Aerosol and Air Quality Research

ORIGINAL RESEARCH
https://doi.org/10.4209/aaqr.220444

Assessment of Malaysia-wide $PM_{2.5}$ Forecasts from a Global Model

Zhixian Tan[1*], Mohd Talib Latif[2], Matthew J. Ashfold[1]

[1] School of Environmental and Geographical Sciences, University of Nottingham Malaysia, 43500 Semenyih, Selangor, Malaysia
[2] Department of Earth Sciences and Environment, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

ABSTRACT

Airborne particulate matter with an aerodynamic diameter of less than 2.5 µm ($PM_{2.5}$) is a major air pollutant worldwide. In Malaysia, transboundary 'haze' episodes with elevated $PM_{2.5}$ concentrations linked to fires are common, causing health and economic harms. To reduce impacts, forecasting $PM_{2.5}$ can enable effective $PM_{2.5}$ management and decision-making. Until now, $PM_{2.5}$ forecasts via a global mechanistic chemical transport model (CTM) have not been evaluated in the setting of Malaysia, where operational $PM_{2.5}$ forecasting systems for preventive warnings are not yet deployed. Hence, this study aims to evaluate the performance of $PM_{2.5}$ forecasts produced by a global CTM and to assess their suitability for use nation-wide in Malaysia. We used the surface $PM_{2.5}$ forecasts from the Copernicus Atmosphere Monitoring Service's (CAMS) global atmospheric composition forecast dataset (CAMS-GACF) and evaluated them against hourly $PM_{2.5}$ observations recorded throughout Malaysia from 2018 to 2020 via exceedance and accuracy analyses. We found that cycle 46r1 CAMS-GACF performance in Malaysia was generally weaker (critical success index (CSI) = 31%, $R^2 = 0.36$) than reported in other studies (CSI = 20–54%, $R^2 = 0.32–0.79$) focused on other countries, across multiple metrics in both analyses. We found CAMS-GACF did not accurately capture local-scale spatiotemporal variations in $PM_{2.5}$ spatially and diurnally. However, we found CAMS-GACF captured better the increased regional $PM_{2.5}$ pollution during the transboundary 'haze' episode of 2019. Based on our findings, we also propose recommendations on integrating CAMS-GACF in early-warning systems in Malaysia and on improving forecasts via bias-correction.

Keywords: CAMS, IFS, Early warnings, Haze episodes, Ground-level air quality monitoring

OPEN ACCESS

Received: December 8, 2022
Revised: April 13, 2023
Accepted: June 23, 2023

*** Corresponding Author:**
tanzhixian1215@gmail.com;
matthew.ashfold@nottingham.edu.my

Publisher:
Taiwan Association for Aerosol Research
ISSN: 1680-8584 print
ISSN: 2071-1409 online

1 INTRODUCTION

Airborne particulate matter (PM) is a major air pollutant. It has natural (e.g., wind-blown dust, sea-salt) and anthropogenic sources (e.g., combustion, forest fires) (Amil *et al.*, 2016; Amin Jaafar *et al.*, 2018; Ooi *et al.*, 2015; Roberts and Wooster, 2021; Suradi *et al.*, 2021). Globally, PM pollution reduces up to 10 million years of life expectancy every year, affects regional water availability, and contributes to climate change (Lelieveld *et al.*, 2019). Southeast Asia is no exception to these, with frequent biomass burning and subsequent PM pollution throughout the whole region (Adam *et al.*, 2021). In Malaysia, haze episodes (mostly transboundary) with heightened PM mass concentrations are regular, occurring almost annually in the last decade (Department of Environment Malaysia (DOE), 2022). These episodes have severe health and economic consequences (Amil *et al.*, 2016; Phung *et al.*, 2022; Sahani *et al.*, 2014): the 2013 Southeast Asian Haze alone cost Malaysia MYR 410 million in hospitalisation bills, medical leaves, and personal protective equipment (PPE), and up to MYR 1 billion more to lost income opportunities (Manan *et al.*, 2018). It is known that PM with an aerodynamic diameter of less than 2.5 µm ($PM_{2.5}$) dominates the mix during these episodes (Adam *et al.*, 2021; Kusumaningtyas and Aldrian, 2016). The U.S. Environmental Protection Agency (U.S. EPA, 2022) marked $PM_{2.5}$ as "the greatest health risks", higher than the coarser PM, because it is respirable and can readily enter the bloodstream, affecting the respiratory and cardiovascular systems (Manan *et al.*, 2018). Hence, $PM_{2.5}$ is not just the major air pollutant in Malaysia, it is also very detrimental to Malaysia's health and economy.

One way to reduce the impacts of $PM_{2.5}$ pollution is through $PM_{2.5}$ forecasting. $PM_{2.5}$ early-warning systems have been implemented in many countries and cities worldwide, often using forecasts derived from mechanistic, computer-driven chemical transport models (CTMs) (Casallas *et al.*, 2020; Celis *et al.*, 2022; Cho *et al.*, 2021; Roux *et al.*, 2020; Savage *et al.*, 2013; Varga-Balogh *et al.*, 2020). Forecasting should aid preparation for bad air quality in advance and early decision-making to reduce exposure and improve resilience via personal and institutional means, e.g., mandate PPEs, institute advanced quarantine orders, and implement dynamic abatement measures (e.g., abating traffic, industrial emission, fire) (Lyu *et al.*, 2017; Zhou *et al.*, 2010). To these ends, forecasting should, at minimum, be able to forecast elevated $PM_{2.5}$ events such as the Southeast Asian haze episodes. Since $PM_{2.5}$ can remain airborne long enough to be a regional and long-duration problem like the Southeast Asian hazes (Dahari *et al.*, 2019; Fujii *et al.*, 2016), CTMs for $PM_{2.5}$ forecasts commonly cover large spatiotemporal scales. For effective $PM_{2.5}$ management, forecasts should also be made 3 to 5 days in advance for decision-making to translate into actions (Lyu *et al.*, 2017). Therefore, a global CTM with a medium-range forecast is appropriate for forecasting $PM_{2.5}$.

However, the few air quality forecasting studies that focus on Malaysia are solely based on statistical and machine learning models (ML) applied over limited scale and resolution in space and time (Koo *et al.*, 2020; Lim *et al.*, 2008; Wong *et al.*, 2021). In fact, Malaysia currently only provides reactionary warnings based on current observed pollution levels, limiting the efficacy of pollution response and decisions (Wong *et al.*, 2021). There are currently no operational $PM_{2.5}$ forecasts for preventive warnings in Malaysia, or any in the literature utilising CTMs, much less evaluating its performance Malaysia-wide. Accordingly, in this study, we aim to evaluate the performance of $PM_{2.5}$ forecasts from a global mechanistic CTM, and to assess their suitability for use nationally in Malaysia. Specifically, we aim to investigate the difference in forecast performances between: (a) geographical regions, (b) non-haze and haze episodes, (c) forecast horizons, and (d) model versions, and to provide implications and recommendations when using a global CTM for $PM_{2.5}$ forecasting in Malaysia.

2 METHODS

2.1 Haze Episodes in Malaysia

Malaysia is located within maritime Southeast Asia, consisting of Peninsular Malaysia and Malaysian Borneo separated by the South China Sea (Fig. 1). PM pollution in Malaysia is thought to be highly dependent on the geographical region and the monsoon seasonality, i.e., the southwest monsoon (SWM) occurring around July, northeast monsoon (NEM) around January, and the inter-monsoons between them (INM) (Juneng *et al.*, 2009). Haze episodes, informally defined as periods with impaired visibility and elevated $PM_{2.5}$ concentrations, occurs almost every year in the last two decades according to DOE (2020, 2019, 2018, 2017a, 2022). Among all haze episodes Malaysia experienced between 2010 and 2019, episodes in June 2013, September–October 2015, and August–September 2019 were particularly severe, affecting Malaysia nationally (see Supplementary Material 1 (SM1)). These episodes all occurred during the regionally drier SWM that increases risks of wildfires, and all are thought to be largely transboundary, sourced from Indonesian forest and peat fires (Reddington *et al.*, 2014; Tacconi, 2016; Zainal *et al.*, 2021). Numerous HYSPLIT backward trajectory analyses also revealed that air parcels usually travelled from Sumatra and Kalimantan to various locations in Malaysia within 1 to 4 days during the SWM hazes (Dahari *et al.*, 2019; Dotse *et al.*, 2016; Kusumaningtyas and Aldrian, 2016; Reddington *et al.*, 2014; Show and Chang, 2016; Zainal *et al.*, 2021). This further highlights the regional nature of extreme $PM_{2.5}$ pollution and the suitability of using a global forecast over several days to predict haze in Malaysia.

2.2 Model Details

We employed forecasts produced from Copernicus Atmosphere Monitoring Service's (CAMS) Integrated Forecasting System (IFS) as our $PM_{2.5}$ forecasts in our analyses. IFS is a global forecast

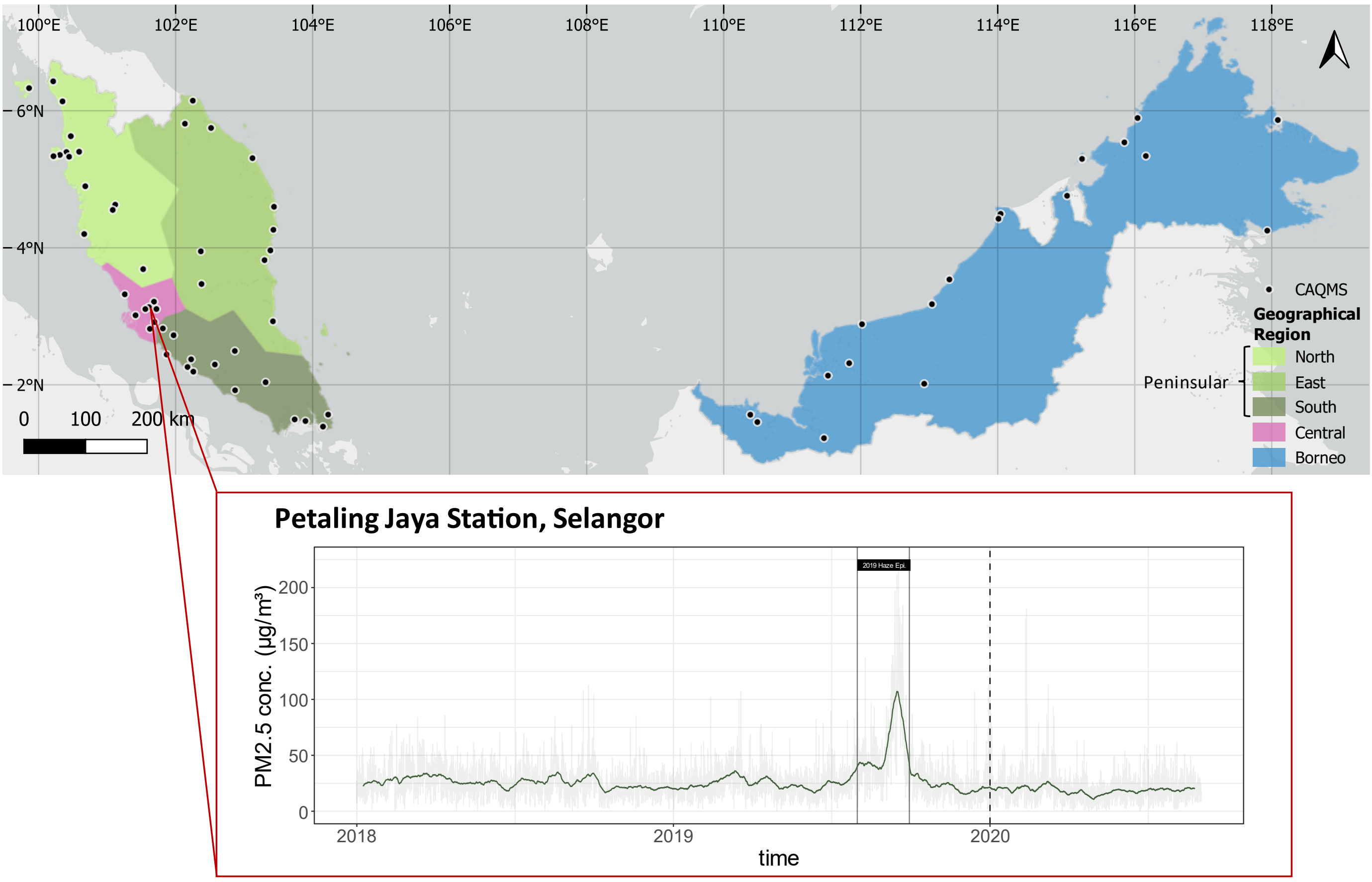

Fig. 1. Malaysia's air quality monitoring network (locations of 65 CAQMS and their geographical regions), and the observed $PM_{2.5}$ concentration at the Petaling Jaya Station (thin line – hourly data; thick line – 15-day running-average).

and assimilation system that was initially developed and used by the European Centre for Medium-range Weather Forecasts (ECMWF) solely for weather-forecasting, but extra modules were developed to also forecast atmospheric composition (henceforth known as CAMS-IFS) (ECMWF, 2022a). CAMS-IFS utilises a four-dimensional variational data assimilation (4D-Var) which combines meteorological and atmospheric composition observations with past forecasts to produce an initial state closer to reality (the analysis), improving the next sets of forecasts (Bannister, 2007; Benedetti *et al.*, 2009). CAMS-IFS may be suitable for forecasting $PM_{2.5}$ in Malaysia because it has a global extent, and the time-horizon is relevant for timely decision-making and for forecasting the SWM hazes.

The major module of interest within the CAMS-IFS is the IFS-AER. It models the aerosol components, including chemical transformations, transport, and deposition (Rémy *et al.*, 2019). The aerosol emissions are obtained from a combination of natural and anthropogenic emission inventories derived from pre-established inventories (Granier *et al.*, 2019). For example, the monthly anthropogenic emission inventory used (CAMS-GLOB-ANT) was derived by extrapolating EDGAR emissions using trends from CEDS, with approximately 10 km resolution (Crippa *et al.*, 2018; Hoesly *et al.*, 2018). Given our interest in fire-sourced haze, CAMS-IFS-AER also uses daily biomass burning emissions estimated by the Global Fire Assimilation System (GFAS) using real-time remotely sensed fires (ECMWF, 2022b, 2022a). The transport, chemical transformation, and deposition of emitted aerosols were then modelled according to their size and chemical characteristics (Rémy *et al.*, 2019). The resulting $PM_{2.5}$ concentrations are calculated based on the simulated concentrations of different aerosols and their sizes at that time step.

Near-term forecasts provided by the previous run are first constrained via satellite aerosol optical depth (AOD) observations over a 12-hour assimilation window. CAMS-IFS-AER then provides 120 hours (5 days) of surface $PM_{2.5}$ concentration forecasts with approximately 40 km spatial resolution every 12 hours at 08:00 and 20:00MYT, made available through CAMS's global atmospheric composition forecast dataset (CAMS-GACF) (https://ads.atmosphere.copernicus.eu/cdsapp#!/dataset/cams-global-atmospheric-composition-forecasts?tab=overview). Only the 20:00MYT forecasts were used in this study because they provide forecasts closest to the next day in Malaysia. CAMS-IFS underwent a major upgrade to cycle 46r1 in 2019, including increasing vertical resolutions, coupling CAMS-IFS-AER with chemistry modules to model nitrate and ammonium aerosols, and added diurnal cycles to emissions (ECMWF, 2019; Rémy *et al.*, 2019), drastically affecting $PM_{2.5}$ forecasts (Basart *et al.*, 2019). The major 2019 transboundary haze episode also occurred during the operation of 46r1 CAMS-IFS. Therefore, this study will focus on forecasts produced by the 46r1 model, but other model versions within the study period were also assessed. More details on the model, its configurations and upgrades, and the CAMS-GACF are provided by Rémy *et al.* (2019) and ECMWF (2022a).

2.3 Datasets

(a) *Ground observations.* Malaysia has an air quality monitoring network since 1995, but $PM_{2.5}$ was only included as an air quality monitoring parameter in 2017, and as a subindex in the air pollutant index (API) in 2018 (DOE, 2018, 2017b). Hence, long-term consistent $PM_{2.5}$ records are limited. As of 2022, there are 65 continuous air quality monitoring stations (CAQMS) currently in operation throughout Malaysia. All CAQMS sample $PM_{2.5}$ using TEOM™ 1405-DF Continuous Dichotomous Ambient Air Monitors (https://www.thermofisher.com/order/catalog/product/TEOM1405DF?SID=srch-srp-TEOM1405DF), with the data undergoing reasonable quality control and assurance by the operating company before being published. The $PM_{2.5}$ concentration data are used with the national air quality index system and are also commonly used for air quality research in Malaysia (e.g., Ahmad Mohtar *et al.*, 2022; Sobri *et al.*, 2021).

We obtained hourly $PM_{2.5}$ concentrations for all CAQMS from 1 January 2018 to 31 August 2020, provided by DOE. The 65 CAQMS were grouped by the DOE into five geographical regions, i.e., North, Central, South, East, and Borneo (Fig. 1). In this study, North, South, and East were combined into one region named 'Peninsular'. The Central region was isolated as a distinct region as it contains the largest urban area in Malaysia, the Greater Kuala Lumpur region. We thus proceed with three regions defined: 'Peninsular', 'Central', and 'Borneo'.

(b) *Model forecasts.* The 20:00MYT hourly surface $PM_{2.5}$ forecasts were obtained from CAMS-GACF. The forecasts were then bilinearly interpolated to the latitude-longitude coordinates of each CAQMS. During the two-and-a-half-year study period, CAMS-IFS was upgraded twice, i.e., on 26 June 2018 and 9 July 2019 (ECMWF, 2022a), resulting in forecasts produced by three model cycles: 43r3, 45r1, and 46r1. As mentioned before, this study will focus on the later model version, 46r1, but 43r3 and 45r1 are also evaluated and compared.

2.4 Forecast Evaluation

The ground-observed and model-forecasted hourly $PM_{2.5}$ concentrations were first averaged through each time-horizon day, i.e., the 24 hours from 21:00MYT to 20:00MYT the next day. The first through to the fifth time-horizon days are known as F1–F5, covering 120 hours of the forecasts. Although we are particularly interested in the larger timescales of transboundary hazes, diurnal variations are also important when assessing CAMS-GACF (e.g., Wu *et al.*, 2020). Thus, diurnal accuracies of the forecasts are also evaluated separately as described in Section 2.4.2.

The forecasts were then evaluated against the observations (assumed as true benchmarks) via exceedance and accuracy analyses, which are described in the next section. Our evaluation considers all five time-horizon days, except when explicitly evaluating differences across time-horizons (see Section 2.4.3).

2.4.1 Exceedance analysis

Exceedance analysis dichotomously classifies the $PM_{2.5}$ status into 'normal' and 'bad' $PM_{2.5}$ air quality and assesses the forecasts' ability to predict them. This dichotomous classification is typically intrinsic to decision-making and early warnings. The analysis would evaluate whether the model can produce functional forecasts that may be useful in $PM_{2.5}$ management.

The threshold concentration levels to classify 'bad' and 'normal' $PM_{2.5}$ levels can be rather arbitrary, defined more by policies (Doswell, 2004). Past health studies classify 'bad' PM levels as concentrations above pre-defined standards or guidelines (Phung *et al.*, 2022; Sahani *et al.*, 2014). In this study, we defined the threshold according to Malaysia's National Ambient Air Quality Standards (MAAQS). The MAAQS were set up based on three different standards: 75 (IT-1, 2015), 50 (IT-2, 2018), and 35 $\mu g\ m^{-3}$ (IT-3, 2020) (DOE, 2014). Days with averaged $PM_{2.5}$ concentrations (rounded to the nearest $\mu g\ m^{-3}$) above these thresholds are considered exceedances. Using these thresholds, the forecasts' performances were evaluated using three metrics: probability of detection (POD), false alarm ratio (FAR), and critical success index (CSI) (see SM2). They reveal whether users can be confident in the forecasts to predict bad $PM_{2.5}$ days.

2.4.2 Accuracy analysis

Accuracy analysis utilises $PM_{2.5}$ concentration values directly to compute and aggregate some measures of accuracies using different metrics. While this analysis does not directly evaluate the forecasts' use in the policy domain, it links the forecasts' performance and improvement to specific areas in the forecasts. Since $PM_{2.5}$ pollution is spatiotemporally heterogenous in Malaysia, we need to fully evaluate the forecasts in space and time. This was done via four characterisation methods (adapted and altered from Meroni *et al.*, 2013) (Fig. 2):

(1) *Total characterisation.* Accuracies were aggregated across all CAQMS and time: overall forecast performance is characterised by a number;
(2) *Spatial characterisation.* Accuracies were aggregated across time: temporal performances are characterised for each CAQMS, represented by a map of Thiessen polygons;
(3) *Temporal characterisation.* Accuracies were aggregated across CAQMS: spatial performances are characterised for each day, represented by a timeline; and
(4) *Diurnal characterisation.* Since diurnal components were removed via the daily-averaging, diurnal variations in the raw hourly observed and forecasted $PM_{2.5}$ were averaged across CAQMS and time.

The accuracies were aggregated using five accuracy metrics: mean bias (MB), modified normalised mean bias (MNMB), root mean square error (RMSE), fractional gross error (FGE), and coefficient of determinant (R^2) (see SM3). They are all used by CAMS validation server in Europe (CAMS, 2022), while some are also commonly used to report model performance in the literature (e.g., Savage *et al.* (2013); see SM7). These metrics provide a comprehensive comparison between different regions and models in other studies. While MB and RMSE are intuitive because they

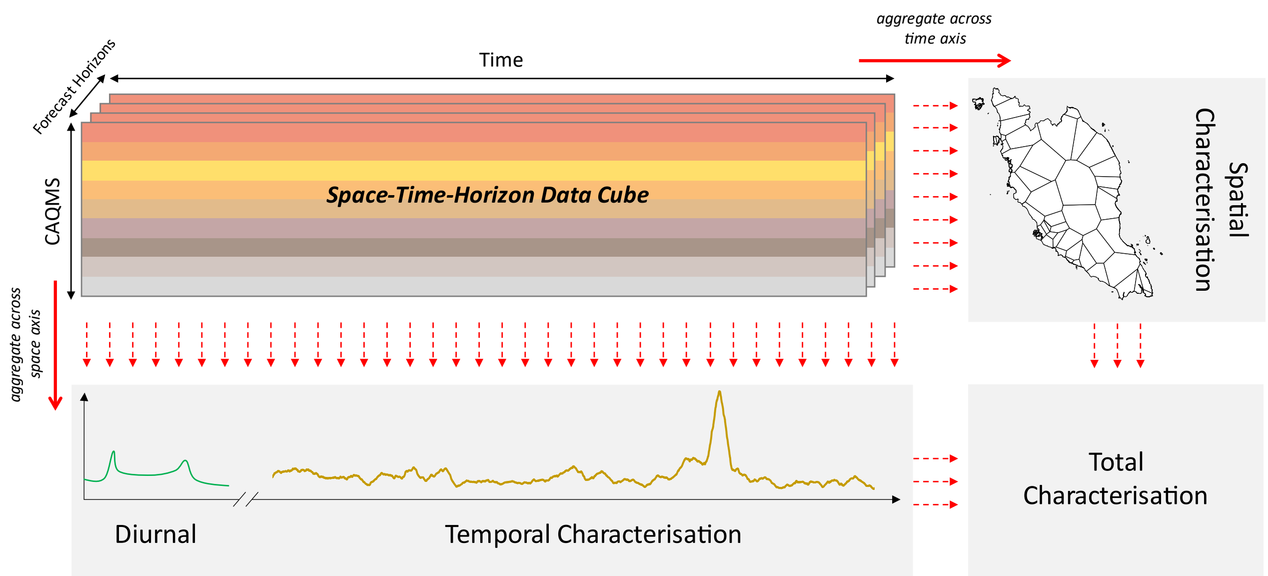

Fig. 2. A simple schematic of the accuracy analysis characterisation methods. Accuracies are aggregated across the time- (rightward) and space-axis (downward) to evaluate CAMS-GACF (space-time-horizon data cube).

are expressed in real units (μg m^{-3}), the normalised metrics, i.e., MNMB, FGE, and R^2, are better for comparisons between CAQMS and periods with different observed $PM_{2.5}$ concentrations. Therefore, we only used the normalised metrics for spatial and temporal characterisations. Since diurnal variations are important when evaluating CAMS-GACF, we also assessed diurnal characteristics by simple visual inspection of pattern differences between observations and forecasts, rather than through accuracy metrics.

2.4.3 Comparing variables

The forecasts' exceedance and accuracy performances were also evaluated against different comparing variables. Differences between our three geographical regions were assessed as a primary comparing variable, in addition to three secondary comparing variables (within which differences between regions were also assessed):

(a) *Non-haze and haze episodes*. Only the main haze episodes were considered. The months of August and September in the 2019 haze were marked as haze episodes (other periods are non-haze). Forecast performances during non-haze and haze episodes were compared.

(b) *Time-horizon days*. Model performances for the five different time-horizon days (i.e., F1–F5) were compared; and

(c) *Model versions*. Since no severe haze occurred during the operations of 43r3 and 45r1, only non-haze periods were compared between the model versions. Performances of different model versions were compared.

3 RESULTS

3.1 Overall Performance

The results of the exceedance analysis are shown in Table 1. Overall, the 46r1 CAMS-GACF performed better with lower exceedance thresholds. The highest POD, FAR, and CSI was obtained for the most stringent threshold (IT-3) at 46%, 52%, and 31%, respectively. POD and CSI increased while FAR decreased (i.e., all improved) with more stringent $PM_{2.5}$ thresholds for all regions except Central, where FAR and CSI were the highest and lowest (i.e., both worst) when using IT-2. The poorer performance in Central using IT-2 can be attributed to the overall overprediction here, where IT-2 threshold labelled forecasted levels as exceedances and observed levels labelled as non-exceedances, causing higher (poorer) FAR.

The results of the accuracy analysis total characterisation are shown in Table2. During the study period, overall mean observed $PM_{2.5}$ ($\bar{o}$) was 17.6 μg m^{-3}, while overall mean forecasted ($\bar{f}$) was lower at 14.3 μg m^{-3}. The 46r1 CAMS-GACF had an overall negative bias in Malaysia (underprediction; MB = –3.3 μg m^{-3}, MNMB = –0.38). In fact, Peninsular and Borneo had a negative bias (MB, MNMB) while Central had a positive bias. While RMSE was highest at Central followed by Borneo, FGE was highest at Borneo followed by Peninsular. But FGE, which measures proportional errors, are simply higher at regions with lower $PM_{2.5}$ concentrations despite similar or lower RMSE, which

Table 1. The 46r1 CAMS-GACF performance in forecasting $PM_{2.5}$ exceedances Malaysia-wide (overall) and in the three regions over three exceedance thresholds (IT-1, IT-2, IT-3). Performance was evaluated for all five time-horizon days.

	Threshold	Overall	Peninsular	Central	Borneo
POD (%)	IT-1	27	15	66	16
	IT-2	38	25	80	29
	IT-3	46	35	90	33
FAR (%)	IT-1	45	35	45	57
	IT-2	59	26	72	40
	IT-3	52	23	68	32
CSI (%)	IT-1	22	14	43	13
	IT-2	25	23	27	24
	IT-3	31	32	31	28

Table 2. The 46r1 CAMS-GACF performance in forecasting $PM_{2.5}$ concentrations Malaysia-wide (Malaysia) and in the three regions. Performance was evaluated for all five time-horizon days.

	$\bar{o}$ (µg m^{-3})	$\bar{f}$ (µg m^{-3})	MB (µg m^{-3})	RMSE (µg m^{-3})	MNMB	FGE	R^2
Overall	17.6	14.3	–3.3	15.1	–0.28	0.51	0.36
Peninsular	17.9	12.6	–5.2	12.0	–0.34	0.49	0.46
Central	25.1	35.6	10.4	19.4	0.36	0.46	0.44
Borneo	13.8	8.6	–5.2	18.6	–0.43	0.59	0.25

$\bar{o}$ and $\bar{f}$ are mean observed and forecasted $PM_{2.5}$ concentrations over the study period, respectively.

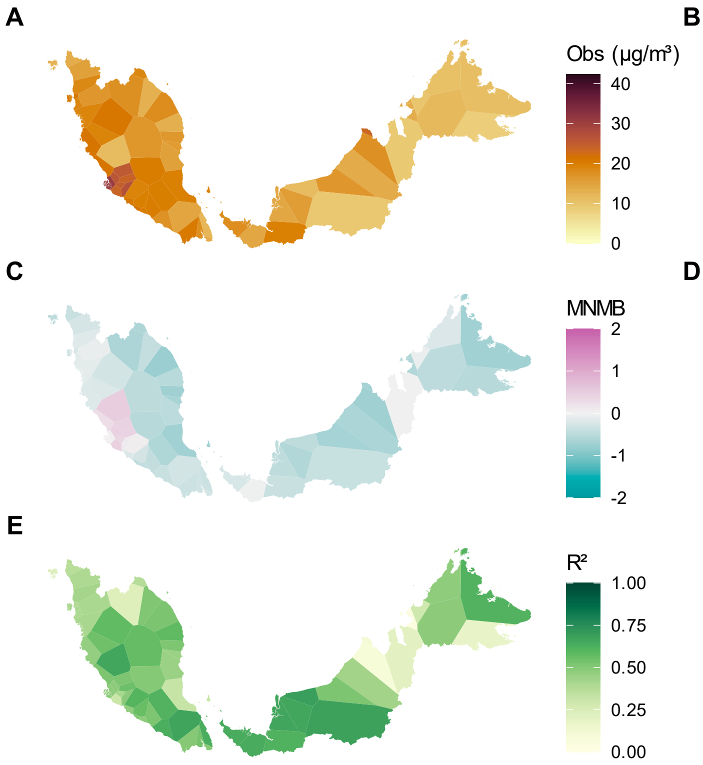

Fig. 3. The performance of 46r1 CAMS-GACF at each CAQMS (spatial characterisation): (a) mean observed and (b) mean forecasted $PM_{2.5}$ concentrations, (c) MNMB, (d) FGE, and (e) R^2. Performance was evaluated for all five time-horizon days.

measures additive errors. The overall R^2 is 0.36: 36% of the spatiotemporal variations in observed $PM_{2.5}$ can be explained by the model. However, the R^2 at Peninsular and Central were more than 0.4, but was only 0.25 at Borneo.

These regional differences were also evident in the forecasts' spatial characters (Fig. 3). Forecasted $PM_{2.5}$ looked evidently higher than observed around the Central region in Figs. 3(a) and 3(b), but vice versa elsewhere. Higher FGE were found at eastern Peninsular and central Borneo. The R^2 at each CAQMS were generally high around 0.75, but certain CAQMS at the northern and eastern Peninsular and central Borneo had low R^2; the temporal variations at these stations are not well represented in the forecasts.

When aggregating model accuracy spatially for each day (temporal characterisation), there were some temporal variations in MNMB Malaysia-wide and in the three regions (Fig. 4). FGE remained relatively constant except in Central and Borneo, at which FGE were slightly higher and lower during INMs, respectively. Malaysia-wide daily spatial R^2 were consistent around 0.3, similar to the total R^2, except during the 2019 haze episode and the March 2020 INM when R^2 were lower. Regionally, spatial R^2 were generally lower and had more temporal variations than Malaysia-wide R^2. Peninsular generally followed the Malaysia-wide trend in spatial R^2. Central

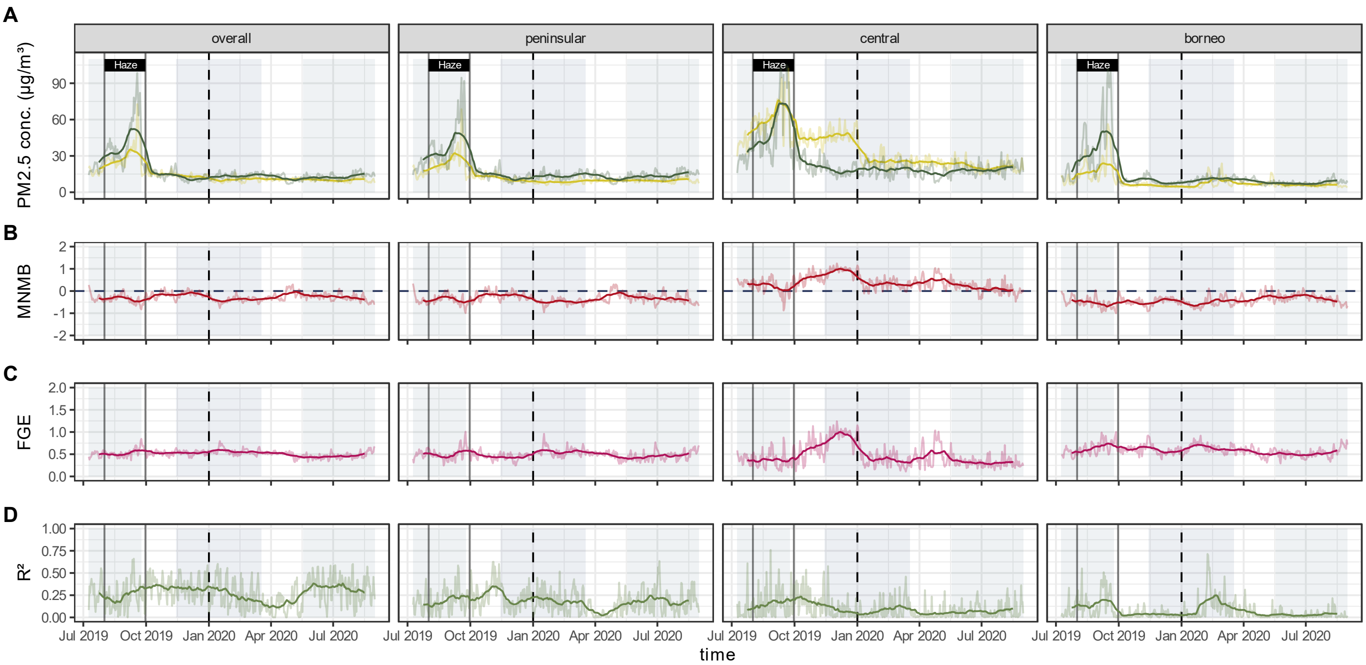

Fig. 4. The performance of 46r1 CAMS-GACF at each day (temporal characterisation) Malaysia-wide (overall) and in the three regions: (a) observed (green) and forecasted (yellow) $PM_{2.5}$ concentrations, (b) MNMB, (c) FGE, and (d) R^2 (thin line – daily data; thick line – 31-day running-average; shaded background – SWM/NEM; unshaded – INM). Performance was evaluated for all five time-horizon days.

and Borneo R^2 were around 0.1 but were higher during the end of the haze and during February–March 2020. We also found anomalous overpredicted forecasts at Central from October to December 2019, causing high MNMB, FGE, and low R^2 during these periods. This anomaly is only briefly discussed below (more information in SM5).

Finally, overall diurnal variations in 120 hours of forecasted $PM_{2.5}$ concentrations mostly fit the observations (Fig. 5). Forecasts were generally lower than observed throughout the whole day for all regions except Central, where night-time forecasts were higher. Regardless, the peaks in observed $PM_{2.5}$ concentration at 9 am and 8 pm were not present in the forecasts.

3.2 Non-haze and Haze Episodes

Next, we compared CAMS-GACF performances during non-haze and haze episodes, with the results shown in Table S2. Firstly, the forecasts predicted exceedances better during haze than non-haze periods for all thresholds. POD, FAR, and CSI behaved as found above according to the different thresholds: overall, POD and CSI increased and FAR decreased (i.e., all improved) with more stringent $PM_{2.5}$ thresholds Malaysia-wide and in the three regions during both non-haze and haze episodes. The forecasts predicted exceedances better using IT-3 during both periods.

CAQMS recorded three times higher $\bar{o}$ during haze (42.1 µg m^{-3}) than non-haze periods (13.5 µg m^{-3}), while $\bar{f}$ during both periods were lower (29.2 and 11.8 µg m^{-3}, respectively). Accuracy patterns during both non-haze and haze episodes were found to be similar to the overall performance: there were negative (positive) biases during both periods in Peninsular and Borneo (Central). However, negative biases were more negative while positive biases were less positive during haze. Although errors (RMSE, FGE) were higher during haze than non-haze episodes in Peninsular and Borneo, R^2 were higher during haze episodes across Malaysia and all regions.

In spatial characterisation, patterns observed during both periods were similar to that of the overall performance above, with positive (negative) bias around the Central region (elsewhere) and FGE being higher at eastern Peninsular and Borneo (Fig. 6). R^2 were visibly higher at all except a few CAQMS during the haze episode, suggesting that temporal variations were better represented in the forecasts during haze than non-haze periods.

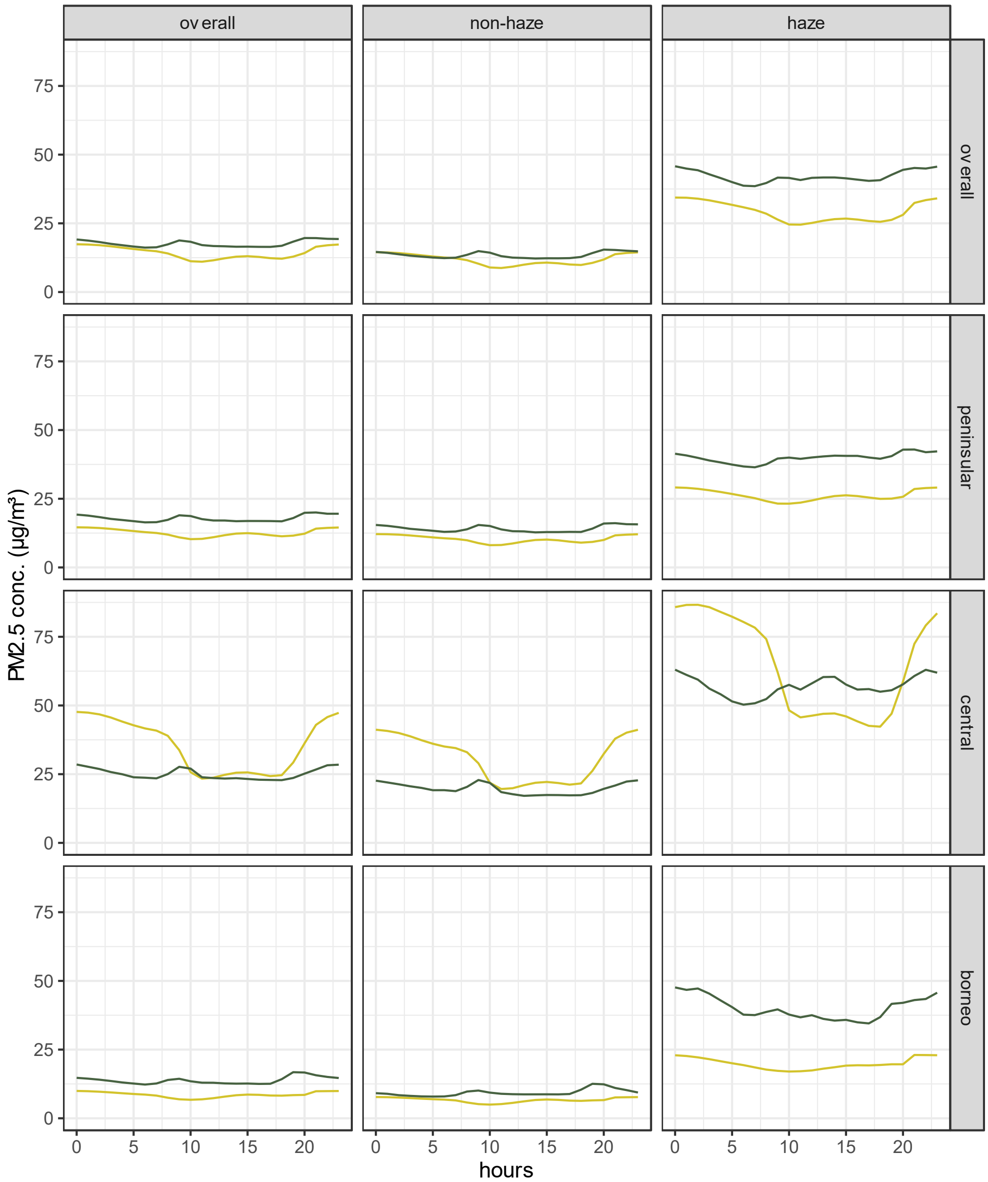

Fig. 5. Averaged diurnal variations in observed (green) and 120 hours of 46r1 model forecasted (yellow) $PM_{2.5}$ concentrations Malaysia-wide (overall) and in the three regions during all periods (overall), non-haze periods, and the 2019 haze.

Lastly, the overall forecasts' diurnal variation largely followed that of the observed during both haze and non-haze periods (Fig. 5). However, the forecasted rise in $PM_{2.5}$ concentrations during haze from normal levels was less than the observed rise (also found in Table S2). Besides that, forecasts at Central overestimated the rise in night-time concentration from daytime during haze, while underestimated the rise at Borneo.

3.3 Time-horizon Days

To assess how forecast accuracy varies with forecast time-horizon, Fig. S2 shows the results of the exceedance analysis (using IT-3) and accuracy analysis, segregated by different time-horizon days. In general, the first time-horizon day (F1) i.e., the first 24 hours of the forecasts, performed the best while the last time-horizon day (F5) performed the worst. Looking at the exceedance metrics, POD and CSI increased and FAR decreased (i.e., all improved) with decreasing time-horizons (from F5 to F1). All regions followed the same trend, except in Central where FAR increased and

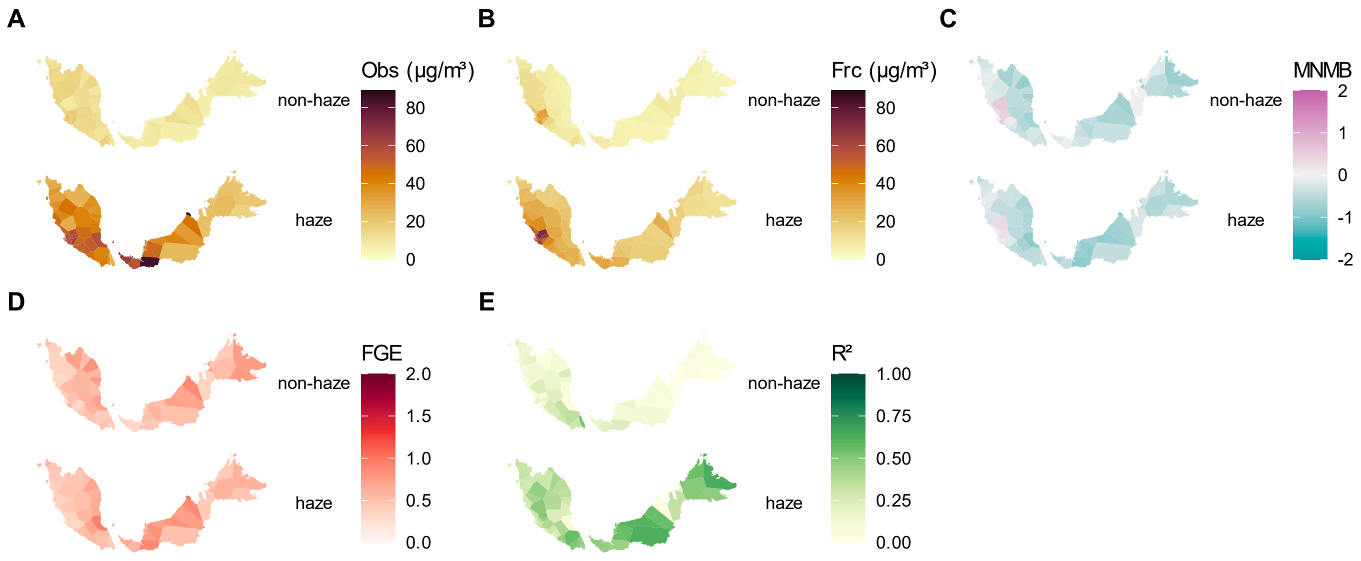

Fig. 6. The performance of 46r1 CAMS-GACF at each CAQMS (spatial characterisation) during non-haze and haze episodes: (a) mean observed and (b) mean forecasted $PM_{2.5}$ concentrations, (c) MNMB, (d) FGE, and (e) R^2. Performance was evaluated for all five time-horizon days.

CSI decreased (i.e., both worsened) with decreasing time-horizons instead. Similar patterns were observed for IT-1 and IT-2.

From the accuracy analysis, biases increased with decreasing time-horizons—forecasts at Peninsular and Borneo were less underpredicted, while forecasts at Central were more overpredicted. While errors decreased with decreasing time-horizons for most regions, they were the lowest at F2 and F3 in Central instead of F1. R^2 also increased with decreasing time-horizons, but R^2 at Peninsular and Central were highest at F2.

Comparing the time-horizons temporally (Fig. S3), the forecasted concentration tended to increase with decreasing time-horizons, i.e., forecasts made more recently tended to be higher than those made further back in the past. In fact, forecasted $PM_{2.5}$ appeared to move closer to the observed with decreasing time-horizons. This F1–F5 gap was more pronounced during the 2019 haze than during non-haze and reflects the forecasts' diurnal variations (Fig. S4). Nevertheless, F1 forecasts in Central appeared distinctly higher than forecasts made at other time-horizons during both haze and non-haze periods, particularly at night when forecasted $PM_{2.5}$ were much higher than other time-horizons.

3.4 Model Versions

Finally, we compared the performances of the three different model versions within our study period, i.e., 43r3, 45r1, and 46r1 (Table S3). Recall that there were no severe haze episodes occurring during cycles 43r3 and 45r1; only non-haze periods were compared. In general, both POD and FAR decreased (i.e., worsened and improved, respectively) with each new model version. To untangle this opposing trend, we used the CSI to determine 'good' or 'bad' model forecasts. Overall, 45r1 performed the best, followed by 46r1. However, the version that produced the best forecasts differed for different regions: 45r1 for Central and Borneo, and 46r1 for Peninsular.

Looking at the accuracy metrics, the 43r3 and 45r1 forecasts overall overpredicted $PM_{2.5}$ concentrations, while 46r1 forecasts overall underpredicted $PM_{2.5}$ concentrations. While RMSE were higher for 43r3 and 45r1, FGE of 46r1 was higher (again, lower forecasts tend to have higher FGE). 43r3 forecasts had the highest R^2, while R^2 of 45r1 and 46r1 were similar. However, regionally, 45r1 forecasts' R^2 was higher than 46r1 in Borneo, but vice versa elsewhere. Regionally, biases were positive (overprediction) at all regions for 43r3 and 45r1; only 46r1 produced forecasts with negative biases (underprediction) at Peninsular and Borneo. Errors generally increased in Peninsular and Borneo and decreased in Central with each new version.

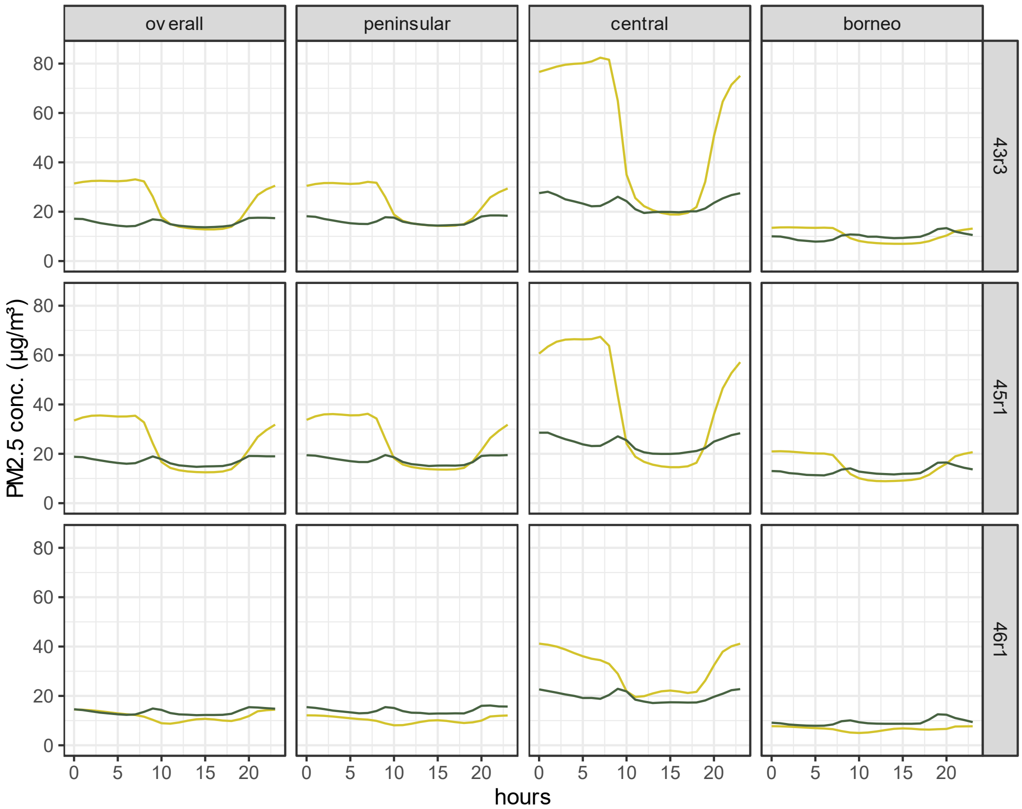

Fig. 7. Averaged diurnal variations in observed (green) and 120 hours of forecasted (yellow) $PM_{2.5}$ concentrations from three different model versions (43r3, 45r1, 46r1) throughout Malaysia (overall) and in the three regions during non-haze periods.

Spatially, there were obvious changes with the 46r1 upgrade (Fig. S5). Firstly, as noted above, there was a switch from overprediction to underprediction at most CAQMS. Similarly, FGE was higher at areas that were previously low (eastern Peninsular, Borneo etc.), and were lower at areas previously high (Central), probably due to the lowered $PM_{2.5}$ forecasts. Lastly, R^2 were lower at most CAQMS, suggesting that forecasts from the 46r1 model captured less of the temporal variations at most places than past versions.

Finally, the diurnal variations of cycle 46r1 were distinct from that of past versions, with lower night-time forecasted $PM_{2.5}$ concentrations across all regions (Fig. 7). In fact, the diurnal cycle in 46r1 forecasts fitted observations better than past versions. Nevertheless, 46r1 still overpredicted night-time $PM_{2.5}$ in Central, but to a lesser degree than past versions. The observed morning and evening $PM_{2.5}$ peaks were not captured in all three versions.

4 DISCUSSION

4.1 Exceedances and Early Warnings

Firstly, we assessed CAMS-GACF fitness for use in the policy domain, such as in early-warning systems and the broader scope of $PM_{2.5}$ management. Malaysia-wide, CAMS-GACF performed best in the exceedance analysis during both haze and non-haze periods when delineating exceedance levels using IT-3 (35 μg m^{-3}), the newer MAAQS. This suggests the introduction of the new MAAQS, as well as its promising health benefits, would be associated with improved CAMS-GACF performance in predicting exceedances of $PM_{2.5}$ in Malaysia. However, overall CAMS-GACF performed less well during non-haze periods (8% CSI), with an exceedance performance worse than found for regional CTMs and statistical forecasting models in other countries (38–54% CSI) (Celis *et al.*, 2022; Cho *et al.*, 2021; Huang *et al.*, 2017) (see SM6). In contrast, CAMS-GACF performed on-par or better (44% CSI) than these studies during haze episodes (20–54% CSI), when $PM_{2.5}$ levels are elevated and forecasts are of most value. The weaker performance of CAMS-GACF during non-haze periods should thus not devalue its potential in providing early warnings of extreme $PM_{2.5}$ events in Malaysia.

4.2 Characteristics of CAMS-GACF in Malaysia

In terms of accuracy analysis, CAMS-GACF performance in Malaysia (MB = –3.3 μg m^{-3}, MNMB = –0.28, RMSE = 15.1 μg m^{-3}, FGE = 0.51, R^2 = 0.36) is poorer than its performances in mid-latitude settings like Europe, United States, and China that has lower biases and errors and higher R^2 (MB_{median} = –1.6 μg m^{-3}, $MNMB_{median}$ = –0.13, $RMSE_{median}$ = 3.7 μg m^{-3}, FGE_{median} = 0.32, R^2_{median} = 0.44) (CAMS, 2022; Wu *et al.*, 2020). The global CAMS-GACF also performed poorer in Malaysia than many regional CTMs in other countries that has higher R^2 (0.39–0.79) (Huang *et al.*, 2017; Lyu *et al.*, 2017; Neal *et al.*, 2014; Savage *et al.*, 2013) (see full comparison in SM7). To scrutinise this difference, we highlight key characteristics of CAMS-GACF in Malaysia in three subsections below.

4.2.1 Large- and small-scale variability

CAMS-GACF performed as expected for a global model forecast—it performed better at larger spatiotemporal scales than at smaller ones. We assessed this using the R^2 metric, which measures the proportion of variability explained by the forecasts. We found higher Malaysia-wide spatial R^2 (in temporal characterisation) than for the smaller regions of Malaysia at most times. The 40 km and 10 km resolutions of CAMS-IFS and its emission inventories hinder representation of processes with smaller spatial-scales which often affect local-scale variations in $PM_{2.5}$.

The reduced robustness of CAMS-GACF at local scales is limited to the spatial dimension. R^2 was not improved by removing the temporal component from the data through temporal characterisation. Rather, Malaysia-wide total R^2 was lower than the regional ones; and regional total R^2 were lower than local ones (i.e., at individual CAQMS). Hence, a large proportion of the reduced robustness (or R^2) at local scales can be attributed to poor representation of Malaysia's $PM_{2.5}$ spatial heterogeneity in the forecasts.

Nevertheless, some temporal variations were also not accounted for in the forecasts. MNMB and FGE showed some intra-annual variations, while the diurnal variations and the local-scale processes affecting it (e.g., traffic peaks) were poorly captured by the forecasts. This conformed to other studies that also used CAMS products (Varga-Balogh *et al.*, 2020; Wu *et al.*, 2020). However, the poor intra-annual temporal representation is only limited to CAQMS in eastern Peninsular and central Borneo with high FGE and low temporal R^2.

Conversely, larger scale variability was captured by the forecast. The forecast performed better in all regions in terms of R^2 during haze than non-haze periods. This highlights the key strength of CAMS-IFS in that emissions from fires are captured through GFAS and their regional transport are modelled well.

4.2.2 Emission sources and diurnal cycle

Our analysis also showed the differing accuracy of CAMS-GACF between periods where $PM_{2.5}$ pollution is either most influenced by local or external emission sources. For example, during the 2019 transboundary haze event, with external sources dominant, the Central region $PM_{2.5}$ was less overpredicted (i.e., improved accuracy) while elsewhere became more underpredicted (i.e., less accurate). R^2 was also higher during the 2019 haze than non-haze periods. We can make two related deductions: (1) 46r1 CAMS-IFS underestimated the amount of $PM_{2.5}$ transported away from pollution sources, and would underpredict $PM_{2.5}$ concentrations during transboundary haze and overpredict when local pollution is dominant; and (2) CAMS-IFS can forecast well regional-scale $PM_{2.5}$ pollution like the 2019 transboundary haze, but is less adept at forecasting non-haze periods when local factors dominate. This conformed to our exceedance analysis results, in that CAMS-GACF can detect major transboundary haze but not regular, minor locally driven exceedances.

CAMS-GACF also overpredicted $PM_{2.5}$ in Central during both haze and non-haze periods, but underpredicted elsewhere. Wu *et al.* (2020) also found similar overprediction for more populous and polluted areas in China. CAMS-GACF appeared to overestimate the retention of $PM_{2.5}$ at pollution sources. The night-time overprediction only in the urban, more polluted Central region points towards inaccurate diurnal modelling of a nocturnal inversion layer (NIL). NIL can inhibit vertical mixing and cause accumulation of pollutants on the surface at night. Figs. S8 and S9 reinforce the hypothesis on NIL modelling, where greater $PM_{2.5}$ retention was found for areas with higher emissions in Central and Peninsular. While high observed night-time $PM_{2.5}$ is theoretically possible in the Central region, local-scale processes (e.g., potentially greater rainfall, shorter-lasting or less shallow NIL than modelled) probably caused lower observed night-time $PM_{2.5}$ (Sani, 1977). Similarly, CAMS-IFS uses extrapolated monthly emission inventories which are likely unrepresentative of current emissions in Malaysia. Diurnal emissions are also likely modelled through a simple function that did not capture local reality. Again, CAMS-IFS and its inventories are not designed to capture these local-scale processes. This finding is consistent with past CAMS-related studies (Marécal *et al.*, 2015; Wu *et al.*, 2020).

4.2.3 Assimilation across time-horizons

The difference in forecasts between time-horizon days can likely be attributed to the 4D-Var assimilation in CAMS-IFS. As forecast time-horizon decreases (i.e., F5 to F1), forecasted $PM_{2.5}$ concentration agrees better with observations (Fig. S3). F1 forecasts performed the best, which is expected of any forecasts that employ data assimilation. The F1–F5 gap showed improvements in forecasts due to assimilation—if no gap was observed i.e., similar forecasts produced over five days, either the forecasts were accurate, or there are insufficient satellite observations that can be assimilated. The wider F1–F5 gap during haze suggests that haze forecasts were most benefited by the assimilation system. Since the gap is present for all regions at most times, it suggests useful satellite observations e.g., from MODIS (Benedetti *et al.*, 2009; Rémy *et al.*, 2019), commonly exist in the region. However, periods like the March INMs and regions like Central and Borneo with poor spatial R^2 might be suffering from poor satellite observations (e.g., due to cloud cover). Wu *et al.* (2020) also found increasing accuracies from F5 to F1 but with an evident diurnal variation (also seen in this study). Satellite assimilation might not be sufficient to improve diurnal variations of $PM_{2.5}$ concentrations, particularly at night when aerosol-related observations are poor or unavailable.

Peculiarly, F2 and F3 forecasts were better than F1 in Central despite data assimilation. The higher $PM_{2.5}$ forecasted during F1 were mainly caused by the higher night-time forecasts, pointing again to a problem with diurnal and NIL modelling, satellite AOD assimilation methods (e.g., vertical distributions that amplify modelled night-time concentrations), and/or erroneous emission inventories. These issues might also be the cause of the October–December 2019 anomaly. However, the sudden change in $PM_{2.5}$ forecasts right on 1st January 2020 is more indicative of a change in emission inventory, though the true cause remains uncertain (see details in SM5).

4.3 Past and Future Performances

Finally in this section, we assess the implications of the model upgrades for the prospect of using CAMS-GACF for $PM_{2.5}$ forecasts in Malaysia. Among the three model versions we studied, cycle 45r1 performed the best when we considered only 'non-haze' periods. However, a minor transboundary haze not defined in this study as a 'haze episode' occurred during 45r1 operation and might inflate performance. Cycle 45r1 may also have been more calibrated after continual upgrades (hence, more accurate) before a major change that was 46r1. Nevertheless, regional R^2 of 46r1 forecasts at Peninsular and Central regions were higher than 45r1 despite lower R^2 at each CAQMS, suggesting that spatial variations of $PM_{2.5}$ in Peninsular and Central were better represented with the newer CAMS-IFS version. The opposite was true for Borneo. Regardless, 46r1 improved $PM_{2.5}$ forecasts through an improved diurnal cycle. While poor diurnal representation still persists in more urban and/or polluted environments, it was improved in other regions (e.g., from changes leading to better particle-size binning, better emission, chemistry, and deposition diurnal cycles).

Given CAMS-GACF past changes, we can expect frequent upgrades to the forecasting system. Since at the time of writing 46r1 is no longer the newest version, some characteristics highlighted in this study might change (see SM5). Nevertheless, we can expect CAMS-GACF to improve in the future.

4.4 Recommendations

Overall, we found that CAMS-GACF performance was weaker in Malaysia than is reported in other studies, focused on different countries and/or with different forecasting models, in both exceedance and accuracy analyses. CAMS-GACF cannot be expected to capture local-scale variability, which hinders accurate forecasting of $PM_{2.5}$ pollution that normally have both regional and local influences. Nevertheless, CAMS-GACF performed on par or better than forecasts from those studies when we consider haze periods only, can capture regional-scale $PM_{2.5}$ variability, employs data assimilation to improve forecasts, and provides suitable forecasting time-horizons for timely decision-making. It is thus interesting to provide some recommendations on how best to take advantage of CAMS-GACF despite its shortcomings.

Firstly, we recommend development of a robust early-warning system for use in Malaysia. CAMS-GACF should not be used as the sole information for early warning, with 100% confidence given to the forecasts. Confidence of the forecasts depends on the forecasts' systematic errors (which can be reduced via local bias-correction; see below) and on other auxiliary information (e.g., known hotspot locations, consistent exceedances predicted by most time-horizons). Uncertainty from forecast background cannot be reduced, only quantified based on human interpretation and/or good early-warning framework incorporating relevant auxiliary information (Doswell, 2004; Gerapetritis and Pelissier, 2004). Early warnings are then issued when confidence of exceedance events exceeds a threshold defined based on early-warning goals, the basic units (e.g., CAMQS, district, states), and the evaluation metrics (e.g., CSI, potential economic loss prevented by warnings). With a robust early-warning system in place, CAMS-GACF can contribute to improved early warnings' performance than reported here simply as exceedances, even without bias-correction.

Secondly, CAMS-GACF can be improved by incorporating the uncaptured local-scale variability via local bias-correction techniques. Being a global model, CAMS-GACF will not change to capture local-scale variability in the foreseeable future. Hence, the influences of local-scale processes can be incorporated via two methods. First, we can downscale CAMS-IFS by using a regional/local CTM. Cho *et al.* (2021) downscaled CAMS-IFS output via a regional CTM and found satisfactory results. Second, we can incorporate local-scale variability via statistical techniques. Many studies employ this method, ranging from simple statistical models like regression linking past forecasts/biases to future (corrected) forecasts (Konovalov *et al.*, 2009), to more complex frameworks classifying forecasts to past analogues and associated bias-correction models (Huang *et al.*, 2017; Lyu *et al.*, 2017; Neal *et al.*, 2014). These studies reported improved performance, particularly in correcting over- and underpredictions. Therefore, we recommend exploration of downscaling CAMS-GACF using a regional CTM, and of correcting the resulting bias via statistical models and bias-correction frameworks. Statistical and ML models used in Malaysia-focused forecasting studies (Koo *et al.*, 2020; Lim *et al.*, 2008; Wong *et al.*, 2021) can be repurposed for the latter. Regardless of the correction methods, model users that wish to derive maximum benefits from CAMS-GACF might want to prioritise the improvements at more polluted and populous regions like Central, while those that wish to improve overall CAMS-GACF accuracy might want to prioritise improvements at Borneo and eastern Peninsular where FGE were high.

Finally, some additional considerations were provided. Since $PM_{2.5}$ chemical composition can be used to determine its source(s) (Adam *et al.*, 2021), forecasted and observed composition can aid in both early warnings and bias-corrections as auxiliary information. Similarly, since PM (as aerosols) and meteorology are interdependent (Adam *et al.*, 2021; Dahari *et al.*, 2020; Ku Yusof *et al.*, 2019; Sobri *et al.*, 2021), existing weather forecasting and $PM_{2.5}$ forecasting are synergistic and can mutually improve each forecast. Finally, due to CAMS-IFS frequent upgrades, we recommend developing robust early-warning systems and simple bias-correction frameworks that ensure easy re-calibration to new upgrades. The forecast analogue approach might help in this regard.

5 CONCLUSION

In this study, we evaluated the performance of a global mechanistic CTM forecast, CAMS-GACF, in forecasting $PM_{2.5}$ in Malaysia qualitatively and quantitatively. It provided a regional outlook on

CAMS-GACF $PM_{2.5}$ forecasting performance in Malaysia for the first time. In summary, the change in MAAQS would not jeopardise but rather improve CAMS-GACF performance in predicting exceedance events. The model performed slightly worse than forecasting models used in other countries, but it performed on-par or better when forecasting is of most value, i.e., during the 2019 haze event. Accuracy-wise, CAMS-GACF performed worse in Malaysia than in other countries. It tended to overpredict $PM_{2.5}$ in polluted urban areas but underpredict elsewhere, likely due to emission inventory limitations, and challenges in diurnal and NIL modelling. Data assimilation of CAMS-IFS has proved effective, with improving forecasts from F5 to F1; haze forecasts benefited most from this feature. However, CAMS-GACF performed poorly at small scales in Malaysia, particularly in the spatial dimension. Short-term temporal variations were also not fully represented in the forecasts, particularly in the diurnal variations in polluted urban areas. CAMS-GACF also performed poorly when forecasting local pollution and exceedances. Nevertheless, CAMS-IFS receives frequent upgrades, and we can expect improvements to $PM_{2.5}$ forecasts in Malaysia in the future.

$PM_{2.5}$ is a dangerous and prevalent pollutant in Malaysia that has both local and external factors influencing its concentrations. Currently, Malaysia only issues air quality deterioration warnings based on observed concentrations (Wong *et al.*, 2021). Preventive warnings based on forecasts can benefit Malaysia in mitigating the impacts of elevated $PM_{2.5}$ by allowing governments and individuals to plan for exceedances (Celis *et al.*, 2022; Lyu *et al.*, 2017). CAMS-GACF is suitable for forecasting $PM_{2.5}$, with relevant time-horizons for decision-making and its considerations of regional processes. Hence, we provided recommendations to allow us to take advantage of CAMS-GACF despite its shortcomings: (1) develop a robust early-warning system around CAMS-GACF to maximise early warning efficacies; and (2) correct inaccuracies of CAMS-GACF via downscaling and utilising statistical bias-correction techniques. Potential focus areas and synergies were also highlighted.

This study provides a comprehensive initial review on CAMS-GACF performance in forecasting $PM_{2.5}$ in Malaysia. Future studies should further quantify the degree CAMS-GACF capture local-scale variations by scrutinising forecasts at individual CAQMS (particularly on the diurnal variations), and the improvements after applying bias-correction techniques recommended above. Future studies should also evaluate CAMS-GACF performances in forecasting $PM_{2.5}$ composition species, and indeed for other air pollutants. If the outcomes are satisfactory, CAMS-GACF could form the basis for a working air quality forecasting system for Malaysia.

ACKNOWLEDGEMENT

The authors wish to thank the Malaysian Department of Environment for providing the air quality data, and Ahmad Mirza bin Mokhtar for technical assistance.

The forecasted $PM_{2.5}$ concentration data were provided by Copernicus Atmosphere Monitoring Service (CAMS), and the outcome of this study contains modified CAMS information [2018-2022]. Neither the European Commission nor the European Centre for Medium-Range Weather Forecasts is responsible for any use that may be made of the Copernicus information or the data it contains.

SUPPLEMENTARY MATERIAL

Supplementary material for this article can be found in the online version at https://doi.org/10.4209/aaqr.220444

REFERENCES

Adam, M.G., Tran, P.T.M., Bolan, N., Balasubramanian, R. (2021). Biomass burning-derived airborne particulate matter in Southeast Asia: A critical review. J. Hazard. Mater. 407, 124760. https://doi.org/10.1016/j.jhazmat.2020.124760

Ahmad Mohtar, A.A., Latif, M.T., Dominick, D., Chel Gee Ooi, M., Azhari, A., Baharudin, N.H., Hanif, N.M., Chung, J.X., Juneng, L. (2022). Spatiotemporal variations of particulate matter and their association with criteria pollutants and meteorology in Malaysia. Aerosol Air Qual. Res. 22, 220124. https://doi.org/10.4209/aaqr.220124

Amil, N., Latif, M.T., Khan, M.F., Mohamad, M. (2016). Seasonal variability of $PM_{2.5}$ composition and sources in the Klang Valley urban-industrial environment. Atmos. Chem. Phys. 16, 5357–5381. https://doi.org/10.5194/ACP-16-5357-2016

Amin Jaafar, S., Latif, T., Razak, I.S., Bahiyah, N., Wahid, A., Khan, F., Srithawirat, T. (2018). Composition of carbohydrates, surfactants, major elements and anions in $PM_{2.5}$ during the 2013 Southeast Asia high pollution episode in Malaysia. Particuology 37, 119–126. https://doi.org/10.1016/j.partic.2017.04.012

Bannister, R.N. (2007). Elementary 4D-Var. DARC Technical Report (No. 2). Data Assimilation Research Centre (DARC), University of Reading, UK.

Basart, S., Benedictow, A., Bennouna, Y., Blechschmidt, A.M., Chabrillat, S., Christophe, Y., Cuevas, E., Eskes, H.J., Hansen, K.M., Jorba, O., Kapsomenakis, J., Langerock, B., Pay, T., Richter, A., Sudarchikova, N., Schulz, M., Wagner, A., Zerefos, C. (2019). Upgrade verification note for the CAMS real-time global atmospheric composition service: Evaluation of the e-suite for the CAMS upgrade of July 2019. Copernicus Atmosphere Monitoring Service (CAMS) report. https://doi.org/10.24380/fcwq-yp50

Benedetti, A., Morcrette, J.J., Boucher, O., Dethof, A., Engelen, R.J., Fisher, M., Flentje, H., Huneeus, N., Jones, L., Kaiser, J.W., Kinne, S., Mangold, A., Razinger, M., Simmons, A.J., Suttie, M. (2009). Aerosol analysis and forecast in the European Centre for Medium-Range Weather Forecasts integrated forecast system: 2. data assimilation. J. Geophys. Res. 114, D13205. https://doi.org/10.1029/2008JD011115

CAMS (2022). Evaluation of global forecasts. CAMS Validation Server. https://global-evaluation.atmosphere.copernicus.eu/ (accessed 23 April 2022).

Casallas, A., Celis, N., Ferro, C., López Barrera, E., Peña, C., Corredor, J., Ballen Segura, M. (2020). Validation of PM_{10} and $PM_{2.5}$ early alert in Bogotá, Colombia, through the modelling software WRF-CHEM. Environ. Sci. Pollut. Res. 27, 35930–35940. https://doi.org/10.1007/S11356-019-06997-9

Celis, N., Casallas, A., López-Barrera, E.A., Martínez, H., Peña Rincón, C.A., Arenas, R., Ferro, C. (2022). Design of an early alert system for $PM_{2.5}$ through a stochastic method and machine learning models. Environ. Sci. Policy 127, 241–252. https://doi.org/10.1016/j.envsci.2021.10.030

Cho, S., Park, H., Son, J., Chang, L. (2021). Development of the global to mesoscale air quality forecast and analysis system (GMAF) and its application to $PM_{2.5}$ forecast in Korea. Atmosphere 12, 411. https://doi.org/10.3390/atmos12030411

Crippa, M., Guizzardi, D., Muntean, M, Schaaf, E., Dentener, F., van Aardenne, J.A., Monni, S., Doering, U., Oliver, J.G.J., Pagliari, V., Janssens-Maenhout, G. (2018). Gridded Emissions of Air Pollutants for the period 1970-2012 within EDGAR v4.3.2. Earth Syst. Sci. Data 10, 1987–2013. https://doi.org/10.5194/essd-2018-31

Dahari, N., Muda, K., Hussein, N., Latif, M.T., Khan, M.F., Mohamad Khir, M.S. (2019). Long-range transport and local emission of atmospheric $PM_{2.5}$ in southern region of Peninsular Malaysia. IOP Conf. Ser.: Mater. Sci. Eng. 636, 012005. https://doi.org/10.1088/1757-899X/636/1/012005

Dahari, N., Latif, M.T., Muda, K., Hussein, N. (2020). Influence of meteorological variables on suburban atmospheric $PM_{2.5}$ in the southern region of Peninsular Malaysia. Aerosol Air Qual. Res. 20, 14–25. https://doi.org/10.4209/aaqr.2019.06.0313

Department of Environment Malaysia (DOE) (2014). New Malaysia Ambient Air Quality Standard. DOE Official Portal. https://www.doe.gov.my/2021/10/04/standard-kualiti-udara/ (accessed 18 April 2022).

Department of Environment Malaysia (DOE) (2017a). Environmental Quality Report 2017. Ministry of Energy, Science, Technology, Environment & Climate Change Malaysia (MESTECC). Shah Alam, Selangor.

Department of Environment Malaysia (DOE) (2017b). Launch of National Environmental Command Center (NECC). DOE Official Portal. https://www.doe.gov.my/en/2021/10/26/ministry-of-natural-resources-press-statement-launch-of-national-environmental-command-center-necc-11-april-2017/ (accessed 14 July 2022).

Department of Environment Malaysia (DOE) (2018). Environmental Quality Report 2018. Ministry of Energy, Science, Technology, Environment & Climate Change Malaysia (MESTECC). Subang Jaya, Selangor.

Department of Environment Malaysia (DOE) (2019). Environmental Quality Report 2019. Ministry of Environment and Water Malaysia (KASA). Bandar Baru Bangi, Selangor.

Department of Environment Malaysia (DOE) (2020). Environmental Quality Report 2020. Ministry of Environment and Water Malaysia (KASA). Petaling Jaya, Selangor.

Department of Environment Malaysia (DOE) (2022). Chronology of Haze Episodes in Malaysia. DOE Official Portal. https://www.doe.gov.my/2021/10/04/kronologi-episod-jerebu-di-malaysia-2/ (accessed 17 April 2022).

Doswell, C.A.I. (2004). Weather forecasting by humans - Heuristics and decision making. Weather Forecasting 19, 1115–1126. https://doi.org/10.1175/WAF-821.1

Dotse, S.Q., Dagar, L., Petra, M.I., de Silva, L.C. (2016). Influence of Southeast Asian Haze episodes on high PM_{10} concentrations across Brunei Darussalam. Environ. Pollut. 219, 337–352. https://doi.org/10.1016/j.envpol.2016.10.059

European Centre for Medium-range Weather Forecast (ECMWF) (2019). Implementation of IFS cycle 46r1. ECMWF Confluence Wiki. https://confluence.ecmwf.int/display/COPSRV/Implementation+of+IFS+cycle+46r1 (accessed 24 April 2022).

European Centre for Medium-range Weather Forecast (ECMWF) (2022a). CAMS: Global atmospheric composition forecast data documentation. ECMWF Confluence Wiki. https://confluence.ecmwf.int/display/CKB/CAMS%3A+Global+atmospheric+composition+forecast+data+documentation (accessed 19 April 2022).

European Centre for Medium-range Weather Forecast (ECMWF) (2022b). CAMS: Global Fire Assimilation System (GFAS) data documentation. ECMWF Confluence Wiki. https://confluence.ecmwf.int/display/CKB/CAMS%3A+Global+Fire+Assimilation+System+%28GFAS%29+data+documentation (accessed 9 July 2022).

Fujii, Y., Mahmud, M., Tohno, S., Okuda, T., Mizohata, A. (2016). A case study of $PM_{2.5}$ characterization in Bangi, Selangor, Malaysia during the southwest monsoon season. Aerosol Air Qual. Res. 16, 2685–2691. https://doi.org/10.4209/aaqr.2015.04.0277

Gerapetritis, H., Pelissier, J.M. (2004). On the Behavior of the Critical Success Index. Eastern Regional Technical Attachment (No. 2004-03). NOAA & NWS of the US. Greer, South Carolina.

Granier, C., Darras, S., Denier van der Gon, H., Doubalova, J., Elguindi, N., Galle, B., Gauss, M., Guevara, M., Jalkanen, J.P., Kuenen, J., Liousse, C., Quack, B., Simpson, D., Sindelarova, K. (2019). The Copernicus Atmosphere Monitoring Service global and regional emissions (April 2019 version). Copernicus Atmosphere Monitoring Service (CAMS) report. https://doi.org/10.24380/D0BN-KX16

Hoesly, R.M., Smith, S.J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., Seibert, J.J., Vu, L., Andres, R.J., Bolt, R.M., Bond, T.C., Dawidowski, L., Kholod, N., Kurokawa, J., Li, M., Liu, L., Lu, Z., Moura, M.C.P., O'Rourke, P.R., Zhang, Q. (2018). Historical (1750-2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS). Geosci. Model Dev. 11, 369–408. https://doi.org/10.5194/gmd-11-369-2018

Huang, J., McQueen, J., Wilczak, J., Djalalova, I., Stajner, I., Shafran, P., Allured, D., Lee, P., Pan, L., Tong, D., Huang, H.C., DiMego, G., Upadhayay, S., Monache, L.D. (2017). Improving NOAA NAQFC $PM_{2.5}$ predictions with a bias correction approach. Weather Forecasting 32, 407–421. https://doi.org/10.1175/WAF-D-16-0118.1

Juneng, L., Latif, M.T., Tangang, F.T., Mansor, H. (2009). Spatio-temporal characteristics of PM_{10} concentration across Malaysia. Atmos. Environ. 43, 4584–4594. https://doi.org/10.1016/j.atmosenv.2009.06.018

Konovalov, I.B., Beekmann, M., Meleux, F., Dutot, A., Foret, G. (2009). Combining deterministic and statistical approaches for PM_{10} forecasting in Europe. Atmos. Environ. 43, 6425–6434. https://doi.org/10.1016/j.atmosenv.2009.06.039

Koo, J.W., Wong, S.W., Selvachandran, G., Long, H.V., Son, L.H. (2020). Prediction of Air Pollution Index in Kuala Lumpur using fuzzy time series and statistical models. Air Qual. Atmos. Health 13, 77–88. https://doi.org/10.1007/S11869-019-00772-Y

Ku Yusof, K.M.K., Azid, A., Abdullah Sani, M.S., Samsudin, M.S., Muhammad Amin, S.N.S., Abd

Rani, N.L., Jamalani, M.A. (2019). The evaluation on artificial neural networks (ANN) and multiple linear regressions (MLR) models over particulate matter (PM_{10}) variability during haze and non-haze episodes: A decade case study. Malaysian J. Fundam. Appl. Sci. 15, 164–172. https://doi.org/10.11113/mjfas.v15n2.1004

Kusumaningtyas, S.D.A., Aldrian, E. (2016). Impact of the June 2013 Riau province Sumatera smoke haze event on regional air pollution. Environ. Res. Lett. 11, 075007. https://doi.org/10.1088/1748-9326/11/7/075007

Lelieveld, J., Klingmüller, K., Pozzer, A., Burnett, R.T., Haines, A., Ramanathan, V. (2019). Effects of fossil fuel and total anthropogenic emission removal on public health and climate. Proc. Natl. Acad. Sci. U.S.A. 116, 7192–7197. https://doi.org/10.1073/pnas.1819989116

Lim, Y.S., Lim, Y.C., Mah, P.J.W. (2008). ARIMA and Integrated ARFIMA Models for Forecasting Air Pollution Index in Shah Alam, Selangor. The Malaysian J. Anal. Sci. 12, 257–263. http://pkukmweb.ukm.my/mjas/v12_n1/36.pdf

Lyu, B., Zhang, Y., Hu, Y. (2017). Improving $PM_{2.5}$ air quality model forecasts in China using a bias-correction framework. Atmosphere 8, 147. https://doi.org/10.3390/ATMOS8080147

Manan, N.A., Rizal, M., Manaf, A., Hod, R. (2018). The Malaysia haze and its health economic impact: a literature review. Malaysian J. Public Health Med. 18, 38–45.

Marécal, V., Peuch, V.H., Andersson, C., Andersson, S., Arteta, J., Beekmann, M., Benedictow, A., Bergström, R., Bessagnet, B., Cansado, A., Chéroux, F., Colette, A., Coman, A., Curier, R.L., Denier Van Der Gon, H.A.C., Drouin, A., Elbern, H., Emili, E., Engelen, R.J., Eskes, H.J., *et al.* (2015). A regional air quality forecasting system over Europe: the MACC-II daily ensemble production. Geosci. Model Dev. 8, 2777–2813. https://doi.org/10.5194/gmd-8-2777-2015

Meroni, M., Atzberger, C., Vancutsem, C., Gobron, N., Baret, F., Lacaze, R., Eerens, H., Leo, O. (2013). valuation of agreement between space remote sensing SPOT-VEGETATION fAPAR time series. IEEE Trans. Geosci. Remote Sens. 51, 1951–1962. https://doi.org/10.1109/TGRS.2012.2212447

Neal, L.S., Agnew, P., Moseley, S., Ordóñez, C., Savage, N.H., Tilbee, M. (2014). Application of a statistical post-processing technique to a gridded, operational, air quality forecast. Atmos. Environ. 98, 385–393. https://doi.org/10.1016/j.atmosenv.2014.09.004

Ooi, E.L., Mustaffa, N.I.H., Amil, N., Khan, M.F., Latif, M.T. (2015). Source contribution of $PM_{2.5}$ at different locations on the Malaysian peninsula. Bull. Environ. Contam. Toxicol. 94, 537–542. https://doi.org/10.1007/S00128-015-1477-9

Phung, V.L.H., Ueda, K., Sahani, M., Seposo, X.T., Wan Mahiyuddin, W.R., Honda, A., Takano, H. (2022). Investigation of association between smoke haze and under-five mortality in Malaysia, accounting for time lag, duration and intensity. Int. J. Epidemiol. 51, 155–165. https://doi.org/10.1093/ije/dyab100

Reddington, C.L., Yoshioka, M., Balasubramanian, R., Ridley, D., Toh, Y.Y., Arnold, S.R., Spracklen, D.V. (2014). Contribution of vegetation and peat fires to particulate air pollution in Southeast Asia. Environ. Res. Lett. 9, 094006. https://doi.org/10.1088/1748-9326/9/9/094006

Rémy, S., Kipling, Z., Flemming, J., Boucher, O., Nabat, P., Michou, M., Bozzo, A., Ades, M., Huijnen, V., Benedetti, A., Engelen, R., Peuch, V.H., Morcrette, J.J. (2019). Description and evaluation of the tropospheric aerosol scheme in the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS-AER, cycle 45R1). Geosci. Model Dev. 12, 4627–4659. https://doi.org/10.5194/GMD-12-4627-2019

Roberts, G., Wooster, M.J. (2021). Global impact of landscape fire emissions on surface level $PM_{2.5}$ concentrations, air quality exposure and population mortality. Atmos. Environ. 252, 118210. https://doi.org/10.1016/j.atmosenv.2021.118210

Roux, E., Ignotti, E., Bègue, N., Bencherif, H., Catry, T., Dessay, N., Gracie, R., Gurgel, H., Hacon, S. de S., Magalhães, M. de A.F.M., Monteiro, A.M.V., Revillion, C., Villela, D.A.M., Xavier, D., Barcellos, C. (2020). Toward an early warning system for health issues related to particulate matter exposure in Brazil: the feasibility of using global $PM_{2.5}$ concentration forecast products. Remote Sens. 12, 4074. https://doi.org/10.3390/RS12244074

Sahani, M., Zainon, N.A., Wan Mahiyuddin, W.R., Latif, M.T., Hod, R., Khan, M.F., Tahir, N.M., Chan, C.C. (2014). A case-crossover analysis of forest fire haze events and mortality in Malaysia. Atmos. Environ. 96, 257–265. https://doi.org/10.1016/j.atmosenv.2014.07.043

Sani, S. (1977). Aspects of air pollution climatology in the Kuala Lumpur-Petaling Jaya area, Malaysia. University of Canterbury. Christchurch, New Zealand.

Savage, N.H., Agnew, P., Davis, L.S., Ordóñez, C., Thorpe, R., Johnson, C.E., O'Connor, F.M., Dalvi, M. (2013). Air quality modelling using the Met Office Unified Model (AQUM OS24-26): Model description and initial evaluation. Geosci. Model Dev. 6, 353–372. https://doi.org/10.5194/GMD-6-353-2013

Show, D.L., Chang, S.C. (2016). Atmospheric impacts of Indonesian fire emissions: Assessing Remote Sensing Data and Air Quality During 2013 Malaysian Haze. Procedia Environ. Sci. 36, 176–179. https://doi.org/10.1016/J.PROENV.2016.09.029

Sobri, N.M., Wan Yaacob, W.F., Ismail, N.A., Malik, M.A.A., Rahman, R.A., Baser, N.A., Sukhairi, S.A.M. (2021). Predicting particulate matter ($PM_{2.5}$) in Malaysia using multiple linear regression and artificial neural network. J. Phys. Conf. Ser. 2084, 012010. https://doi.org/10.1088/1742-6596/2084/1/012010

Suradi, H., Khan, M.F., Sairi, N.A., Rahim, H.A., Yusoff, S., Fujii, Y., Qin, K., Bari, Md.A., Othman, M., Latif, M.T. (2021). Ambient levels, emission sources and health effect of $PM_{2.5}$-bound carbonaceous particles and polycyclic aromatic hydrocarbons in the city of Kuala Lumpur, Malaysia. Atmosphere 12, 549. https://doi.org/10.3390/ATMOS12050549

Tacconi, L. (2016). Preventing fires and haze in Southeast Asia. Nat. Clim. Change 6, 640–643. https://doi.org/10.1038/nclimate3008

United State Environmental Protection Agency (U.S. EPA) (2022). How Does PM Affect Human Health? https://www3.epa.gov/region1/airquality/pm-human-health.html (accessed 15 April 2022).

Varga-Balogh, A., Leelőssy, Á., Lagzi, I., Mészáros, R. (2020). Time-dependent downscaling of $PM_{2.5}$ predictions from CAMS air quality models to urban monitoring sites in Budapest. Atmosphere 11, 669. https://doi.org/10.3390/atmos11060669

Wong, K.S., Chew, Y.J., Ooi, S.Y., Pang, Y.H. (2021). Toward forecasting future day air pollutant index in Malaysia. J. Supercomput. 77, 4813–4830. https://doi.org/10.1007/S11227-020-03463-Z

Wu, C., Li, K., Bai, K. (2020). Validation and calibration of CAMS $PM_{2.5}$ forecasts using in situ $PM_{2.5}$ measurements in China and United States. Remote Sens. 12, 3813. https://doi.org/10.3390/rs12223813

Zainal, S., Zamre, N.M., Khan, Md.F. (2021). Emission level of air pollutants during 2019 pre-haze, haze, and post-haze episodes in Kuala Lumpur and Putrajaya. Malaysian J. Chem. Eng. Tech. 4, 137–154. https://doi.org/10.24191/MJCET.V4I2.14299

Zhou, Y., Wu, Y., Yang, L., Fu, L., He, K., Wang, S., Hao, J., Chen, J., Li, C. (2010). The impact of transportation control measures on emission reductions during the 2008 Olympic Games in Beijing, China. Atmos. Environ. 44, 285–293. https://doi.org/10.1016/j.atmosenv.2009.10.040