

Hybrid Model-Driven Spectroscopic Network for Rapid Retrieval of Turbine Exhaust Temperature

Yalei Fu, *Student Member, IEEE*, Rui Zhang, Jiangnan Xia, Andrew Gough, Stuart Clark, Abhishek Upadhyay, Godwin Enemali, Ian Armstrong, Ihab Ahmed, Mohamed Pourkashanian, Paul Wright, Krikor Ozanyan, *Senior Member, IEEE*, Michael Lengden, Walter Johnstone, Nick Polydorides, Hugh McCann, Chang Liu, *Senior Member, IEEE*

Abstract—Exhaust Gas Temperature (EGT) is a key parameter in diagnosing the health of gas turbine engines (GTEs). In this paper, we propose a model-driven spectroscopic network with strong generalizability to monitor the EGT rapidly and accurately. The proposed network relies on data obtained from a well-proven temperature measurement technique, i.e., wavelength modulation spectroscopy, with the novelty of introducing underlying physical absorption model and building a hybrid dataset from simulation and experiment. This hybrid model-driven network enables strong noise resistance of the neural network against real-world experimental data. The proposed network is assessed by *in situ* measurements of EGT on an aero-GTE at millisecond-level temporal response. Experimental results indicate that the proposed network substantially outperforms previous neural-network methods in terms of accuracy and precision of the measured EGT when the GTE is steadily loaded.

Index Terms—deep neural network, signal processing, gas turbine engine, exhaust gas temperature, wavelength modulation spectroscopy

I. INTRODUCTION

GAS Turbine Engines (GTEs) provide efficient energy solutions for many industrial sectors, such as electrical power generation and propulsion systems, particularly in aviation. As an indicator of the thermal health condition of a GTE, Exhaust Gas Temperature (EGT) is critically important and must be monitored with good temporal resolution [1, 2]. Although thermocouples (TCs) have been employed for EGT monitoring, their slow response is inadequate for the assessment of the instantaneous energy conversion process, which strongly relates to the combustion characteristics and control efficiency of the GTE. In contrast, optical diagnostics offer non-intrusiveness and much faster responses and thus facilitate timely engine fault diagnosis [3, 4].

Among the various optical diagnostic tools, Wavelength Modulation Spectroscopy (WMS) is unique for robust, rapid, accurate and sensitive temperature measurement in harsh

environments [5, 6]. WMS is implemented by extracting harmonics of the laser transmission signal, which are then fitted using a spectral database, typically using an algorithm based on a least-squares metric, to retrieve the mean temperature along the laser path [7]. The fast wavelength-scanning capabilities of available laser sources, combined with the excellent temporal response characteristics of photodetectors [8] and associated electronics [9-11], mean that WMS can be carried out at kHz-level temporal resolution [12], and recently up to 1 MHz [13]. However, the iteration-based spectral fitting process limits the real-time response of WMS in industrial applications to the second or sub-second level, even using modern high-level processors. Therefore, acceleration of the WMS signal processing stage is a critical requirement to enable its application on rapid EGT retrieval.

Deep neural networks (DNNs) have been utilized for rapid signal processing in absorption spectroscopy [14-17]. Some attempts rely on training the DNN using simulated datasets established by the spectral database. For example, Liu *et al* [14] used this approach to establish a back-propagation neural network to retrieve isotopic CH₄ abundance with enhanced sensitivity. Tian *et al* [15] introduced a pre-trained DNN for gas concentration retrieval using a simulated dataset and employed transfer learning to tackle the shortage of experimental data. Since the above neural networks are mainly trained with simulated data, the result strongly depends on the match between the trained model and the physical circumstances when acquiring the real-world data. However, such models rely heavily on the dataset and thus have poor generalizability for more unpredictable circumstances, such as GTE exhaust.

Recent efforts have been made on integrating physical models and experimental data into the training process to improve the generalizability and robustness of DNNs. For example, Xie *et al* [16] integrated the absorption model into a neural network via outputting the temperature-dependent line-

This work was supported in part by the UK Engineering and Physical Sciences Research Council under Grants EP/T012595/1 and EP/P001661/1. (*Corresponding author: Chang Liu.*)

Y. Fu, R. Zhang, J. Xia, G. Enemali, N. Polydorides, H. McCann and C. Liu are with the School of Engineering, University of Edinburgh, Edinburgh EH9 3JL, U.K. (e-mail: C.Liu@ed.ac.uk).

A. Gough, S. Clark, A. Upadhyay, I. Armstrong, M. Lengden and W. Johnstone are with Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow G1 1XQ, U.K.

I. Ahmed and M. Pourkashanian are with Department of Mechanical Engineering, University of Sheffield, Sheffield S10 2TN, U.K.

P. Wright and K. Ozanyan are with School of Electrical and Electronic Engineering, University of Manchester, Manchester M13 9PL, U.K.

shape function with ultra-low spectral sampling. Sun *et al* [17] introduced a physically-motivated recurrent neural network (RNN), in which de-noised spectra were incorporated into the training process. These state-of-the-art algorithms can effectively obtain features from experimental absorption spectra to calculate gas parameters. However, the experimental data used in the datasets were sampled under well-controlled lab conditions. In harsh industrial environments, for example, the test cell of a GTE, these algorithms suffer from significant uncertainties in their performance against real-world measurement data, in which unpredictable noise arises from mechanical vibration and optical interference. Moreover, industrial environment can also make the laser performance, e.g., linearity and wavelength stability, different than that in the lab, and thus make the lab-based neural network no longer effective.

To address the aforementioned limitations, this paper proposes a new hybrid model-driven neural network, named HMD-WMS. The proposed algorithm is evaluated against TC measurements in the retrieval of EGT behind an Auxiliary Power Unit (APU), a GTE that provides energy for purposes other than propulsion on large aircraft. The main novelties of this paper are:

- 1) A new HMD-WMS network is designed for real-time spectra recovery and EGT monitoring. It combines a transformer-based encoder and a Convolutional Neural Network (CNN)-structured decoder. Such an architecture efficiently captures temporal and shape information of the WMS harmonic spectra, de-noises spectral sequences and enables efficient neural network testing.
- 2) The developed HMD-WMS, for the first time, integrates both the physical model of absorption spectroscopy and industrially-relevant experimental WMS data. The hybrid model enables strong noise resistance against real-world experimental data.
- 3) An experiment is carried out on an APU under steady-state load to examine the performance of the developed HMD-WMS on retrieving both the spectra and the EGT at 250 Hz. The retrieved spectra fit well with the least-square metric, while the EGT agrees with the TC measurements in its temporal and spatial coverage.

The remainder of this paper is organized as follows: In Section II, the WMS principles and the detailed architecture of the HMD-WMS are briefly presented. Section III illustrates the experiment setup and dataset construction. Section IV presents the experimental results and validation of the HMD-WMS. Finally, conclusions are presented in section V.

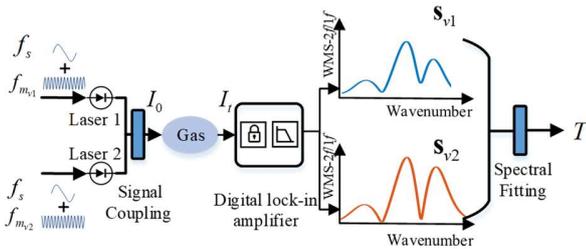


Fig. 1. Flowchart of WMS measurement and the spectral fitting for temperature retrieval.

II. METHODS

A. Fundamentals of WMS

In our case, frequency-division multiplexing (FDM) is applied in WMS by simultaneously measuring two absorption transitions of the same chemical species to calculate EGT via ratio thermometry [18, 19]. As shown in Fig. 1, two lasers at the central wavenumbers of $\bar{\nu}_1$ [cm⁻¹] and $\bar{\nu}_2$ [cm⁻¹] are driven by sinusoidal signals at $f_{m_{v1}}$ [Hz] and $f_{m_{v2}}$ [Hz], respectively, superimposed on a low-frequency carrier at f_s [Hz]. In general, $f_{m_v} \geq 100 \times f_s$, where v is either v_1 or v_2 . The emitted laser intensity of either laser, $I_{0,v}$, can be expressed as

$$I_{0,v}(t) = \bar{I}_{0,v} [1 + i_{0,v} \cos(2\pi f_{m_v} t + \phi_{1,v}) + i_{2,v} \cos(4\pi f_{m_v} t + \phi_{2,v})] \quad (1)$$

where $\bar{I}_{0,v}$ refers to the time-average emitted laser intensity, $i_{0,v}$ is the amplitude of linear intensity modulation with linear phase shift $\phi_{1,v}$, and $i_{2,v}$ and $\phi_{2,v}$ are the second-harmonic intensity modulation amplitude and phase shift, respectively.

The two emitted laser signals are combined with a fiber coupler and collimated to penetrate the target plume with a path length of L_{abs} [cm]. The transmitted laser intensity $I_{t,v}$ is received by a photodetector and can be expressed as:

$$I_{t,v}(t) = I_{0,v}(t) \cdot \tau(v_v(t)) = I_{0,v}(t) \cdot \exp(-P \int_0^{L_{\text{abs}}} X \cdot S_v(T) \cdot \phi_v dl) \quad (2)$$

where τ is the absorbance, $v_v(t)$ the instantaneous laser wavenumber, P [atm] the pressure and X the mole fraction of targeted gas species. $S_v(T)$ [cm⁻²atm⁻¹] is the temperature-dependent line strength when the temperature is T [K]. ϕ_v is the line-shape function and can be integrated against wavenumber to unity, i.e., $\int_{-\infty}^{\infty} \phi_v dv = 1$.

Further, $I_{t,v}$ is demodulated by a digital lock-in amplifier (DLIA) at the frequencies of f_{m_v} and $2 \times f_{m_v}$, resulting in the extraction of the first- and second-order harmonics of $I_{t,v}(t)$, i.e., $1f$ and $2f$, respectively. The absorption spectrum at frequency $2f$, normalized to the absorption spectrum at $1f$, noted as WMS- $2f/1f$, can be expressed as [20]:

$$\text{WMS-}2f/1f = \frac{S_v(T) \cdot P \cdot X \cdot L_{\text{abs}}}{i_{0,v} \cdot \pi} \int_{-\pi}^{\pi} \phi_v(\bar{\nu}_v + a \cos \theta) \cos 2\theta d\theta \quad (3)$$

where a is the amplitude of modulation depth. The WMS- $2f/1f$ spectra of the two transitions are saved as two vectors, $\mathbf{s}_{v1} \in$

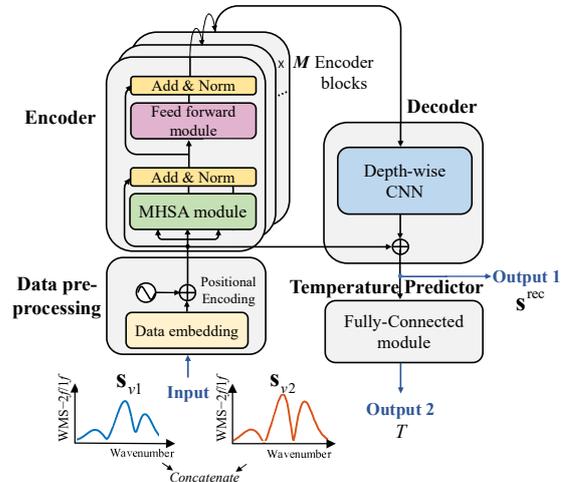


Fig. 2. Architecture of the proposed HMD-WMS.

$\mathbb{R}^{L \times 1}$ and $\mathbf{s}_{v2} \in \mathbb{R}^{L \times 1}$. L is the number of demodulated wavenumber samples of each transition. Traditionally, T will be retrieved after least-square fitting \mathbf{s}_{v1} and \mathbf{s}_{v2} to the spectroscopic models generated by the HITRAN database [21]. It generally costs a few hundred milliseconds to a few seconds for fitting a single transition even using modern processors, such as those with 2.5 GHz (or higher) base frequency and 8 MB cache. Such computational cost limits the real-time temporal resolution of WMS even if f_s is set at kHz or above. To solve this issue, HMD-WMS is designed in the next subsection to replace the least-square fitting by effectively and rapidly mapping both the temporal and shape features of \mathbf{s}_{v1} and \mathbf{s}_{v2} to T with high accuracy and generalizability.

B. HMD-WMS architecture

The architecture of the HMD-WMS is shown in Fig. 2. It contains four modules, i.e., a data pre-processing module, an encoder, a decoder, and a temperature predictor. Overall, the measured $\{\mathbf{s}_{v1}, \mathbf{s}_{v2}\}$ are concatenated as $\mathbf{s} \in \mathbb{R}^{2L \times 1}$, as the input of the data pre-processing module. Then, the pre-processed data go through a transformer-based encoder, followed by a CNN-structured decoder. These three modules aim to recover a de-noised WMS-2f/1f spectral sequence $\mathbf{s}^{\text{rec}} \in \mathbb{R}^{2L \times 1}$, which is physically constrained by the line-shape function. Finally, the temperature predictor maps \mathbf{s}^{rec} to T via fully-connected layers. The functionalities of these four modules are detailed below.

1. Data pre-processing

This module aims to extract data features to maximize the efficiency of the subsequent encoder. Fig. 3 shows a simulated WMS-2f/1f spectral sequence \mathbf{s} of water vapor (H_2O) absorption centred at $\bar{\nu}_1=7185.6 \text{ cm}^{-1}$ and $\bar{\nu}_2=7444.4 \text{ cm}^{-1}$. The primary features of such spectral sequences are manifested as a) the potential temporal correlation between neighboring wavenumber samples arising from time-dependent wavelength scanning; b) the shape features of the whole WMS-2f/1f spectra, such as the temperature-related absorption peaks, broadened lineshape and phase shift [22]. In this module, data embedding and positional encoding are introduced to pre-process \mathbf{s} and then output $\mathbf{P} \in \mathbb{R}^{2L \times d_{\text{emb}}}$ which retains the temporal and shape features. d_{emb} is the embedding dimension.

Data embedding will convert \mathbf{s} to a higher-dimensional matrix. Data embedding has been widely used in Natural Language Processing for converting the categorical variable to

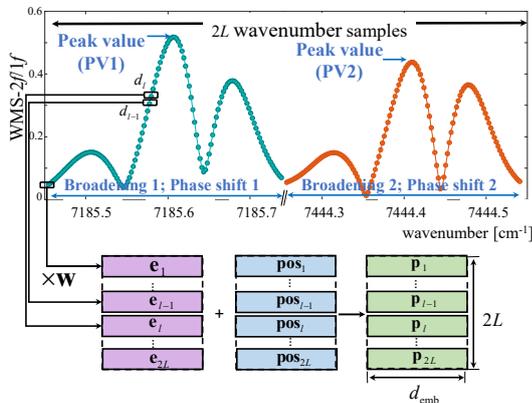


Fig. 3. Pre-process of concatenated $\{\mathbf{s}_{v1}, \mathbf{s}_{v2}\}$ via data embedding and positional encoding.

a vector of numerical values for semantic meaning representation [23]. In our case, data embedding is mainly used for maintaining the spectral information of each wavenumber sample \mathbf{s}_l ($l = 1, 2, \dots, 2L$) by an embedded vector $\mathbf{e}_l \in \mathbb{R}^{1 \times d_{\text{emb}}}$ for the following positional encoding, where \mathbf{e}_l is the l -th row vector of an embedded matrix $\mathbf{E} \in \mathbb{R}^{2L \times d_{\text{emb}}}$. \mathbf{E} is given by:

$$\mathbf{E} = \mathbf{s} \times \mathbf{w} \quad (4)$$

where $\mathbf{w} \in \mathbb{R}^{1 \times d_{\text{emb}}}$ is the trainable weight.

Positional encoding generates a unique code for the embedded matrix to preserve and enhance its position information [24]. The positional encoding matrix $\mathbf{Pos} \in \mathbb{R}^{2L \times d_{\text{emb}}}$ is obtained for \mathbf{E} . The element of the l -th row vector \mathbf{pos}_l can be expressed as:

$$\text{pos}_{l,2j} = \sin\left(\frac{l}{10000^{2j/d_{\text{emb}}}}\right) \quad (5)$$

$$\text{pos}_{l,2j+1} = \cos\left(\frac{l}{10000^{2j/d_{\text{emb}}}}\right) \quad (6)$$

where $j = 1, 2, \dots, d_{\text{emb}}/2$. d_{emb} is an even number. The use of sinusoidal functions here is due to a) their nature of continuity and differentiability to benefit model training; and b) their reliability of frequency adjustment to generate unique position-involved vectors. Finally, the data pre-processing module will output $\mathbf{P} \in \mathbb{R}^{2L \times d_{\text{emb}}}$:

$$\mathbf{P} = \mathbf{E} + \mathbf{Pos} \quad (7)$$

2. Encoder

A transformer benefitting from its effectiveness on learning long-term dependencies via parallel computation [24], is used in HMD-WMS encoder to capture the temporal features of \mathbf{P} . Here, M series of encoder blocks are concatenated, as shown in Fig. 2, to form an encoder. In each block, there are a multi-head self-attention (MHSA) sublayer and a feedforward sublayer. The principles of MHSA have been demonstrated in [24]. In our case, the structure of the encoder can be simplified as:

$$\mathbf{O}_{\text{sublayer1}} = \text{LayerNorm}(\text{MHSA}(\mathbf{P})) + \mathbf{P} \quad (8)$$

$$\mathbf{O}_{\text{sublayer2}} = \text{LayerNorm}(\sigma(\mathbf{O}_{\text{sublayer1}} \mathbf{W}_1 + \mathbf{b}_1) \mathbf{W}_2 + \mathbf{b}_2) + \mathbf{P} \quad (9)$$

where $\mathbf{O}_{\text{sublayer1}} \in \mathbb{R}^{2L \times d_{\text{emb}}}$ and $\mathbf{O}_{\text{sublayer2}} \in \mathbb{R}^{2L \times d_{\text{emb}}}$ are the outputs of MHSA and the feedforward sublayer, respectively. $\mathbf{W}_1 \in \mathbb{R}^{d_{\text{emb}} \times d_{\text{ff}}}$ and $\mathbf{W}_2 \in \mathbb{R}^{d_{\text{ff}} \times d_{\text{emb}}}$ represent the trainable weights for the first and the second linear layers within the feedforward sublayer. $\mathbf{b}_1 \in \mathbb{R}^{1 \times d_{\text{ff}}}$ and $\mathbf{b}_2 \in \mathbb{R}^{1 \times d_{\text{emb}}}$ are bias vectors. σ is the ReLU activation function.

3. Decoder

The decoder consists of a depth-wise CNN (DW-CNN) and a residual submodule. It further learns and refines the shape of the WMS-2f/1f spectra to recover the de-noised spectra.

1) Depth-wise Convolutional Neural Network

CNNs are well-suited for discovering intricate shape features of matrices. Here, we choose DW-CNN to prevent crosstalk among different channels by setting the number of kernels as $2L$ to avoid the distortion on the extracted temporal information from the sequence. In standard CNNs, each convolutional layer performs convolution across all input channels using a shared set of learnable kernels. In contrast, DW-CNN will employ an individual kernel set for each input channel, i.e., each

wavenumber sample in the spectral sequence, thus preserving the existing temporal features among different channels. In our model, DW-CNN is composed of two depth-wise convolutional layers, i.e., Conv1 and Conv2, each followed by a batch normalization layer and an activation function. $\mathbf{O}_{\text{sublayer2}}$ is reshaped to $\mathbf{C} \in \mathbb{R}^{\sqrt{d_{\text{emb}}} \times \sqrt{d_{\text{emb}}} \times 2L}$ as the input. The forward propagation is expressed as:

$$\mathbf{O}_{\text{Conv1}} = \sigma \left\{ \mathbf{BN}(\mathbf{W}_{\text{Conv1}} * \mathbf{C} + \mathbf{b}_{\text{Conv1}}) \right\} \quad (10)$$

$$\mathbf{O}_{\text{Conv2}} = \sigma \left\{ \mathbf{BN}(\mathbf{W}_{\text{Conv2}} * \mathbf{O}_{\text{Conv1}} + \mathbf{b}_{\text{Conv2}}) \right\} \quad (11)$$

$$\mathbf{o}^{\text{dec}} = \sigma \left\{ \mathbf{FC}(\mathbf{O}_{\text{Conv2}}) \right\} \quad (12)$$

where the operator $*$ represents convolution, \mathbf{BN} and \mathbf{FC} the batch normalization and the fully-connected calculation, respectively. $\mathbf{O}_{\text{Conv1}} \in \mathbb{R}^{H_{O1} \times W_{O1} \times C_{O1}}$ ($\mathbf{O}_{\text{Conv2}} \in \mathbb{R}^{H_{O2} \times W_{O2} \times C_{O2}}$), $\mathbf{W}_{\text{Conv1}} \in \mathbb{R}^{H_{W1} \times W_{W1} \times C_{W1}}$ ($\mathbf{W}_{\text{Conv2}} \in \mathbb{R}^{H_{W2} \times W_{W2} \times C_{W2}}$) and $\mathbf{b}_{\text{Conv1}} \in \mathbb{R}^{1 \times C_{O1}}$ ($\mathbf{b}_{\text{Conv2}} \in \mathbb{R}^{1 \times C_{O2}}$) are the output maps, convolutional kernels and bias vectors of Conv1 (Conv2), respectively. H_{O1} (H_{O2}), W_{O1} (W_{O2}), are the height and width of $\mathbf{O}_{\text{Conv1}}$ ($\mathbf{O}_{\text{Conv2}}$), respectively. C_{O1} (C_{O2}) is the number of channels. H_{W1} (H_{W2}), W_{W1} (W_{W2}), and C_{W1} (C_{W2}) are the kernel height, width, and the number of kernels of $\mathbf{W}_{\text{Conv1}}$ and $\mathbf{W}_{\text{Conv2}}$, respectively. $C_{O1} = C_{O2} = C_{W1} = C_{W2} = 2L$. $\mathbf{o}^{\text{dec}} \in \mathbb{R}^{2L \times 1}$ is the reconstructed feature vector from the DW-CNN module.

2) Residual module

This module tackles the issues of vanishing gradient and degradation in accuracy via short-cut connections when the depth of the neural network increases [25]. Here, the input \mathbf{s} will shortcut to the output of DW-CNN, \mathbf{o}^{dec} , for residual mapping. The output of the residual module is defined as:

$$\mathbf{s}^{\text{rec}} = \sigma \left\{ \mathbf{o}^{\text{dec}} + \mathbf{s} \right\} \quad (13)$$

where $\mathbf{s}^{\text{rec}} \in \mathbb{R}^{2L \times 1}$ is the reconstructed WMS-2f/1f spectral sequence, ‘+’ the operand of element-wise addition.

4. Temperature Predictor

The Temperature Predictor retrieves the mean path temperature T^{rec} based on \mathbf{s}^{rec} . According to Eq. (3), the dependence of \mathbf{s}^{rec} on T^{rec} can be represented by the following mapping:

$$\mathbf{s}^{\text{rec}} = g(T^{\text{rec}}) \quad (14)$$

We construct three fully-connected layers TP1, TP2 and TP3 to train this mapping. Finally, T^{rec} can be expressed as:

$$T^{\text{rec}} = \tanh \left\{ \mathbf{w}_{\text{TP3}} \left\{ \sigma \left(\mathbf{W}_{\text{TP2}} \left\{ \sigma \left(\mathbf{W}_{\text{TP1}} \mathbf{s}^{\text{rec}} + \mathbf{b}_{\text{TP1}} \right) \right\} + \mathbf{b}_{\text{TP2}} \right) \right\} + \mathbf{b}_{\text{TP3}} \right\} \quad (15)$$

where $\mathbf{W}_{\text{TP1}} \in \mathbb{R}^{d_{\text{TP1}} \times 2L}$, $\mathbf{W}_{\text{TP2}} \in \mathbb{R}^{d_{\text{TP2}} \times d_{\text{TP1}}}$ and $\mathbf{w}_{\text{TP3}} \in \mathbb{R}^{1 \times d_{\text{TP2}}}$ are the trainable weights for the first, second and third layers, respectively, $\mathbf{b}_{\text{TP1}} \in \mathbb{R}^{1 \times d_{\text{TP1}}}$, $\mathbf{b}_{\text{TP2}} \in \mathbb{R}^{1 \times d_{\text{TP2}}}$ and \mathbf{b}_{TP3} are the bias of each layer, and ‘tanh’ is the activation function Tanh. Since \mathbf{s}^{rec} is noise-free, T^{rec} retrieval using \mathbf{s}^{rec} gives more stable results compared to those obtained from the noise-contaminated raw \mathbf{s} .

III. EXPERIMENT AND RESULTS

A. Experiment setup

To assess the performance of the proposed HMD-WMS method, an industrial-scale experiment was carried out by measuring the EGT of an aircraft APU (Honeywell, Model: GTC85-129) mounted in a test cell at the University of

Sheffield, UK. Fig. 4 shows the APU exit, laser sensors and the layout of the optical path. In this experiment, we selected H_2O as the target gas species, as it is a common combustion product and has rich absorption spectra in the near infrared. Here, H_2O transitions at $\bar{\nu}_1=7185.6 \text{ cm}^{-1}$ and $\bar{\nu}_2=7444.4 \text{ cm}^{-1}$ are used for ratio thermometry due to their good sensitivity in the target temperature range [3]. The two FDM laser diodes at $\bar{\nu}_1$ (NTT, NLK1E5GAAA) and $\bar{\nu}_2$ (NTT, NLK1B5EAAA) were both scanned at $f_s=1 \text{ kHz}$, and modulated at $f_{m_{\nu_1}}=100 \text{ kHz}$ and $f_{m_{\nu_2}}=130 \text{ kHz}$, respectively. The transmitted laser signal was received by a photodetector (Hamamatsu, G12182-110K) and then digitized by a customized DAQ system [9] at $f_{\text{samp}}=15.625 \text{ Mega samples/second}$. The DLIA is integrated into the DAQ. It utilizes a cascaded integrator-comb (CIC) filter [26][27] to down-sample the raw transmission signals with a decimator of $c=32$. As a result, each of the $1f$ and $2f$ signals for each transition has $N_{\text{samp}}=488$ wavenumber samples, where $N_{\text{samp}}=f_{\text{samp}}/(f_s \cdot c)$. Here, we use the central 120 wavenumber samples for each absorption transition, i.e., $L=120$, that adequately cover all the features of the absorption spectra as the input to the HMD-WMS, i.e., $\mathbf{s}_{\nu_1} \in \mathbb{R}^{120 \times 1}$, $\mathbf{s}_{\nu_2} \in \mathbb{R}^{120 \times 1}$. The total laser path length L_{abs} is 50 cm, whereas the estimated path length L_{plu} through the APU exhaust plume is 24 cm. To enable operation of the optics under ambient conditions, both the laser emitter and detector are located at a distance along the radius of 13 cm (denoted L_{amb}) outside the edge of the APU exit (Fig. 4). In downstream direction, the laser beam path is positioned 3 cm from the APU exit. For line-of-sight WMS measurement, ratio thermometry retrieves mean path temperature $\bar{T} = f(A_{\nu_1, \text{abs}}, A_{\nu_2, \text{abs}})$, where $A_{\nu, \text{abs}} = P \cdot S_{\nu}(\bar{T}) \cdot \bar{X} \cdot L_{\text{abs}}$ is the integrated absorption from the transition ν and $f(\cdot)$ represents the mapping between the integrated absorption of the two transitions and the temperature. Transmitted laser signals were also collected when the APU was not in operation in order to calculate the ambient temperature T_{amb} and H_2O concentration X_{amb} via the traditional fitting method. Then, the integrated absorption along the ambient path length L_{amb} is calculated as $A_{\nu, \text{amb}} = P \cdot S_{\nu}(T_{\text{amb}}) \cdot X_{\text{amb}} \cdot L_{\text{amb}}$, leaving that in the plume as $A_{\nu, \text{plu}} = A_{\nu, \text{abs}} - 2 \times A_{\nu, \text{amb}}$. Finally, the EGT is calculated by $T_{\text{plu}} = f(A_{\nu_1, \text{plu}}, A_{\nu_2, \text{plu}})$. As the standard instrument on the APU, a Class 2 K-type TC (IEC 60584-2, OMEGA Engineering Ltd, UK) was used to record T_{plu} . As demonstrated in [28], the plume temperature is nearly uniform and consistent with the TC measurement. Therefore, the TC-measured T_{plu} can be regarded as a reference for the mean path temperature measurement.

B. Dataset construction

As shown in Fig. 5, a hybrid dataset was built by combining both simulated data and a small portion of experimental data. Among different APU working conditions, the measured

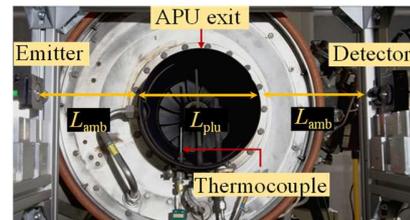


Fig. 4. Auxiliary Power Unit and laser system setup

temperature range by TC was from 400 K to 880 K, whereas the mole fraction measured by a gas sampler was 0.00052 to 0.02. To improve the diversity of training samples, we generated a simulated training dataset utilizing the HITRAN database, allowing simulated average temperature and H₂O concentration to cover the above range with the step size of 4 K for temperature and 0.0002 for H₂O mole fraction. As a result, 14400 simulated WMS-2f/1f spectral sequences $\mathbf{s}_m^{\text{sim}} = \{\mathbf{s}_{v_1,m}^{\text{sim}}, \mathbf{s}_{v_2,m}^{\text{sim}}\} \in \mathbb{R}^{240 \times 1}$, $m = 1, 2, \dots, 14400$, were generated. To further improve noise robustness, 20 dB, 30 dB and 40 dB white noises (the reasonable noise levels in real experiments [5, 9]) were added to $\mathbf{s}_m^{\text{sim}}$, resulting in 57600 noise-contaminated simulated spectra $\mathbf{s}_m^{\text{sim}}$. Each $\mathbf{s}_m^{\text{sim}}$ spectrum is associated with its corresponding reconstruction $\mathbf{s}_m^{\text{rec,sim}}$, i.e., the output of the decoder, and with the ground-truth temperature T_m^{sim} , to be a simulated training sample.

For the experimental training dataset, we randomly selected 100 measured spectra for each absorption feature when the APU was working at steady-state, noted as $\mathbf{s}_n^{\text{exp}} = \{\mathbf{s}_{v_1,n}^{\text{exp}}, \mathbf{s}_{v_2,n}^{\text{exp}}\} \in \mathbb{R}^{240 \times 1}$, $n = 1, 2, \dots, 100$. The traditional least-squares fitting method is used to obtain a de-noised spectral sequence $\mathbf{s}_n^{\text{rec,exp}}$ and thus retrieve T_n^{exp} . The combination of $\mathbf{s}_n^{\text{exp}}$, $\mathbf{s}_n^{\text{rec,exp}}$ and T_n^{exp} are regarded as the experimental training samples. Therefore, there are $\mathcal{N}_{\text{train}}$ training samples in total, denoted as $((\mathbf{s}_n^{\text{sim}}, \mathbf{s}_n^{\text{exp}}), (\mathbf{s}_n^{\text{rec,sim}}, \mathbf{s}_n^{\text{rec,exp}}), (T_n^{\text{sim}}, T_n^{\text{exp}}))$, $\mathcal{N}_{\text{train}} = 57700$.

Using the above training dataset, empirical risk minimization (ERM) and structural risk minimization (SRM) are deployed on

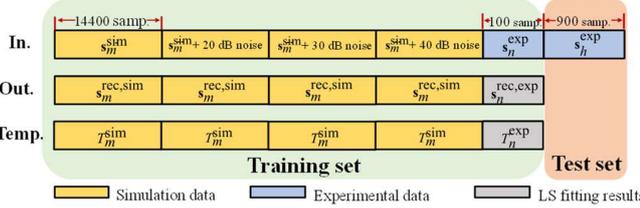


Fig. 5. Construction of the training set and the test set.

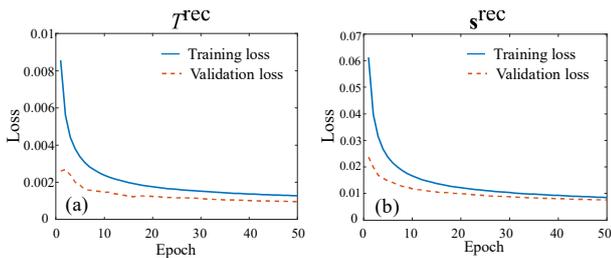


Fig. 6. Training loss and validation loss for (a) temperature prediction and (b) WMS-2f/1f spectra retrieval.

the HMD-WMS for model training. Specifically, ERM is applied to minimize the weighted mean-square-error (MSE) loss in terms of the reconstructed spectral sequence and temperature, whereas SRM is used to avoid overfitting. Given the training dataset with a batch size of B , the total loss function $L_{\text{tot,loss}}$ is defined as:

$$L_{\text{tot,loss}} = \alpha \times \frac{1}{B} \sum_{b=1}^B (\widehat{\mathbf{s}}_b^{\text{rec}} - \mathbf{s}_b^{\text{tru}})^2 + (1-\alpha) \times \frac{1}{B} \sum_{b=1}^B (\widehat{T}_b^{\text{rec}} - T_b^{\text{tru}})^2 + \lambda \|\psi\|_2^2 \quad (16)$$

where α is the hyperparameter to trade off MSE loss between the b -th reconstructed spectral sequence $\widehat{\mathbf{s}}_b^{\text{rec}}$ and temperature $\widehat{T}_b^{\text{rec}}$, $\mathbf{s}_b^{\text{tru}} \in (\mathbf{s}_m^{\text{rec,sim}}, \mathbf{s}_n^{\text{rec,exp}})$ and $T_b^{\text{tru}} \in (T_m^{\text{sim}}, T_n^{\text{exp}})$ are the corresponding ground-truth values, λ is the penalty parameter for L2 regularization $\|\cdot\|_2^2$. ψ is the set of trainable weights. The other hyper-parameters in HMD-WMS are shown in Table I.

The ablation study is carried out when designing the HMD-WMS to examine the robustness of the proposed network architecture. To further assess the HMD-WMS, the traditional fitting method and the TC measurement are regarded as benchmarks. The results obtained by the HMD-WMS are compared to the benchmarks, and to two other state-of-the-art data-driven methods that utilize the same dataset and are pre-processed in the same way. The details of the three alternative methods are given below:

- Traditional fitting method [7]: a least-squares fit to the measured WMS-2f/1f spectra is carried out using gas parameters in the HITRAN database, i.e., Doppler/collision broadening, line strength, etc. to calculate the integrated absorption for each transition. The mean path temperature is retrieved using ratio thermometry.
- Adaptively Optimized Gas Analysis Model (AOGAM) [17]: attention-aided long short-term memory (LSTM) algorithm is used as its framework to reconstruct gas parameters via absorption spectroscopy. Here, WMS-2f/1f spectra are input to the AOGAM and flow through the same data pre-processing module, an attention-aided LSTM-structured encoder and decoder to retrieve the de-noised spectral sequence and thus the temperature.
- Convolutional autoencoder (ConV-AE): this method is widely recognized in the field of absorption spectroscopy, i.e., Y-Net [29]. Another reason to use this framework for temperature retrieval is to compare the pure ConV-AE to the HMD-WMS. Here, the WMS-2f/1f spectra are fed into a DW-CNN for retrieval.

All the data-driven models were trained using PyTorch on a single GPU of NVIDIA A100-SXM4-40GB and deployed Adam [30] for optimization with a learning rate of 2×10^{-3} . $\alpha = 0.5$. $B = 32$. $\lambda = 2 \times 10^{-6}$. Epoch is 50.

TABLE I
HYPER-PARAMETERS OF HMD-WMS

| Encoder | | Decoder | | | | Temperature Predictor | | | |
|------------------|-----|--------------------|-------------------------|-------------------------|----------------|-----------------------|------------------|-----------------|---------------|
| d_{emb} | 64 | | Conv1 | Conv2 | FC | | TP1 | TP2 | TP3 |
| θ | 8 | Input dim. | $8 \times 8 \times 240$ | $5 \times 5 \times 240$ | 240×4 | Input dim. | 240 | 128 | 32 |
| M | 4 | Weight matrix size | 4×4 | 4×4 | 4×1 | Weight matrix size | 240×128 | 128×32 | 32×1 |
| d_{ff} | 128 | Stride | (1,1) | (1,1) | – | | | | |
| d_Q, d_K, d_V | 8 | Output dim. | $5 \times 5 \times 240$ | $2 \times 2 \times 240$ | 240×1 | Output dim. | 128 | 32 | 1 |

IV. EXPERIMENTAL VALIDATION

When the APU worked at steady-state, 900 s_h^{exp} ($h=1, 2, \dots, 900$) were measured for testing, i.e., $\mathcal{N}_{\text{test}} = 900$. To accelerate convergence of the model, $(s_n^{\text{sim}}, s_n^{\text{exp}})$ and $(s_m^{\text{rec, sim}}, s_n^{\text{rec, exp}})$ are standardized, whereas the temperature is rescaled within the range of (0,1) using the low temperature T_{low} of 300 K and the high temperature T_{high} of 900 K. To examine the proposed HMD-WMS's learning trajectory and its adaptability, Fig. 6 (a) shows both training loss and validation loss for temperature prediction and (b) for spectra retrieval in the training process. It can be seen that the optimization of the HMD-WMS in both cases converges rapidly and steadily. The three data-driven models HMD-WMS, AOGAM and ConV-AE are tested by comparing their outputs in terms of s^{rec} and $T_{\text{plu}}^{\text{rec}}$ against values obtained by the traditional least-squares fitting procedure and the TC data. $T_{\text{plu}}^{\text{rec}}$ represents the retrieved plume temperature.

A. Evaluation of s^{rec}

Fig. 7 shows one example of the reconstructed WMS-2f/1f spectra (at $\bar{\nu}_1$ and $\bar{\nu}_2$) s_h^{rec} ($h = 900$) measured behind the APU, obtained by deploying the least-squares spectral fitting method, i.e. the benchmark method, and shown as the solid curve in each of Fig. 7 (a), (b), (c); those obtained by the three different data-

driven algorithms are given in Fig. 7 (a), (b), and (c) and compared separately against the benchmark.

It can be seen that s_h^{rec} retrieved using the proposed HMD-WMS in Fig. 7 (a) best fits the benchmark. s_h^{rec} retrieved by AOGAM in Fig. 7 (b) is smooth but deviates significantly from the fitting result around the peak of the absorbance. The retrieval using ConV-AE in Fig. 7 (c) is similarly close to the benchmark at the wavenumber around 7185.6 cm^{-1} , in comparison to that using the HMD-WMS. However, ConV-AE results in a much larger residual to the benchmark at the wavenumber around 7444.3 cm^{-1} . Furthermore, the Euclidean norms of the residual for HMD-WMS, AOGAM and ConV-AE are 0.2647, 0.6211 and 0.4485, respectively. This confirms that HMD-WMS gives the best accuracy when compared to the other two deep learning models.

Furthermore, the performance on s_h^{rec} retrieval for the whole experimental test set (i.e., 900 pairs of spectra) is analyzed statistically. We used Curve Error (CE), Correlation Coefficient (CC) and Peak Values (PVs) as metrics to evaluate the models' performance. CE is defined as:

$$\text{CE} = \frac{1}{\mathcal{N}_{\text{test}}} \sum_{h=1}^{\mathcal{N}_{\text{test}}} \frac{\|s_h^{\text{rec}} - s_h^{\text{tru}}\|_2}{\|s_h^{\text{tru}}\|_2} \quad (17)$$

where s_h^{tru} is the vector of WMS-2f/1f spectra recovered using the spectral fitting method. As shown in Table II, the CE of the

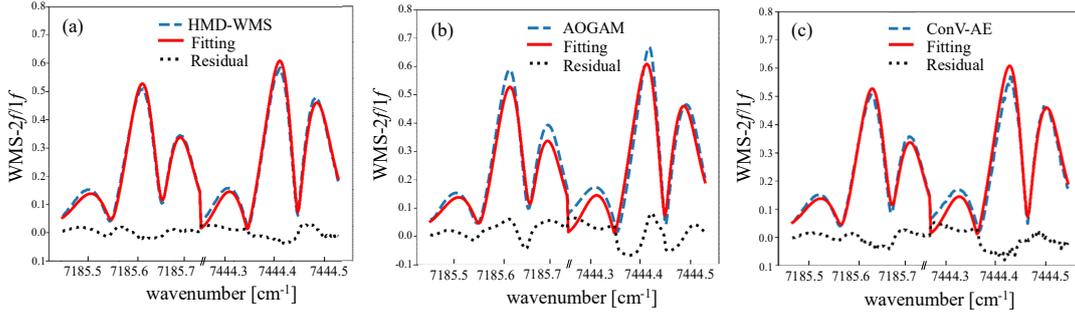


Fig. 7. Comparison of the reconstructed WMS-2f/1f spectra using (a) HMD-WMS (b) AOGAM and (c) ConV-AE to the benchmark obtained from the traditional least-squares spectral fitting method.

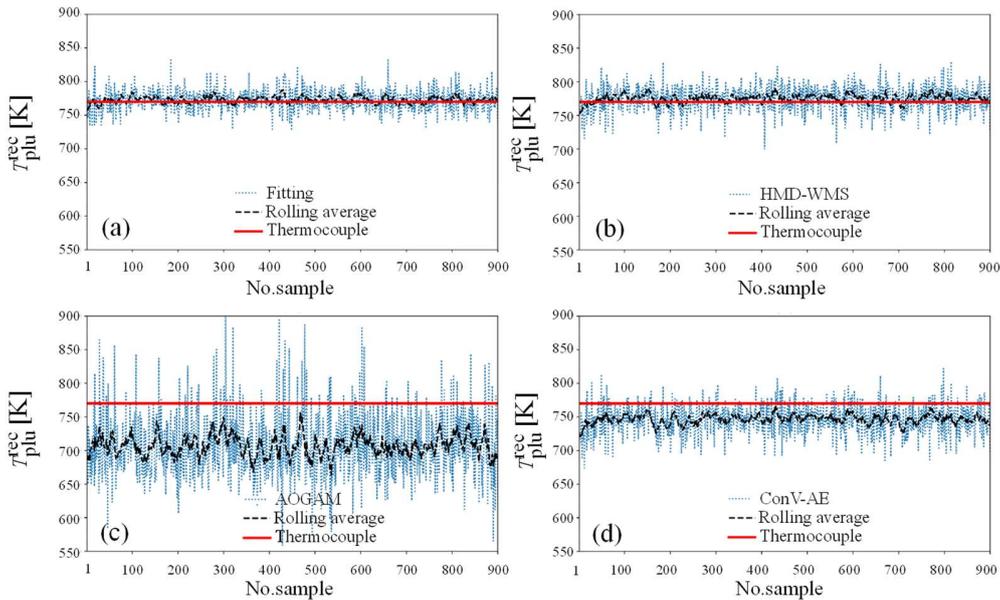


Fig. 8. Reconstructed EGT in the plume $T_{\text{plu}}^{\text{rec}}$, using (a) traditional spectral fitting, (b) HMD-WMS, (c) AOGAM, (d) ConV-AE and their rolling averages compared to the TC measurement.

HMD-WMS is 0.0464, approximately $\frac{1}{4}$ of AOGAM and $\frac{1}{2}$ of ConV-AE. CC represents the similarity between s^{rec} retrieved using the deep learning models and the benchmark. Although CCs obtained from different models are all higher than 0.96, HMD-WMS gives the highest, at 0.9971. The 0.01-0.03 higher CC compared to the AOGAM and ConV-AE is reflected by the improved accuracy on the absorption peaks and the smoothness of the curve. PV mainly represents the local characteristics at the absorption peak of the two WMS-2f/1f spectra. The two peak values around $\bar{\nu}_1$ and $\bar{\nu}_2$ on s^{rec} are named PV1 and PV2, respectively (shown in Fig. 3). Both PV1 and PV2 obtained using HMD-WMS are closest to the result given by the benchmark. Therefore, it can be quantitatively and statistically concluded that HMD-WMS outperforms the other two state-of-the-art models for reconstructing the WMS-2f/1f spectra with better accuracy.

B. Evaluation of $T_{\text{plu}}^{\text{rec}}$

Fig. 8 shows the comparison of the reconstructed EGT in the plume $T_{\text{plu}}^{\text{rec}}$ during the 900 ms data collection period. The least-squares fitting method, HMD-WMS and ConV-AE show a similar extent of short-term temperature fluctuation, while AOGAM suffers from much stronger fluctuations. Regarding the TC measurement as the benchmark in this comparison, a moving average with a window length of 10 on 900 test samples is imposed on the raw $T_{\text{plu}}^{\text{rec}}$ to clearly visualize the offset between the reconstructed temperature and the TC measurement. It can be observed that HMD-WMS gives the smallest offset. To quantitatively evaluate the offset between the least-squares fitting method (the benchmark) and the deep learning algorithms, the Bland-Altman plot is shown in Fig. 9. M_{HW} , M_{CV} and M_{AG} represent the means of temperature offsets for the 900 test samples between the traditional fitting method and HMD-WMS, ConV-AE and AOGAM, respectively. SD_{HW} , SD_{CV} and SD_{AG} are the corresponding standard deviation of the offsets. As shown in Fig. 9, M_{HW} is 2.25 K, which is the smallest compared to those obtained using AOGAM (65.97 K) and ConV-AE (26.48 K). In addition, the results from the HMD-WMS exhibit the narrowest variation and are closest to zero, falling within the range of $\pm 1.96 SD_{\text{HW}}$. Therefore, it can be concluded that HMD-WMS gives the best accuracy on EGT retrieval, among the three deep learning methods.

Table III shows a statistical comparison among the different methods using the same test set. Here, the mean value of the 900 measurements is used to assess the accuracy of each algorithm and the standard deviation (STD) allows examination of their robustness against measurement noise. The Standard Error (SE) of the mean value is used to compare accuracy among all the TDLAS algorithms. Although the APU was working at steady-state, a high-speed plume passing through the laser beam can cause beam steering and thus distortion on the measurement, so there is likely to be some contribution to the STD from variability in the measured gas sample. In addition, the tolerance of the TC is 0.75%, resulting in ± 5.8 K potential offset of the EGT, as measured by the TC, from the true value.

Furthermore, Table III also shows that both the traditional fitting method and the proposed HMD-WMS algorithm recover values of EGT that are consistent with that measured by the TC, and they have very similar dispersion among their individual

measured values, as indicated by their STD values, which are both small, yielding small SE. Adding the two SE values in quadrature indicates that the apparent discrepancy between the two mean values, 2.27K, is not statistically significant, hence their relative performance in recovering EGT is indistinguishable. Both AOGAM and ConV-AE fail to yield EGT values that are consistent with the TC result or with the traditional fitting method, even after consideration of the larger dispersion that they both display in their individual measured values.

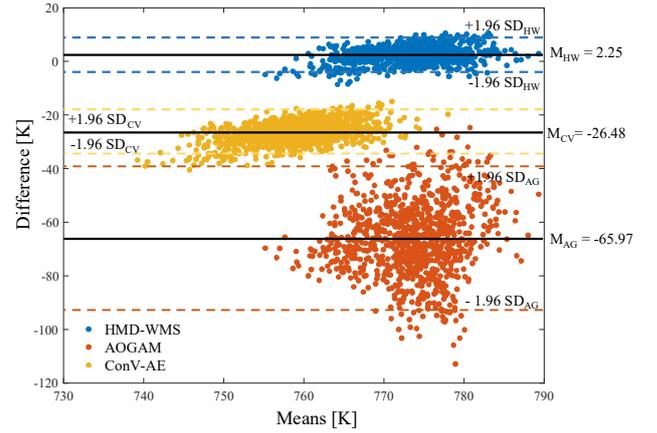


Fig. 9. The Bland-Altman plot to evaluate $T_{\text{plu}}^{\text{rec}}$ from HMD-WMS, AOGAM and ConV-AE in comparison to the benchmark: least-squares fitting method.

TABLE II
PERFORMANCES OF HMD-WMS, AOGAM AND CONV-AE ON WMS-2F/1F RETRIEVAL AND THEIR COMPARISONS TO THE SPECTRAL FITTING METHOD

| Method | | Fitting | HMD-WMS | AOGAM | ConV-AE |
|------------------|-----|---------|---------|--------|---------|
| s^{rec} | CE | -- | 0.0464 | 0.1613 | 0.0968 |
| | CC | -- | 0.9971 | 0.9685 | 0.9871 |
| | PV1 | 0.608 | 0.582 | 0.6450 | 0.540 |
| | PV2 | 0.528 | 0.507 | 0.589 | 0.504 |

TABLE III
MEAN, STD AND SE OF RETRIEVED EGT FOR THE TEST SET USING THE SPECTRAL FITTING, HMD-WMS, AOGAM AND CONV-AE, AND THEIR COMPARISON TO THE TC MEASUREMENT

| Method | TC | Fitting | HMD-WMS | AOGAM | ConV-AE |
|----------|-------------------|---------|---------|--------|---------|
| Mean (K) | 775 (± 5.8) | 772.32 | 774.59 | 706.36 | 745.82 |
| STD (K) | -- | 15.50 | 18.45 | 48.39 | 22.00 |
| SE (K) | -- | 0.52 | 0.62 | 1.61 | 0.73 |

TABLE IV
COMPARISON ON THE COMPUTATIONAL COMPLEXITY FOR THE TEST SET USING THE SPECTRAL FITTING, HMD-WMS, AOGAM AND CONV-AE

| Method | Fitting | HMD-WMS | AOGAM | ConV-AE |
|-----------------------|----------|---------|--------|-----------|
| Time elapsed (second) | 4.23×900 | 0.104 | 0.233 | 0.016 |
| No. parameters | -- | 78,575 | 68,642 | 2,128,243 |
| GFLOPs | -- | 1.77 | 0.040 | 0.7 |

C. Evaluation of computational cost

Table IV shows the computational inference time for processing the 900 test samples, the number of trainable parameters and Giga Floating-Point Operations per Second (GFLOPs) of these different algorithms. Least-squares fitting takes around 4.23 s for each spectra, resulting in a total of 3807 s for the whole test set, on a computer with the processor Intel(R) Core (TM) i7-6700 CPU @ 3.40GHz. In contrast, all the deep learning algorithms take less than 0.25 s for processing 900 spectral pair samples, i.e., less than 0.3 ms per sample. Although HMD-WMS gives higher GFLOPs, its inference time is shorter than AOGAM. This is benefitted from parallel computation of transformer in HMD-WMS. Once the data-driven model is trained, such inference time enables real-time EGT measurement at kHz temporal resolution or better. Besides, HMD-WMS has comparable trainable parameters with AOGAM, whereas ConV-AE has nearly three times more. Although ConV-AE achieves the shortest processing time, it has the worst estimation accuracy. Therefore, HMD-WMS is the best performer among all four methods presented here.

V. CONCLUSION

A hybrid model-driven WMS, named HMD-WMS, is proposed to enable accurate and rapid monitoring of the EGT of gas turbines. The model is constructed by the combination of data pre-processing, encoder, decoder and temperature predictor. To improve the fidelity of temperature retrieval, the underlying physical absorption spectroscopic model is introduced into the deep learning model and a hybrid dataset is built by including data from both simulation and experiment.

The proposed HMD-WMS is experimentally assessed by retrieving the EGT of an aircraft APU working at steady-state load. In comparison to the other two state-of-the-art deep learning models considered here, i.e., AOGAM and ConV-AE, HMD-WMS improves the accuracy in the reconstructed spectral sequence by 25%-50% under practical noise levels ranging from 20 dB to 40 dB. The new HMD-WMS method yields EGT values that are consistent with TC measurements and with the traditional least-squares fitting method, whereas AOGAM and ConV-AE fail in terms of the error on EGT estimation. The inference time of HMD-WMS for testing 900 samples is 0.104 s, enabling EGT retrieval above the kHz rate, with the potential for real-time gas turbine diagnosis.

REFERENCES

- [1] A. J. Volponi, "Gas Turbine Engine Health Management: Past, Present, and Future Trends," *J. Eng. Gas Turbines Power*, vol. 136, no. 5, 2014.
- [2] M. Prauzek *et al.*, "An Optical-Based Sensor for Automotive Exhaust Gas Temperature Measurement," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1-11, 2022.
- [3] A. Ehn, J. Zhu, X. Li, and J. Kiefer, "Advanced laser-based techniques for gas-phase diagnostics in combustion and aerospace engineering," *Appl. Spectrosc.*, vol. 71, no. 3, pp. 341-366, 2017.
- [4] F. J. Dutz, S. Boje, U. Orth, A. W. Koch, and J. Roths, "High-temperature profile monitoring in gas turbine exhaust-gas diffusers with six-point fiber-optic sensor array," *Int. J. Turbomach.*, vol. 5, no. 4, p. 25, 2020.
- [5] A. Upadhyay *et al.*, "Tomographic imaging of carbon dioxide in the exhaust plume of large commercial aero-engines," *Appl. Opt.*, vol. 61, no. 28, pp. 8540-8552, 2022.
- [6] G. Stewart, W. Johnstone, J. R. Bain, K. Ruxton, and K. Duffin, "Recovery of absolute gas absorption line shapes using tunable diode laser spectroscopy with wavelength modulation—Part I: Theoretical analysis," *J. Lightwave Technol.*, vol. 29, no. 6, pp. 811-821, 2011.
- [7] C. S. Goldenstein, C. L. Strand, I. A. Schultz, K. Sun, J. B. Jeffries, and R. K. Hanson, "Fitting of calibration-free scanned-wavelength-modulation spectroscopy spectra for determination of gas properties and absorption lineshapes," *Appl. Opt.*, vol. 53, no. 3, pp. 356-367, 2014.
- [8] P. Wright, K. B. Ozanyan, S. J. Carey, and H. McCann, "Design of high-performance photodiode receivers for optical tomography," *IEEE Sens. J.*, vol. 5, no. 2, pp. 281-288, 2005.
- [9] G. Enemali, R. Zhang, H. McCann, and C. Liu, "Cost-effective quasi-parallel sensing instrumentation for industrial chemical species tomography," *IEEE Trans. Ind. Electron.*, vol. 69, no. 2, pp. 2107-2116, 2021.
- [10] A. Huang, Z. Cao, C. Wang, J. Wen, F. Lu, and L. Xu, "An FPGA-based on-chip neural network for TDLAS tomography in dynamic flames," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1-11, 2021.
- [11] E. M. Fisher *et al.*, "A custom, high-channel count data acquisition system for chemical species tomography of aero-jet engine exhaust plumes," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 2, pp. 549-558, 2019.
- [12] S. J. Cassady *et al.*, "Time-resolved, single-ended laser absorption thermometry and H₂O, CO₂, and CO speciation in a H₂/C₂H₄-fueled rotating detonation engine," *P. Combust. Inst.*, vol. 38, no. 1, pp. 1719-1727, 2021.
- [13] G. C. Mathews, M. G. Blaisdell, A. I. Lemcherfi, C. D. Slabaugh, and C. S. Goldenstein, "High-bandwidth absorption-spectroscopy measurements of temperature, pressure, CO, and H₂O in the annulus of a rotating detonation rocket engine," *Appl. Phys. B-Lasers O.*, vol. 127, no. 12, p. 165, 2021.
- [14] Z. Liu, C. Zheng, T. Zhang, Y. Zhang, Y. Wang, and F. K. Tittel, "High-precision methane isotopic abundance analysis using near-infrared absorption spectroscopy at 100 Torr," *Analyst*, vol. 146, no. 2, pp. 698-705, 2021.
- [15] L. Tian *et al.*, "Retrieval of gas concentrations in optical spectroscopy with deep learning," *Meas.*, p. 109739, 2021.
- [16] H. Xie, L. Xu, Y. Tan, G. Hou, and Z. Cao, "Ultra-Low Sampled and High Precision TDLAS Thermometry Via Artificial Neural Network," *IEEE Photonics J.*, vol. 13, no. 3, pp. 1-9, 2021.
- [17] J. Sun *et al.*, "Adaptively Optimized Gas Analysis Model with Deep Learning for Near-Infrared Methane Sensors," *Anal. Chem.*, 2022.
- [18] C. S. Goldenstein, C. A. Almodóvar, J. B. Jeffries, R. K. Hanson, and C. M. Brophy, "High-bandwidth scanned-wavelength-modulation spectroscopy sensors for temperature and H₂O in a rotating detonation engine," *Meas. Sci. Technol.*, vol. 25, p. 105104, 2014.
- [19] A. Huang, Z. Cao, W. Zhao, H. Zhang, and L. Xu, "Frequency-division multiplexing and main peak scanning WMS method for TDLAS tomography in flame monitoring," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 11, pp. 9087-9096, 2020.
- [20] G. B. Rieker, J. B. Jeffries, and R. K. Hanson, "Calibration-free wavelength-modulation spectroscopy for measurements of gas temperature and concentration in harsh environments," *Appl. Opt.*, vol. 48, no. 29, pp. 5546-5560, 2009.
- [21] I. Gordon *et al.*, "The HITRAN2020 molecular spectroscopic database," *J. Quant. Spectrosc. Radiat. Transf.*, vol. 277, p. 107949, 2022.
- [22] R. K. Hanson, R. M. Spearrin, and C. S. Goldenstein, "Spectroscopy and optical diagnostics for gases," Springer International Publishing, 2016.
- [23] F. Almeida and G. Xexéo, "Word embeddings: A survey," *arXiv preprint arXiv:1901.09069*, 2019.
- [24] A. Vaswani *et al.*, "Attention is all you need," *Adv. Neural Inf. Processing Syst.*, vol. 30, 2017.
- [25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 770-778, 2016.
- [26] G. Li, E. Dong, and W.-h. Ji, "A Near-Infrared Trace CO₂ Detection System Based on an 1,580 nm Tunable Diode Laser Using a Cascaded Integrator Comb (CIC) Filter-Assisted Wavelength Modulation Technique and a Digital Lock-in Amplifier," *Front. Phys.*, vol. 7, p. 199, 2019.

- [27] J. Xia *et al.*, "FPGA-Accelerated Distributed Sensing System for Real-Time Industrial Laser Absorption Spectroscopy Tomography at Kilo-Hertz," *IEEE Trans. Ind. Inform.*, 2023. (Online Published. DOI: 10.1109/TII.2023.3292971)
- [28] R. Zhang *et al.*, "A fast sensor for non-intrusive measurement of concentration and temperature in turbine exhaust," *Sens. Actuators B Chem.*, p. 134500, 2023. (Online Published. DOI: 10.1016/j.snb.2023.134500)
- [29] Z. Wang, N. Zhu, W. Wang, and X. Chao, "Y-Net: a dual-branch deep learning network for nonlinear absorption tomography with wavelength modulation spectroscopy," *Opt. Express*, vol. 30, no. 2, pp. 2156-2172, 2022.
- [30] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, 2014.