any combination of these factors for various animals. Future work should now
37 look to identify markers of respiratory disease in cattle and assess whether these can be
38 directly detected using the approach outlined here, or with other methods such as
39 randomForests which are more suited to binary classification problems.

40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

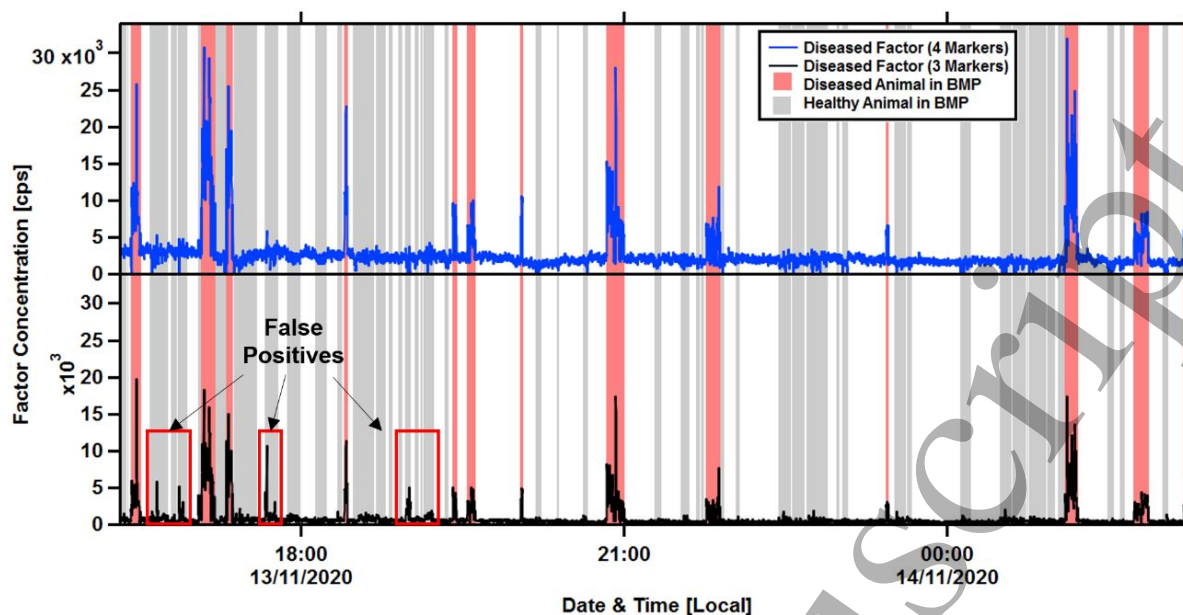


Figure 9. Example time series of a “diseased animal” factor identified by PMF. Shaded green areas represent the time periods where a healthy animal visited the BMP and the red areas represent visits from diseased animals. Panel (a) shows the factor detected on the basis of simulated disease with four biomarkers (methanol, acetaldehyde, CH₂S and acetic acid) and panel (b) shows the detection of a similar factor based on three biomarkers (methanol, CH₂S and acetic acid).

4.4 Limitations of the breath monitoring platform

The breath monitor platform allows for the automated capture of breath samples from individual cattle without the need for handling. We have shown how the collected breath samples can be carefully processed to limit the contributions of non-breath artefacts and how this approach might be used to highlight differences in breath composition from within the monitored herd. This approach is a significant advancement on previous studies that have used single point measurements to assess the general health of a herd (17) rather than individual animals, offering potential for passive breath monitoring to aid in the detection of respiratory diseases in livestock. However, this method also has limitations. For example, our analysis has shown that samples acquired from the BMP and analysed by PMF will never perfectly replicate the breath composition acquired through the use of a respiratory mask. In particular, using respiratory masks allows for a targeted analysis of a particular phase of the exhaled breath, such as the alveolar fraction, where more specific diagnostic information may be located. Accessing such specific information is plainly not possible using the BMP. A further consideration is the fact that the application of PMF to the passively acquired data introduces a degree of subjectivity, with the number of factors chosen in the solution down to the individual processing the data. This means that while comparison between animals from within that set of measurements is possible, comparisons with other data sets and other methods, such as respiratory masks, are likely to be less meaningful.

In this study we use a PTR-QiTOF for the quantification of VOCs, but this instrument is only able to detect certain compounds, for example those with a proton affinity greater than water (when in H₃O⁺ ion mode). Furthermore, the instrument is limited to molecular formula identification of measured ions which means absolute identification of compounds is not possible.

We have shown that the PMF algorithm is, in principle, capable of directly detecting instances of a simulated disease in a herd of cattle, albeit under highly idealised conditions.

1
2
3 However, its success at doing so was (at least in this simulation) related to the magnitude of
4 the changes of the biomarkers relative to the total measured signal, which is often dominated
5 by background odours from within the cattle shed. The sensitivity of the approach could,
6 therefore, be improved by taking steps to reduce this background. Future efforts should see if
7 improvements can be made by placing the BMP outside of the cattle shed where background
8 concentrations are typically orders of magnitude lower due to the natural ventilation of the
9 system by the wind. Where this is not possible, the system could be continually flushed with
10 air from outside the cattle shed or using air that has first past through a charcoal filter to reduce
11 the background concentrations and ensure the breath concentrations dominate the sampled
12 air rather than other odours.
13

14 Finally, the successful separation of breath samples from the cattle shed background
15 requires a sufficient amount of data to be collected for the PMF algorithm to work with e.g. 1
16 day of measurements. This means animals with an abnormal breath composition are unlikely
17 to be flagged at the point of measurement, but only once the full measurement period has
18 been analysed and reviewed.
19
20
21
22

23 5. Conclusion

24
25 Online breath analysis has become an important clinical tool for the diagnosis of
26 numerous respiratory and metabolic illnesses in humans, but its application to livestock is
27 relatively unexplored due to the practical difficulties of sampling from animals. We have shown
28 how the diagnostic potential seen in humans might be extended to livestock using a passive
29 breath monitoring platform (BMP), where exhaled breath accumulates under a hood while an
30 animal accesses the water trough. The volatile organic compounds (VOCs) emitted into this
31 air space were analysed using a proton transfer reaction time-of-flight mass spectrometer
32 (PTR-QiTOF), with a total of 106 ions that could be assigned a molecular formula. Passive
33 monitoring of the breath in this way avoids the requirement for handling livestock and allows
34 the breath composition of individuals to be tracked over time, but comes at the cost of
35 controlled acquisition of the sample, with the breath unavoidably mixed with the complex blend
36 of compounds present within the cattle shed. Our study highlights how the breath sample can
37 be retrospectively separated from the cattle shed background through the application of
38 positive matrix factorisation (PMF). In this study, three breath related components were
39 identified following the removal of the background factors, thought to reflect the average
40 endogenous breath, exogenous breath (breath after feeding) and emissions from the rumen.
41 While the endogenous breath fraction showed no discernible pattern, the exogenous and in
42 particular the rumen fraction exhibited a strong diurnal cycle with breath concentrations largest
43 during the daytime. The background factors were dominated by emissions from the cattle feed,
44 farm traffic and urine /faeces. The time series of each background factor was highly variable
45 and, with the exception of CO₂ and acetone, most of the compounds were present at
46 concentrations far higher than in breath making it impossible to isolate the breath fraction
47 using standard background removal.
48
49
50
51

52 Following the removal of the cattle shed background using PMF, the three identified
53 breath factors were summed together with the residuals to give an average breath sample for
54 each of the 26 animals. The inclusion of the residuals was found to be vital, because each
55 breath factor in isolation represents the average composition of the monitored herd.
56 Introducing an artificial disease signal to three of the healthy cattle, by upregulating several
57 compounds found in the endogenous breath component of each of these animals, we found
58 that the addition resulted in only a modest increase in the endogenous breath component of
59 all cattle, with the upregulation instead captured in the residuals of the three modified cattle.
60

The potential for PMF to identify and remove artefacts from passive breath samples has been demonstrated in this short pilot study, but future work should now explore how the requirement for signal analysis might be reduced. This could be achieved by taking steps to lower background interferences, either by placing the BMP outside of the main cattle shed to increase natural ventilation of the platform, or by flushing the hooded area with VOC free air between visits. Reducing the background relative to the breath sample will increase the sensitivity of the approach.

Finally, our short pilot study worked with 26 cattle free from respiratory illness. Going forward, the method should focus on whether factor analysis techniques can directly identify individuals actually suffering from respiratory disease and, importantly, whether subclinical incidence of disease can be identified. A key step going forward will be to identify the potential compounds indicative of disease so the approach can become more targeted and where possible utilise low-cost electrochemical sensors. The later step will be vital if the BMP is to offer a practical solution for detecting respiratory and metabolic illnesses in cattle and other livestock.

Ethical Statement

This study was reviewed by SRUC's Animal Welfare and Ethical Review Body (AWERB) as required by the UK Home Office. Approval was granted (Approval ID: CAT AE 29-2020). Animals were housed as required under UK Home Office guidelines in respect to stocking density, feeder space allowance etc.

Acknowledgements

We would like to thank Lesley Purcell and Claire Broadbent for management and health scoring the animals throughout the trial and Neil Mullinger for helping to setup the mobile laboratory. SRUC receives core funding from RESAS (Rural and Environment Science and Analytical Services, Scottish Government). We thank the DelhiFlux project (NE/P016502/1) for funding the PTR-QiTOF. We thank two anonymous reviewers for their helpful comments on the manuscript.

References

1. Bank W (2010) *People, Pathogens, and Our Planet : Volume One - Towards a One Health Approach for Controlling Zoonotic Diseases*.
2. Currin J.F WD (2009) Recognition and treatment of bovine respiratory disease complex. in *Virginia Cooperative Extension Publication* (Virginia Cooperative Extension Publication).
3. Sivula NJ, Ames TR, Marsh WE, & Werdin RE (1996) Descriptive epidemiology of morbidity and mortality in Minnesota dairy heifer calves. *Preventive Veterinary Medicine* 27(3):155-171.
4. Schaefer AL, et al. (2007) The use of infrared thermography as an early indicator of bovine respiratory disease complex in calves. *Research in Veterinary Science* 83(3):376-384.
5. Nickell JS & White BJ (2010) Metaphylactic Antimicrobial Therapy for Bovine Respiratory Disease in Stocker and Feedlot Cattle. *Veterinary Clinics: Food Animal Practice* 26(2):285-301.
6. Hateley G (2018) *Mycoplasma bovis*: current knowledge, industry challenges and gaps for the UK cattle industry. (SRUC, Animal & Plant Health Agency), pp 1-5.
7. Glover ID (2017) Improving calf-side diagnosis of respiratory disease on UK dairy farms. *Livestock* 22(5):230-238.

8. Timsit E, Assié S, Quiniou R, Seegers H, & Bareille N (2011) Early detection of bovine respiratory disease in young bulls using reticulo-rumen temperature boluses. *The Veterinary Journal* 190(1):136-142.
9. Rose-Dye TK, et al. (2011) Rumen temperature change monitored with remote rumen temperature boluses after challenges with bovine viral diarrhoea virus and Mannheimia haemolytica. *Journal of Animal Science* 89(4):1193-1200.
10. Mahendran SA, Booth R, Burge M, & Bell NJ (2017) Randomised positive control trial of NSAID and antimicrobial treatment for calf fever caused by pneumonia. *Veterinary Record* 181(2):45.
11. Duthie CA, et al. (2021) Feeding behaviour and activity as early indicators of disease in pre-weaned dairy calves. *Animal : an international journal of animal bioscience* 15(3):100150.
12. Peled N, et al. (2012) Detection of volatile organic compounds in cattle naturally infected with Mycobacterium bovis. *Sensors and Actuators B: Chemical* 171-172:588-594.
13. Ophelders D, et al. (2021) Screening of Chorioamnionitis Using Volatile Organic Compound Detection in Exhaled Breath: A Pre-clinical Proof of Concept Study. *Frontiers in pediatrics* 9:617906.
14. Traxler S, et al. (2018) VOC breath profile in spontaneously breathing awake swine during Influenza A infection. *Scientific Reports* 8(1):14857.
15. Purkhart R, et al. (2011) Chronic intestinal Mycobacteria infection: discrimination via VOC analysis in exhaled breath and headspace of feces using differential ion mobility spectrometry. *Journal of breath research* 5(2):027103.
16. Maurer DL, Koziel JA, Engelken TJ, Cooper VL, & Funk JL (2018) Detection of Volatile Compounds Emitted from Nasal Secretions and Serum: Towards Non-Invasive Identification of Diseased Cattle Biomarkers. *Separations* 5(1):18.
17. Gierschner P, et al. (2019) Crowd monitoring in dairy cattle—real-time VOC profiling by direct mass spectrometry. *Journal of Breath Research* 13(4):046006.
18. Jordan A, et al. (2009) A high resolution and high sensitivity proton-transfer-reaction time-of-flight mass spectrometer (PTR-TOF-MS). *Int. J. Mass Spectrom.* 286(2-3):122-128.
19. Inomata S, et al. (2008) Technical Note: Determination of formaldehyde mixing ratios in air with PTR-MS: laboratory experiments and field measurements. *Atmos. Chem. Phys.* 8(2):273-284.
20. Smolinska A, et al. (2014) Current breathomics—a review on data pre-processing techniques and machine learning in metabolomics breath analysis. *Journal of Breath Research* 8(2):027105.
21. Warneke C, De Gouw JA, Kuster WC, Goldan PD, & Fall R (2003) Validation of atmospheric VOC measurements by proton-transfer-reaction mass spectrometry using a gas-chromatographic pre-separation method. *Environmental science & technology* 37(11):2494-2501.
22. Paatero P & Tapper U (1994) Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics* 5(2):111-126.
23. Allan JD, et al. (2003) Quantitative sampling using an Aerodyne aerosol mass spectrometer - 1. Techniques of data interpretation and error analysis. *Journal of Geophysical Research-Atmospheres* 108(D3).
24. Ulbrich IM, Canagaratna MR, Zhang Q, Worsnop DR, & Jimenez JL (2009) Interpretation of organic components from Positive Matrix Factorization of aerosol mass spectrometric data. *Atmos. Chem. Phys.* 9(9):2891-2918.
25. Oertel P, et al. (2018) Continuous real-time breath analysis in ruminants: effect of eructation on exhaled VOC profiles. *J Breath Res* 12(3):036014.
26. Erickson MH, Gueneron M, & Jobson BT (2014) Measuring long chain alkanes in diesel engine exhaust by thermal desorption PTR-MS. *Atmos. Meas. Tech.* 7(1):225-239.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
27. Jobson BT, Alexander ML, Maupin GD, & Muntean GG (2005) On-line analysis of organic compounds in diesel exhaust using a proton transfer reaction mass spectrometer (PTR-MS). *Int. J. Mass Spectrom.* 245(1-3):78-89.
 28. Sintermann J, *et al.* (2014) Trimethylamine emissions in animal husbandry. *Biogeosciences* 11(18):5073-5085.
 29. Ellis CK, *et al.* (2014) A Pilot Study Exploring the Use of Breath Analysis to Differentiate Healthy Cattle from Cattle Experimentally Infected with *Mycobacterium bovis*. *PLOS ONE* 9(2):e89280.