The set-up and evaluation of fine-scale data assimilation for the urban climate of Amsterdam

Koopmans, Sytse; van Haren, Ronald; Theeuwes, Natalie; Ronda, Reinder; Uijlenhoet, Remko; Holtslag, Albert A.M.; Steeneveld, Gert Jan

DOI
10.1002/qj.4401

Publication date
2022

Document Version
Final published version

Published in
Quarterly Journal of the Royal Meteorological Society

Citation (APA)
Koopmans, S., van Haren, R., Theeuwes, N., Ronda, R., Uijlenhoet, R., Holtslag, A. A. M., & Steeneveld, G. J. (2022). The set-up and evaluation of fine-scale data assimilation for the urban climate of Amsterdam. *Quarterly Journal of the Royal Meteorological Society*, *149*(750), 171-191. https://doi.org/10.1002/qj.4401

Received: 24 September 2021 | Revised: 6 September 2022 | Accepted: 10 November 2022

DOI: 10.1002/qj.4401

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

The set-up and evaluation of fine-scale data assimilation for the urban climate of Amsterdam

Sytse Koopmans[1] | **Ronald van Haren**[2] | **Natalie Theeuwes**[1,3] | **Reinder Ronda**[1] | **Remko Uijlenhoet**[4] | **Albert A. M. Holtslag**[1] | **Gert-Jan Steeneveld**[1]

[1]Meteorology and Air Quality Section, Wageningen University, Wageningen, The Netherlands

[2]Netherlands eScience Center, Amsterdam, Amsterdam, The Netherlands

[3]Royal Netherlands Meteorological Institute (KNMI), De Bilt, The Netherlands

[4]Hydrology Section, Delft University of Technology, Delft, The Netherlands

Correspondence
Sytse Koopmans, Meteorology and Air Quality Section, Wageningen University, PO box 47, 6700 AA Wageningen, The Netherlands.
Email: sytse.koopmans@wur.nl

Gert-Jan Steeneveld, Meteorology and Air Quality Section, Wageningen University, PO box 47, 6700 AA Wageningen, The Netherlands.
Email: gert-jan.steeneveld@wur.nl

Funding information
AMS Institute, Grant/Award Number: VIR16002; NWO, Grant/Award Number: 864.14.007; EPSRC (DARE), Grant/Award Number: EP/P002331/1

Abstract

Ongoing urbanization highlights the need for a better understanding and high resolution modelling of the urban climate. In this study, we combine rural observations by WMO surface stations, weather radar data and urban crowd-sourced observations with very fine-scale modelling efforts for Amsterdam, The Netherlands. As a model, we use the Weather Research and Forecasting (WRF) mesoscale model with 3D variational data assimilation at a 100-m resolution in the innermost model domain. In order to enable the assimilation of observations within the urban canopy, we develop a scheme to reduce urban temperature biases by adjusting urban fabric temperatures. The scheme is tested against independent urban observations for the summer month of July 2014 and specifically for a hot period and an extreme precipitation event. We find data assimilation reduces biases in temperature and wind speed. Within the city, the most significant improvement is the reduction of negative temperature biases during clear nights, which implies a better prediction of the Urban Heat Island (UHI). Concerning precipitation, the fractional skill score improves incrementally when additional observations are assimilated, and the largest impact is seen from the assimilation of weather radar observations.

KEYWORDS

1 | INTRODUCTION

Today's weather and climate models are progressing towards ever finer spatiotemporal scales, which is largely due to the increased computational capabilities that have become available. Recently, a global grid spacing of 1 km was employed for numerical weather prediction (NWP) models (Wedi *et al.*, 2020). In the transition to these high resolutions, the representation of cities in NWP models becomes crucial. The complex structures of cities — which are made up of buildings, streets and vegetation of variable sizes and shapes — require a different modelling approach than the typically more homogeneous rural surroundings. In mesoscale atmospheric models (resolution: 0.5–2 km), urban canopy-layer schemes are primarily used to model the radiation and energy balance of the urban surfaces in the modelling domain (Masson, 2000; Kusaka

et al., 2001; Martilli *et al.*, 2002; Chen *et al.*, 2011). Most of these parameterizations take the effects of shading, radiation trapping, drag and anthropogenic heat emissions into account. Despite evidence that the models perform well on a case-by-case basis (Miao *et al.*, 2009; Salamanca *et al.*, 2011), most models still show significant bias in the representation of the energy balance and deviate substantially due to the variety of parameterization schemes (Grimmond *et al.*, 2010; Loridan *et al.*, 2010). Therefore, steps are still required to improve the meteorological forecasting and hindcasting of the urban environment.

Improved model performance can be achieved by advancing the physics of the model and/or steering the model using observations, such as by applying data assimilation. The purpose of data assimilation is to accurately describe the state of the atmosphere using observations and model results. By using data assimilation, the model fields and observational fields are compared, and the model is steered in the direction of the observations based on the weighted uncertainties of both observations and model output. Various studies have shown the benefits of data assimilation for the improvement of mesoscale weather forecasts (e.g., Xiao and Sun, 2007; Huang *et al.*, 2009; Dee *et al.*, 2011; Adam *et al.*, 2016). Data assimilation is especially useful in hindcasting and forecasting extreme precipitation as there is rapid error growth the convective scale, which is a well-known challenge for atmospheric models (Sun *et al.*, 2014). Data assimilation techniques have also been applied in meteorological reanalysis data sets, such as Era-Interim and its successor, Era-5 (Dee *et al.*, 2011; Hersbach *et al.*, 2020). Reanalysis data consists of a complete and coherent collection of meteorological gridded data, which is the best-known atmospheric state computed from the model and observations. Era-Interim and Era-5 are commonly used in climate research and have many applications in renewable energy, for example in the siting of favourable areas for wind (Kalverla *et al.*, 2019) or solar power plants (Boilley and Wald, 2015; Lorenz and Barstad, 2016).

Current generation reanalysis products lack accuracy on relatively fine sub-kilometre urban scales. These fine scales are crucial to account for the variability of urban morphology, which directly affects meteorological processes and accordingly improves forecasts of routine weather variables, such as 2-m air temperature (Ching *et al.*, 2009; Ronda *et al.*, 2017). For the fine urban scale, specific studies reporting on data assimilation in the urban canopy or urban boundary layers are absent. In this paper, we explore whether data assimilation techniques have the potential to improve the representation of urban weather characteristics in a NWP system. This includes a novel approach to nudge modelled urban canopy temperatures with quality-controlled, crowd-sourced observations and independent verification against professionally installed urban weather stations.

In this study, we performed a hindcast for July 2014 for Amsterdam (The Netherlands) at a high resolution of 100 m, nested in coarser-resolution surrounding domains. The hindcast benefitted from high-resolution (urban) land-use maps in combination with a dense spatial network of crowd-sourced and professionally installed weather stations. In addition, Amsterdam's complex spatial structure, with its network of canals and proximity to large water bodies (i.e., the North Sea and Marker Lake), makes Amsterdam an excellent test site for model improvement. The month of July 2014 is chosen as test case because it includes several days with intense convective precipitation and a relatively warm, dry period, which are both critical conditions for water management and for human thermal comfort in cities. This work builds on the study by Ronda *et al.* (2017) that developed urban weather forecasts for Amsterdam at the same high-resolution grid spacing, but without data assimilation. Although they were able to reproduce the spatial variation in the urban air temperature, there was a cold bias of 1–2 K. This is a good incentive to explore whether data assimilation can assist in reducing this cold bias. The final goal is to create a reanalysis dataset of 15 years focused on the urban surroundings in order to investigate the urban climatology based on the evaluation of this study.

This paper is organized as follows. Section 2 presents the modelling set-up, a description of the assimilation approach for the urban canopy, the observations used for data assimilation and model verification. Section 3 presents the modelling results and Section 4 summarizes discussion and conclusions.

2 | METHODS

2.1 | Model set-up

This study used the Weather Research and Forecasting (WRF) model version 3.7.1 (Skamarock *et al.*, 2005; Powers *et al.*, 2017). WRF is a non-hydrostatic numerical model that allows for nesting to represent step-by-step the large-scale circulation towards micro-scale transport (Chen *et al.*, 2011). We built on the study by Ronda *et al.* (2017) and used four one-way nested computational domains with gradually increasing resolutions (12.5 km, 2.5 km, 500 m, 100 m, see Figure 1a,b). The European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis data (ERA-Interim) was used as initial and boundary conditions because Era-5 was not fully released when we start performing data assimilation test runs. At the innermost domain 100-m resolution, parks, ponds and runways at Schiphol airport were visible and,

consequently, their specific impact is visible in the model simulation. The basic WRF configuration is in accordance with the urban weather forecasts of Ronda *et al.* (2017). This set-up included 72 vertical eta-levels and made use of the Noah land surface model (Ek *et al.*, 2003). The Grell–Freitas cumulus scheme (Grell and Freitas, 2014) was applied to the outer domain (12.5 km resolution). As Planetary Boundary Layer Scheme, Yonsei University (YSU) was used for the three coarsest domains (Hong *et al.*, 2006). The inner domain turbulence was partly resolved and partly modelled with the Smagorinsky first-order closure scheme. Compared to the set-up of Ronda *et al.* (2017) our innermost domain was enlarged so Schiphol Airport was included (Figure 1). In addition, various data assimilation techniques were applied using the WRFDA module (Barker *et al.*, 2012), which will be described in Section 2.2.

As the urban scheme within WRF, we used the Single-Layer Urban Canopy Model (SLUCM) from Kusaka *et al.* (2001). The most relevant input parameters are listed in Table 1. The physical properties of buildings — including albedo, thermal conductivity and the thickness of roads, walls and roofs — are uniform for all buildings in Amsterdam. These values were determined by calculating averages from typical housing types in different construction year intervals (Agentschap, 2011; De Waag, 2021). The calculated albedo values were checked and marginally adjusted (<0.01) with albedo calculations from remote-sensing bands from the Landsat satellite (Liang, 2001; USGS, 2021). Other building properties — such as mean, area-weighted and standard deviation building height — were grid-cell-specific and adopted from Ronda *et al.* (2017) using the NUDAPT approach (National Urban Data and Access Portal Tool) (Ching *et al.*, 2009) (Figure 1c). Together with other grid-cell-specific geomorphological statistics — such as frontal area index and plan area fraction — it is possible to perform neighbourhood-specific energy balance calculations for the urban canyon. This leads to a more accurate representation of meteorological variables in the urban area (Zhang *et al.*, 2020). For completeness, the urban fraction is displayed (Figure 1d) to demonstrate the urban density differences in the study area.

Following Monaghan *et al.* (2014), the value of the 'Akanda'-parameter was modified from default 1.29 (Kanda *et al.*, 2007) to 0.4 (Table 1). This parameter affects the ratio of the roughness length for heat and the roughness length of momentum, and it enhances sensible heat exchange between the urban canopy and atmosphere above. We simulated and compared two simulations with the default (0.4) and modified (1.29) Akanda-parameter (1.29) to grasp how sensitive this modification was. This had a similar effect on root mean squared error (RMSE) performance of net radiation and sensible heat flux as albedo changes of roof and wall, with 0.09 and 0.15 respectively (Loridan *et al.*, 2010). For the anthropogenic heat flux (Table 1), we accounted for energy consumption, gas consumption (CBS, 2018a), traffic scaled for Amsterdam's population density (Klok *et al.*, 2010) and metabolic rate estimates (Sailor & Lu, 2004).

2.2 | Data assimilation

Data assimilation aims to describe the state of the atmosphere in a statistical manner by steering model fields in the direction of observations by a limiting cost function and accounting for uncertainties in the model results and observations (Barker *et al.*, 2012). The observations used for data assimilation consisted of three types:

1 Two-meter air temperature and air pressure observations from WMO-certified SYNOP stations in The Netherlands, Germany and the UK. These observations have a resolution and accuracy of 0.1 K and 0.1 hPa (KNMI, 2010; Met Office, 2020; DWD, 2022), which is in line with the WMO requirements (WMO, 2018).
2 Volume radar data from the C-band Doppler weather radar as operated by KNMI (Royal Netherlands Meteorological Institute) in The Netherlands (Beekhuis and Holleman, 2008). As product, both radial (wind) velocity and reflectivities were used.
3 Temperature observations from personal weather stations available from the Wunderground database (Weather Underground, 2022), where most brands have an accuracy of <0.3 K even under high exposure to solar radiation, while a few may have a radiation error of up to several degrees (Bell *et al.*, 2013). Therefore additional data selection is applied (see below in Section 2.2.1).

In addition, two types of observations were not used for data assimilation, but as independent evaluation of the model performance:

1 Temperature observations from the Amsterdam Atmospheric Monitoring Supersite (AAMS; Ronda *et al.*, 2017) have an accuracy of less than 0.3 K (Bell *et al.*, 2013).
2 Rain gauge-adjusted radar data (Overeem *et al.*, 2009b). This record has a bias of less than 0.4 mm for the 24-hr precipitation sum.

In the research set-up, three types of data assimilation were applied sequentially in four simulations. The abbreviated experiment names are given between brackets and were used as reference in the following sections:

- No data assimilation (NO-DA)

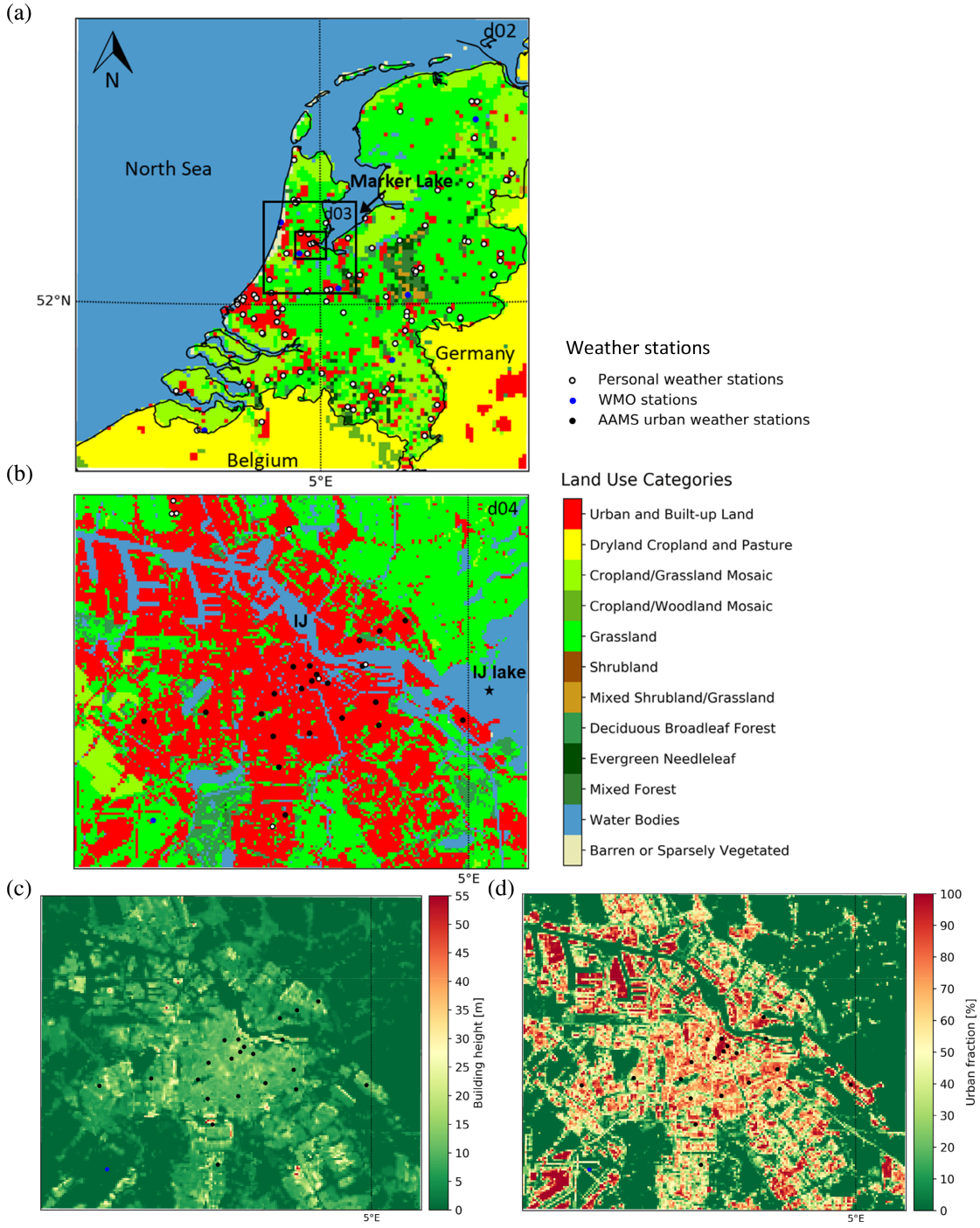

FIGURE 1 (a) Land-use map of the Dutch model domain (domain 2 with domains 3 and 4 nested). The black rectangles represent the four nested domains (100–500–2,500–12,500 m). The white dots indicate the 96 personal weather stations, and the blue dots indicate the 7 WMO stations: Schiphol (WMO code: 6240), Wijk aan Zee (WMO code: 6257), De Bilt (WMO code: 6260), Deelen (WMO code: 6275), Eelde (WMO code: 6280), Westdorpe (WMO code: 6319) and Volkel (WMO code: 6375) used for the model evaluation. (b) Land use map of the inner model domain (domain 4) with 22 independent Amsterdam Atmospheric Monitoring Supersite (AAMS) urban weather stations used for the evaluation (black dots). The black star is an excluded station. The personal weather stations are indicated with white dots. (c, d) Mean building height and urban (built) fraction, including 22 independent AAMS stations as black dots [Colour figure can be viewed at wileyonlinelibrary.com]

- WMO stations data assimilation (WMO-DA)
- WMO stations data assimilation + radar data assimilation (RADAR-DA)
- WMO stations data assimilation + radar data assimilation + urban nudging using personal weather stations (URBAN-DA)

TABLE 1 Configurations used in the Single-Layer Urban Canopy Model (SLUCM) (Kusaka *et al.*, 2001)

TS_scheme	4-layer model
CH_scheme	Narita scheme for the convective heat transfer coefficients within urban canopies
Akanda (exchange variable)	0.4 Monaghan *et al.* (2014), parameter introduced by Kanda *et al.* (2007)
Thickness roof layers	0.23 m
Thickness of building wall	0.179 m
Thermal conductivity of road	0.4 $J \cdot s^{-1} \cdot K^{-1}$
Thermal conductivity of wall	0.4004 $J \cdot s^{-1} \cdot K^{-1}$
Thermal conductivity of roof	0.48 $J \cdot s^{-1} \cdot K^{-1}$
Heat capacity of road	1.4 $MJ \cdot K^{-1}$
Heat capacity of wall	1.3 $MJ \cdot K^{-1}$
Heat capacity of roof	1.65 $MJ \cdot K^{-1}$
Surface albedo of road	0.2
Surface albedo of wall	0.2
Surface albedo of roof	0.175
Roughness length for momentum over ground (road) and roof	0.01 m
Roughness length for momentum over wall	0.0001 m
Maximum anthropogenic heat flux	38.5 $W \cdot m^{-2}$ (average 21.4 $W \cdot m^{-2}$)
Hourly weights of maximum anthropogenic heat flux, starting from 0000 to >2300 UTC	{0.16 0.13 0.08 0.07 0.08 0.26 0.67 0.99 0.89 0.79 0.74 0.73 0.75 0.76 0.82 0.9 1.0 0.95 0.68 0.61 0.53 0.35 0.21 0.18}

The data assimilation process worked as follows. First, the observations from the WMO-certified stations and volume radar data were assimilated every two hours within the WRF module WRFDA 3.8.1 using variational data assimilation (3D-Var). This is presented in the left panel of Figure 2. Second, within each two-hourly 3D-Var update cycle, an inner loop runs until convergence is reached in one data assimilation update cycle (Figure 2, right panel), with a maximum of 400 iterations. The final analysis was used by the WRF model to start the next 2-hr forecast (Figure 2, left panel), after which a 3D-Var update cycle was applied again.

The steps within one data assimilation update cycle are further explained. The 3D-Var solves (iteratively) a prescribed cost function by minimizing differences between analysis and observation/first guess from the background error model, while being weighted by estimates of predefined observation and background errors (Skamarock *et al.*, 2005; Yang *et al.*, 2015). The background error was determined separately, following the so-called 'NMC method' (Parrish and Derber, 1992). Herein, the model variance was estimated at certain time steps by comparing two overlapping 24-hr forecasts with 12 hr shifted lead times. More specifically, forecast 1, starting on day 1 at 0000 UTC and forecast 2, starting on day 1 at 1200 UTC are compared with day 2 at 0000 UTC, both on the largest domain 1 and at leaps of 12 hours.

Finally, the lateral boundaries and the lower boundaries — such as soil moisture/temperature and sea surface temperatures — were updated from ERA-Interim every six hours. These updates of the lower boundaries are important to keep the model boundaries within realistic reanalysis values throughout the monthly run and will be crucial to maintaining consistency for the 15-year reanalysis later on.

From the WMO observations, we assimilated 2-m air temperature and air pressure. Wind was not assimilated since pressure gradients were well-correlated with wind speed. Surface humidity was also not assimilated, because it showed no improved skill in the data assimilation for this case (July 2014). Otherwise, the liquid water in the atmosphere was adjusted through the radar data assimilation in case of precipitation in the model or radar observations.

For the radar data assimilation, we used the C-band Doppler weather radar located in De Bilt (52.099°N 5.176°E). As product, both radial (wind) velocity and reflectivities were used. Data from the lowest elevation angle (0.3°) were discarded since this elevation contained

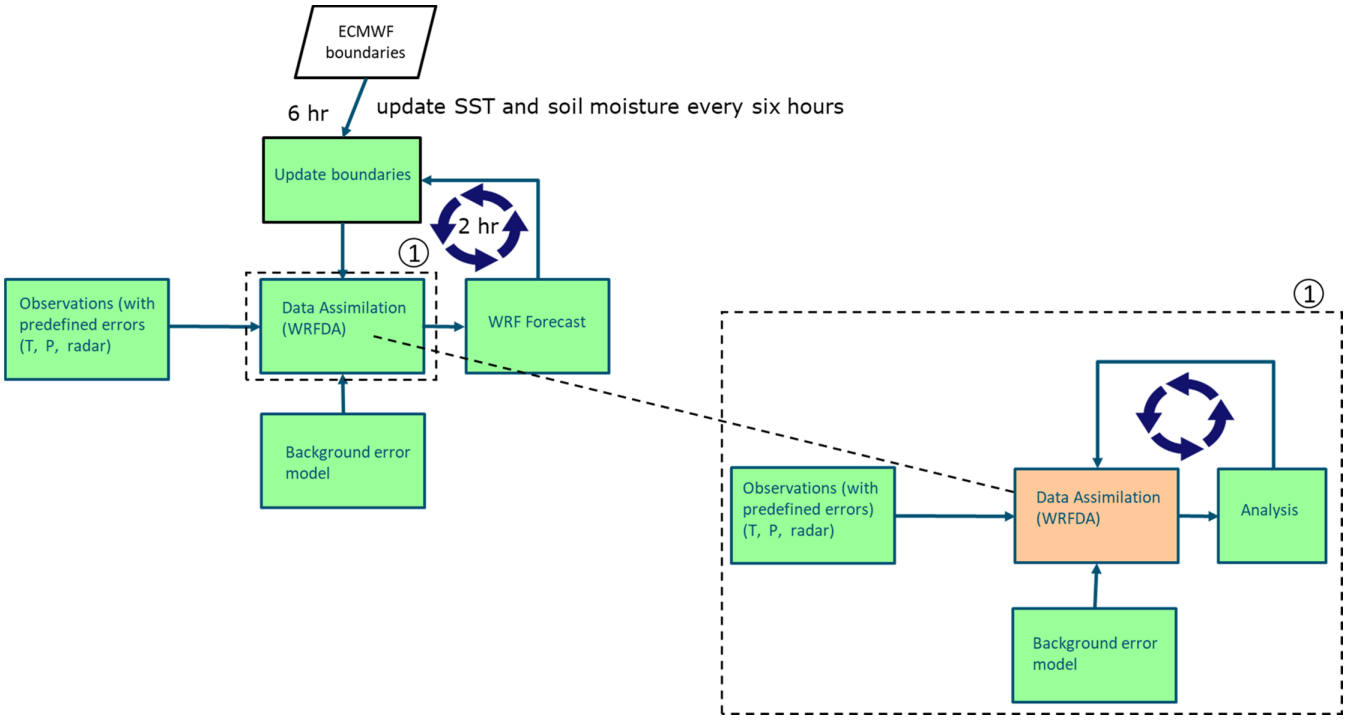

FIGURE 2 Flowchart of 3D-Var update cycle (left) and the internal model cycles (right) for Weather Research and Forecasting Data Assimilation (WRFDA) [Colour figure can be viewed at wileyonlinelibrary.com]

a disproportionate amount of clutter due to nearby buildings. From the remaining elevations, an average variance of 2 dBZ was calculated for the reflectivity employing the parameters of the volume scanning (Beekhuis and Holleman, 2008) and the method of Chumchean *et al.* (2003), see Equation (1). The standard deviation of the radial velocity was supplied with the radar product and is in the order of 2–3 $m \cdot s^{-1}$ (Holleman, 2008).

$$\sigma_Z^2 = 0.13\frac{R}{P} \tag{1}$$

σ_Z^2 is the reflectivity measurement error variance (dB^2); R is the observation range from the radar (km); P is the number of pulses per range bin.

Finally, the reflectivities and radial wind velocities were interpolated to model resolutions. The heights and locations of the C-band Doppler radar observations were calculated with the WRADLIB module (Heistermann *et al.*, 2013). This module converted the data from a polar to a Cartesian representation, taking refraction into account.

WRFDA offers different methods for radar data assimilation: it directly assimilates reflectivity, or first converts to rainwater mixing ratios before data assimilation is applied. The latter is known as the 'indirect method' (Wang *et al.*, 2013). This method has the advantage that the specific humidity remains more realistic in the whole atmosphere including better 2-m dew-point temperatures. We adopted the method of Wang *et al.* (2013) as it has the most constraints for specific humidity in the atmosphere, which should also lead to a better representation of clouds. For further details about this method, we refer to Wang *et al.* (2013).

Referring to data assimilation of observations taken in urban areas, we noticed that the canopy scheme SLUCM did not interact well with WRFDA. This is because WRFDA needs control variables, including pressure and temperature on model levels. These model levels, starting from approximately 50 m above ground level, are present above the urban canopy and are therefore not part of SLUCM. The latter is crucial because the control variables should be in balance with energy fluxes from the urban fabric (i.e., roads and walls). As such, we propose a new method that 'nudges' urban canopy temperatures within the canopy scheme, through adjustments in the wall and road temperatures of the scheme (see Section 2.2.2). To do this, a temperature model bias first had to be determined in the city by using personal weather stations (Section 2.2.1). The personal weather stations were not used to correct other meteorological variables, such as wind. Wind was not used for data assimilation for urban areas, since the spatial and time representativeness of wind measurements at one given moment is too limited for personal weather stations (Droste *et al.*, 2020).

2.2.1 | Model bias correction with personal weather stations

The estimation of the temperature bias differs between personal weather stations and WMO weather stations. While WMO stations are sited according to well-defined standards and obtain a high degree of spatial representativeness, this is not the case for personal weather stations in built-up areas (Bell *et al.*, 2013). The urban environment is very heterogeneous on a local scale, which results in microclimates (Heusinkveld *et al.*, 2014), and therefore different personal weather stations might show little spatial correlation on close distances. This is confirmed by the inter- and intra-local climate zone variability of a large number of personal weather stations in Berlin (Fenner *et al.*, 2017). Local characteristics, such as north- or south-facing gardens and the difference between paved areas and lawns show large variability in diurnal cycles of the 2-m air temperature urban heat island effect (UHI) (Oke, 2006). The uniform signal of quality-controlled personal weather stations was extracted by using a statistical model that depends on the weather at that time. This statistical model excludes specific weather station variability and intends to remove the cold bias in urban areas for WRF simulations without DA (Ronda *et al.*, 2017). The urban canopy temperatures were calculated with a revised SLUCM scheme of (Theeuwes *et al.*, 2014), which applies the stability correction between the first model level (~50 m) and the 2-m level instead of between the first model level and the roughness length of heat. This revision was also used in the set-up of the former study of Ronda *et al.* (2017).

This urban canyon temperature at the 2-m level does not represent yet an urban temperature typical urban grid cell with an urban fraction lower than 1, which consists of a natural fraction (e.g., green backyards, parks and, typically for Amsterdam's centre: water). As in Ronda *et al.* (2017), urban grid cells in WRF consist of a paved urban canopy fraction (buildings and streets) and a separate natural fraction. Energy fluxes and air temperature were calculated separately for the urban tile in SLUCM and for the vegetated tile. A combined area-weighted urban-natural temperature, indicated by urban fraction, was calculated and considered as a representative 2-m temperature for urban grid cells. Hence, the difference between this model temperature and personal weather stations ΔT_c was the response value in the statistical model in Equation (2).

$$\Delta T_c = \alpha * L\downarrow + \beta * U + \gamma \tag{2}$$

ΔT_c is the difference in urban temperature between personal weather stations from the statistical regression model and WRF model (°C); $L\downarrow$ is the WRF modelled surface downward long-wave radiation ($W \cdot m^{-2}$); U is the WRF-modelled 10-m wind speed ($m \cdot s^{-1}$); and α, β and γ are regression coefficients.

The statistical model consists of a linear regression between response value ΔT_c and explanatory variables wind speed and downward long-wave radiation. The fit (derivation of regression coefficients) was determined every two hours on domain 2, covering The Netherlands and downscaled to the other domains. This equation was then applied to all urban grid cells. The estimated ΔT_c was used to correct the model bias, and this is further described in Section 2.2.2 following next.

Regarding the chosen explanatory variables wind speed and long-wave radiation, the ΔT_c were largest on clear sky nights. Thus, personal weather stations facing cloudy conditions showed generally smaller temperature variations. As such, the surface downward long-wave radiation was a suitable variable to correct the model temperature state. Wind was also an important factor, as a high wind speed mixes the air and therefore usually reduces the modelled UHI. On the contrary, urban characteristics like urban fraction and aspect ratio are insignificant in the statistical model. The aspect ratio was taken into account in the implementation of the temperature nudging in the urban canopy model (see Section 2.2.2). The long-wave radiation and wind variables were retrieved from the model at the location of the personal weather stations.

The quality control of personal weather observations contains a number of steps. The full list of the used 96 personal weather stations is included as Supplementary Material (Appendix S1). First, some personal weather station brands do not have proper radiation shields (Bell *et al.*, 2015). During daytime, only Davis Vantage stations were used, since we were confident in their relatively small radiation error. The remaining stations can still report unreliable values due to effects of wrong placement or malfunctions. The number of crowdsourced personal weather stations and the long period of time made it unfeasible to inspect, inquire after or trace these stations and determine whether they met proper placement criteria. Therefore, additional criteria were applied to the selection of acceptable stations for the statistical model:

- The observed temperature should not deviate from the corresponding WRF model grid cell by more than 5°C
- After the previous selection, a median is determined from the observed values of the remaining personal weather stations
- The remaining personal weather stations are discarded if their values deviate from the median ± 2 × standard deviation

- Depending on the number of remaining observations:
  - Equation (2) was fitted when there are enough observations, which was set at more than 5
  - For three or four observations, the median was recalculated
  - With fewer than three observations, the number was too small to statistically correct the model
- If the fitted statistical model was not significant (p value > 0.1), the median was used to nudge the model.

Note that Meier *et al.* (2017) used a less strict $3\times$ standard deviation as threshold for omitting personal weather stations. However, the employed number of stations (Berlin) was substantially larger in their study than available for Amsterdam. The temperatures resulting from the urban data assimilation are verified against 22 independent urban weather stations (AAMS stations) within domain 4. The station on a small island in the IJ lake was excluded (see black star in Figure 1b) because it is non-urban and shows an atypical diminished diurnal cycle above the water surface compared to land stations. The temperature sensors of the AAMS stations are shielded (Decagon VP-3) and ventilated and mounted at 4 m on lantern posts (de Vos *et al.*, 2020).

2.2.2 | Nudging urban temperatures

The correction, ΔT_c, was applied in the urban canopy model SLUCM through a nudging approach and served as forcing for the model bias correction. This ΔT_c engages within SLUCM's original canopy temperature on the level of the roughness length of heat rather than the revised 2-m canyon temperature of Theeuwes *et al.* (2014). This choice is described after the explanation of the nudging concept.

We note that the nudged canopy temperatures could not preserve the effect of nudging until the next data assimilation cycle because no change in heat storage in the canopy was realized. Therefore, we adjusted the urban fabric temperatures in such a way that it was consistent with ΔT_c as derived from Equation (2). This included both walls and roads, which have a large heat storage. Altering the temperature of the urban fabric was also more effective in altering energy fluxes, which were fed back to the atmospheric model levels. Hence, the ΔT_c was redistributed to temperature corrections of the wall and road layers, while maintaining the physical formulation of the resistance scheme in Kusaka *et al.* (2001) (see Appendix A). Thereby, the average diurnal temperature amplitude (difference maximum and minimum daily temperature) in each layer of urban fabric determined the proportion of the temperature correction applied. SLUCM consists of four road and wall layers. Smaller amplitudes belonging to deeper layers received a smaller temperature correction and vice versa. So, urban fabric layers at the surface respond faster to temperature changes than the deeper layers imposed by the nudging approach. Hence, these often-raised temperatures for wall and road surfaces increase the urban canopy temperature and the air temperature of the first model level approximately 50 m above ground level, via an increased sensible heat flux (see Figure A1 in the Appendix A). The average diurnal cycle amplitudes were determined for all urban grid cells in the four-day warm period of July 2014. Subsequently, amplitude factors were derived for the inner layers denoting the diurnal cycle amplitude relative to the largest diurnal cycle amplitude on the wall and road surfaces (Table 2).

Before changes could be applied to the four wall and road layers, ΔT_c was translated to a difference in wall surface and road surface temperatures $\Delta T_{w,g}$ (the suffix g standing for 'ground' is used to be consistent with variable names in WRF). This was calculated with Equation (3), which is based on the SLUCM resistance equations of Kusaka *et al.* (2001), and the complete derivation can be found in the Appendix A. The temperature increments for wall and road layers were calculated by multiplying the $\Delta T_{w,g}$ with the accompanying amplitude factors [see Equation (6) and Table 2]. In addition, in accordance with the amplitude determination for walls and roads, the soil temperatures were also adjusted for the vegetated fraction of urban cells. This is because personal weather stations are often located in green backyards

TABLE 2 Diurnal cycle amplitude of temperature for wall, road and soil layers relative to the diurnal cycle amplitude of the skin (surface) temperature. Wall layers consist of four equally divided 0.0448 m thick layers. Road and soil layers increase with depth: 0.05 m, 0.25 m, 0.50 m, 0.75 m

	Layer 1	Layer 2	Layer 3	Layer 4
Wall (amp_w)	0.709	0.503	0.357	0.253
Road (amp_g)	0.734	0.157	0.007	
Soil (amp_s)	0.499	0.015		

rather than over paved surfaces.

$$\Delta T_{w,g} = \Delta T_c \left(1 + \frac{w}{(2h+w)} \frac{C_{ah}}{C_w}\right) \tag{3}$$

$$C_{ah} = \frac{\kappa u_*}{\psi_h} \tag{4}$$

$$C_w = C_g = \left\{ \begin{matrix} 7.51 U_c^{0.78} & \left(U_c > 5\ \mathrm{m \cdot s^{-1}}\right) \\ 6.15 + 4.18 U_c & \left(U_c \leq 5\ \mathrm{m \cdot s^{-1}}\right) \end{matrix} \right\} \tag{5}$$

(Kusaka *et al.*, 2001). ΔT_c is the difference in urban temperature between model and personal weather stations from the statistical regression model (predicted variable) (°C); $\Delta T_{w,g,s}$ is the wall, road surface and soil temperature (°C); h is the wall height (m); w is the road width (m); κ is the Karman constant (0.4); u_* is the friction velocity ($\mathrm{m \cdot s^{-1}}$); ψ_h is the Monin–Obukhov stability term (dimensionless); $C_{w,g}$ is the heat transfer coefficient from wall and road to the canopy ($\mathrm{m \cdot s^{-1}}$); C_{ah} is the heat transfer coefficient from the canyon to the overlying atmosphere ($\mathrm{m \cdot s^{-1}}$); and U_c is the urban canopy wind speed ($\mathrm{m \cdot s^{-1}}$).

$$\Delta Tlayer_{w,g,s} = \Delta T_{w,g,s} * amp_{w,g,s} \tag{6}$$

The magnitude of $\Delta T_{w,g}$ in Equation (3) depends on the heat exchange between the urban fabric and urban canopy, and the heat exchange between the urban canopy and first model layer on top. These heat exchanges are dependent on wind speed and stability (see Equations 4 and 5). Thus, less stable conditions allow for larger adjustments to road and wall layers with changing ΔT_c. The street width to building height ratio plays a significant role. In narrow urban canyons, the total heat exchange decreases between the urban fabric and the overlying atmosphere (first model level at ~50 m), and this leads to smaller adjustments to road and wall layers with the same ΔT_c. An example of a temperature adjustment of the outer wall is given in Figure 3. Here it is visible that the denser centre of Amsterdam has a smaller adjustment, because of a larger aspect ratio.

We chose to apply ΔT_c in Equation (3) on the level of the roughness length corresponding the original canopy temperatures, because the newly defined 2-m canopy temperature by Theeuwes *et al.* (2014) is an output variable and not implemented in the iterative energy flux calculations within SLUCM. As original canopy temperatures in SLUCM are closer to walls and roads, they are more biased to wall and road surface temperatures. Therefore, adjustments to roads and walls $\Delta T_{w,g}$ were smaller with the current method than when ΔT_c was applied on the 2-m level (the 2-m level is further away from the road and this

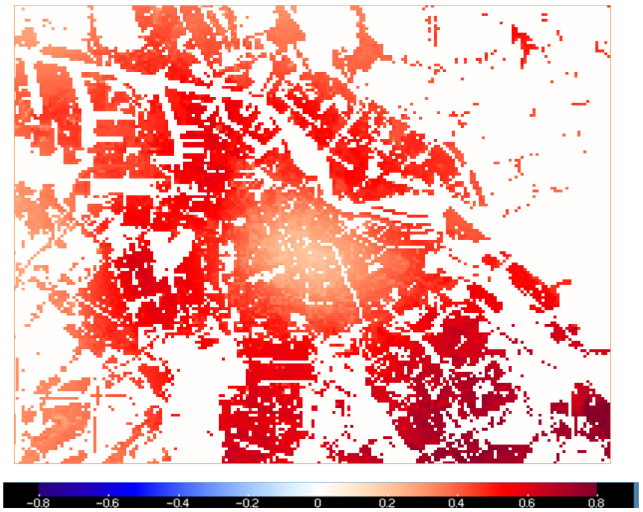

FIGURE 3 Example of outer wall temperature correction ΔT_w (in °C) in the evening on 2 July 2014 at 2200 UTC [Colour figure can be viewed at wileyonlinelibrary.com]

therefore results in a smaller heat transfer coefficient ratio $\frac{C_{ah}}{C_w}$ in Equation 3). Thus, the magnitude of the model correction, resulting from the adjustments of walls and roads is generally on the safe side, which is also a benefit since nudging relies heavily on observations.

2.3 | Model evaluation methods

Different verification methods were applied to evaluate the performance of consecutive data assimilation steps. We used the bias and the root mean squared error to evaluate the model performance for air and dew-point temperature, wind speed and sea-level pressure. This evaluation was performed for seven selected WMO stations across the country (Figure 1a). These stations provided observations suitable for application in the data assimilation, and were also used for model verification one and two hours after the data assimilation cycling. This time delay was considered sufficient to avoid serious dependence between cycling and verification. Nonetheless, we made use of independent weather stations for the verification of urban air temperatures.

$$\mathrm{FSS} = \frac{1}{N_{\mathrm{days}}} \sum_{i=1}^{N_{\mathrm{days}}} \mathrm{FSS}(d) \tag{7}$$

$$\mathrm{FSS}(d) = 1 - \frac{\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[O(d)_{i,j} - M(d)_{i,j}\right]^2}{\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} O(d)_{i,j}^2 + \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} M(d)_{i,j}^2} \tag{8}$$

$O(d)_{i,j}$ is the observed fraction precipitation grid cells above threshold value over window i, j; and $M(d)_{i,j}$ is the model fraction precipitation grid cells above threshold value over window i, j.

To verify the spatial representation of precipitation, the fractional skill score (FSS) was calculated for daily accumulated precipitation of the consecutive data assimilation runs for domain 2 that covers The Netherlands. The FSS is a neighbourhood-based verification method for evaluating the spatial accuracy of precipitation (Roberts and Lean, 2008) and has been applied in real case verification studies (Ebert, 2009; Mittermaier *et al.*, 2013; Skok and Roberts, 2016; Imhoff *et al.*, 2020; Khain *et al.*, 2020). In order to use it, the gauge-adjusted radar data (Overeem *et al.*, 2009a) were transformed with an average weighted interpolation scheme (Met Office, 2018) to match the model resolution. The FSS calculates the skill of precipitation forecasts by comparing fractions of precipitation over a threshold in a certain spatial window between model and observations, see Equations (7) and (8). This window was moved over the entire domain, and the FSS was determined as the spatial average. A FSS of 1 indicates perfect skill and 0 no skill. Evaluating fractions determined over a window has the advantage that the location differences of showers are not double-penalized.

A 90th percentile (90p) threshold was chosen to evaluate the spatial accuracy of extreme precipitation (Skok and Roberts, 2016), and an average threshold to evaluate the spatial accuracy in general. Since evaluating spatial accuracy is the aim of FSS, the model and observation fractions are unbiased (Mittermaier *et al.*, 2013). Hence, threshold values were separately determined for the model and observations. For 90p, this means that precipitation fractions comprised both 10% of domain 2 for model and observations. Correcting for bias was important because a substantial bias would deteriorate the FSS, and positive biases would have had larger impacts than negative biases because of the $M(d)^2_{i,j}$ in the denominator of Equation (8). The window or neighbourhood size is 50 km and, thus, consists of 20×20 cells with a resolution of 2.5 km according to domain 2. This window size is small enough for forecast purposes (Roberts and Lean, 2008).

3 | RESULTS

3.1 | Model evaluation

In the model evaluation, we first present a monthly time series of the nearby KNMI weather station Schiphol Airport (WMO code 6240) (Figure 4) and statistics from other pre-selected WMO stations spread across the country (Figure 5). The diurnal cycles of air temperatures in the NO-DA run and WMO-DA run are well represented for most days. However, WMO-DA was able to remove or substantially reduce cold biases on clear nights, clearly visible on 13 and 29 July 2014. In NO-DA the maximum hourly cold bias amounted 3.5°C on 13 July 2014, which virtually disappeared in WMO-DA, and was reduced from 5.3 to 2.6°C on 29 July 2014. Maximum temperatures are also better modelled in the WMO-DA in terms of biases, as can be seen on 11 July (2.5 vs. 1.0°C), 16 July (3.2 vs. 0.8°C), 27 July (3.3 vs. 0.0°C) and 29 July (3.3 vs. 0.1°C). On the other hand, 9 July and 21 July are clear examples of cases that are both too warm in NO-DA and WMO-DA during the daytime, which is related to a surplus of solar radiation.

Note that no data assimilation was applied for solar downwelling radiation at the surface. Over the whole month, the cloud cover was underestimated in the NO-DA and WMO-DA run, resulting in an overestimation of solar downwelling radiation. In addition, both WRF simulations have a surplus of surface solar radiation on clear days. The modelled surface pressure correspond reasonably well with the model runs and observations, especially in timing and for high-pressure situations. Some low-pressure system occurrences remain too deep in the WMO-DA run. The WMO-DA run partly restores this around 21 July and 27 July 2014. The largest hourly deviations on these days reduced from 6.7 to 3.4 hPa and from 3.2 to 1.4 hPa respectively. The modelling system was unable to resolve the minimum pressure on 9 July (recall the deviations in solar radiation and temperature). Both WRF simulations had difficulties in correctly representing precipitation events, especially high-intensity precipitation. The model runs were generally drier than the observations, with the largest deviation for NO-DA. Comparing precipitation at a single point is not very representative due to the spatial variability of convective precipitation, therefore we perform a spatial analysis in Section 3.3.

The statistics plots (Figure 5) show significantly lower RMSE for the WMO-DA run compared to the base NO-DA run for the rural WMO stations. The bias is also smaller overall for the WMO-DA simulation, although this is less pronounced. The sea-level pressure (SLP) and wind (FF) show smaller biases, (44% and 49% respectively) in WMO-DA. For the dew-point temperature a small positive bias (~0.3°C) was introduced, but this is accompanied with by lower RMSE. The RADAR-DA run only has small changes compared to the WMO-DA run. RADAR-DA adjusts clouds and humidity on different model levels, which has a slightly negative effect on the RMSE of the dew-point temperature (and indirectly on SLP). The bias, on the other hand, is improved for the dew-point temperature and the air temperature. Straightforwardly, the URBAN-DA mimics the RADAR-DA, since the additional

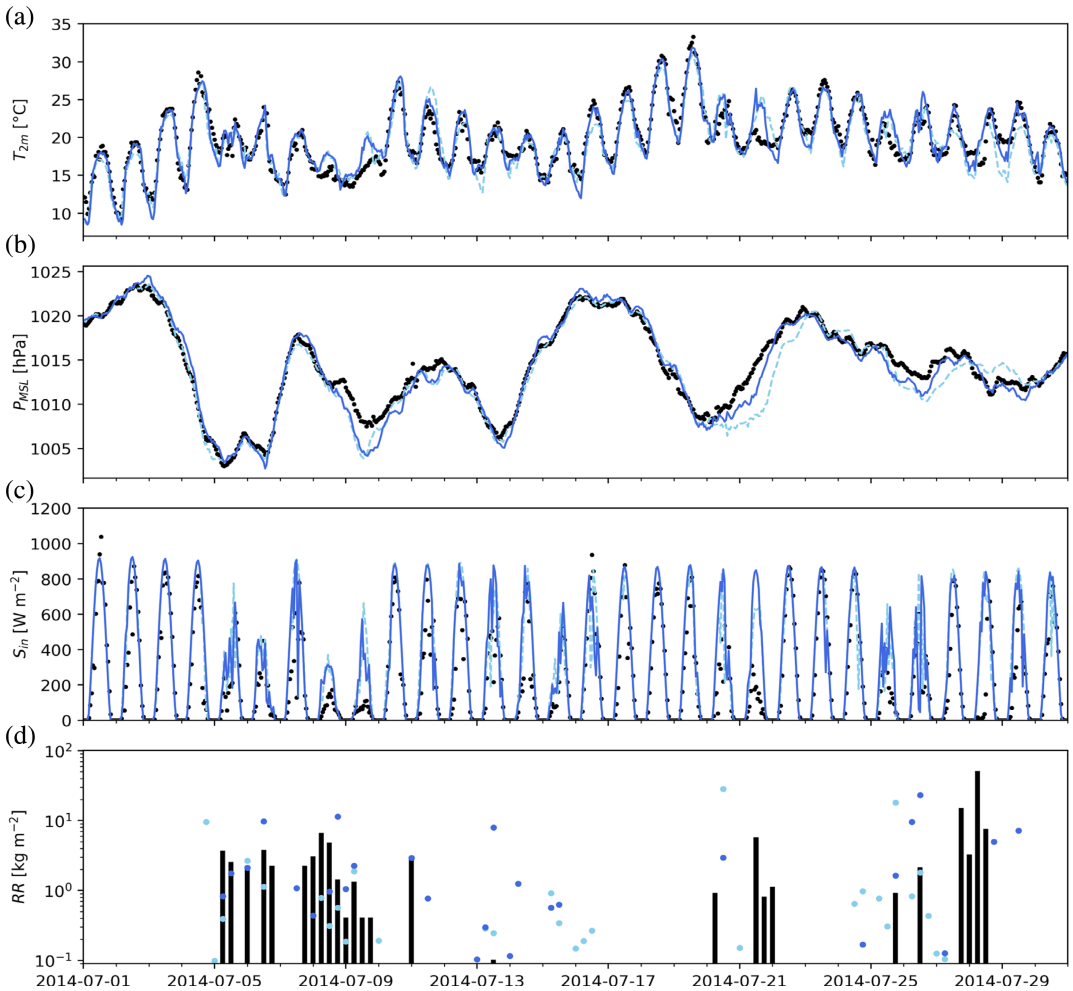

FIGURE 4 Times series of modelled and observed (a) 2-m temperature, (b) mean sea-level pressure, (c) downwelling solar radiation, and (d) precipitation at Schiphol airport weather station (52.301°N, 4.774°E, WMO code 6240) for July 2014. The black dots and bars indicate the observations, the dark blue solid lines and dots the WMO stations data assimilation (WMO-DA) run and the light blue dashed lines and dots the no data assimilation (NO-DA) run [Colour figure can be viewed at wileyonlinelibrary.com]

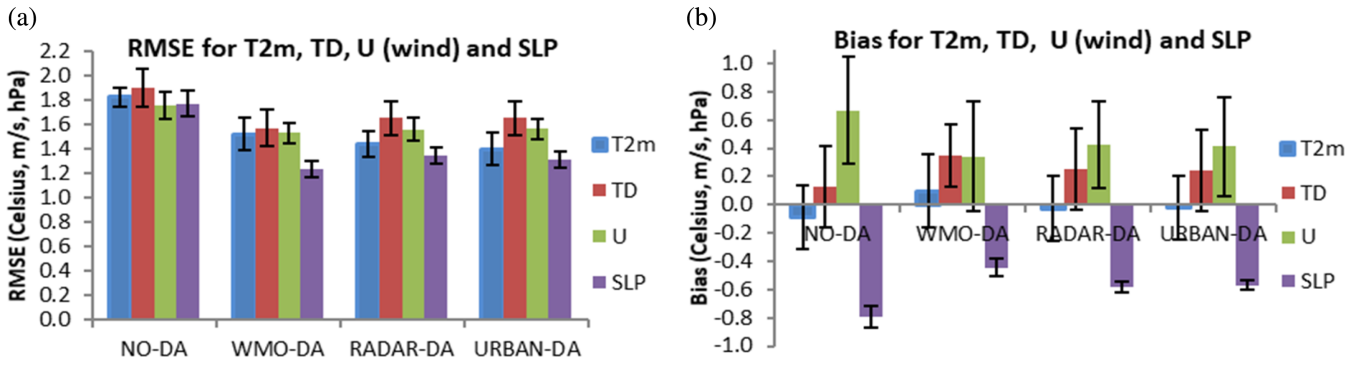

FIGURE 5 Root mean squared error (RMSE) (a) and bias b) of 2-m temperature (T2m), dew-point temperature (TD), wind speed (U) and sea-level pressure (SLP), for the data assimilation model runs at seven WMO stations (see Figure 1a). The analysis was done at hourly intervals between 3 July (two days spin up) and 29 July 2014. The bars indicate the standard deviation [Colour figure can be viewed at wileyonlinelibrary.com]

nudging in the urban canopy has no substantial impact on the rural station outside.

Noting the significant differences between simulations and observations (paired Student's t-test, $\alpha = 0.05$), the RMSE is slightly larger for dew-point temperature (TD) and wind speed (U). The RMSE for air temperature is also smaller (just significant for URBAN-DA and just not significant for RADAR-DA). The URBAN-DA should not have a substantial effect on the evaluated variables on the rural stations, which is confirmed by the large similarities of RADAR-DA and URBAN-DA.

3.2 | Heatwave case

The nudging of urban air temperatures (URBAN-DA) of personal weather stations was also evaluated for a warm episode between 16 July and 20 July 2014. Urban air temperatures are underestimated by WRF on clear nights for the simulations without urban nudging (NO-DA). The nudging process increases the nocturnal temperatures and decreases the bias with respect to independent urban observations (Figure 6a). In addition, the simulated hours in the warm period are closer to the 1:1 line in Figure 6b. With additional urban data assimilation (URBAN-DA) the bias decreases from −0.9°C to −0.2°C and the RMSE decreases from 1.5°C to 1.0°C. During daytime, only small adjustments in urban air temperature are made by the nudging process, since the atmospheric boundary layer is well mixed. The underestimation in maximum temperature is reduced to less than 1°C in URBAN-DA for all four days. The WMO-DA effectively reduces the cold bias in urban areas during daytime (Figure 6a), which means the cold bias is not caused by urban-specific processes. WMO-DA strongly reduces the cold bias for Schiphol (Figure 4a) and other stations during the warm period and affects urban areas via interpolated model fields. In addition, during daytime urban model temperatures in WMO-DA correlate better with rural WMO temperature observations compared to the night, because of daytime-enhanced turbulent mixing and predominant temperature advection from the countryside.

Supplementary to the time series results, contour plots are shown for Amsterdam after urban nudging to provide a spatial overview (Figure 7). The minimum temperatures are clearly higher for the densely built areas in Amsterdam for July 2014. These areas are either largely paved or have a small sky-view factor. The fine resolution allows runways at Schiphol Airport to be distinguished as warmer areas. The canals and the large water body, the IJ, can be distinguished as cooler places than the adjacent urban areas. The parks and rural parts adjacent to the city are relatively cool and only slightly warmer than the rural areas further away (~0.3°C). Therefore, it can be stated that (intra-)urban heat advection does not play a substantial role for modelled minimum temperatures, and corresponding results apply for average temperatures in the model context at this grid spacing. In the average monthly urban temperatures, temperature differences between cities and the rural counterparts are smaller, due to the lower daytime UHI. Water surfaces are relatively cool at daytime. Lastly, maximum temperatures (not shown) are only 0.5–1.0°C warmer for urban areas compared to the rural environment.

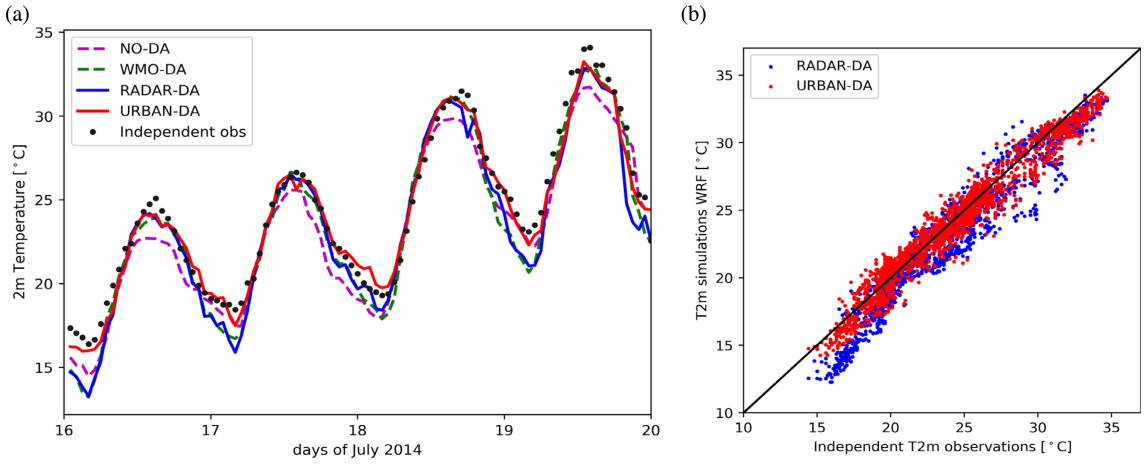

FIGURE 6 (a) Time series of median 2-m air temperature of all model runs with consecutive data assimilation at observation grid-points and median observations from independent Amsterdam Atmospheric Monitoring Supersite (AAMS) urban weather stations in Amsterdam for the warm period of 16–20 July 2014. (b) Scatter plot of modelled WMO stations data assimilation + radar data assimilation (RADAR-DA) and WMO stations data assimilation + radar data assimilation + urban nudging using personal weather stations (URBAN-DA) and observed 2-m air temperature of 16–20 July 2014 [Colour figure can be viewed at wileyonlinelibrary.com]

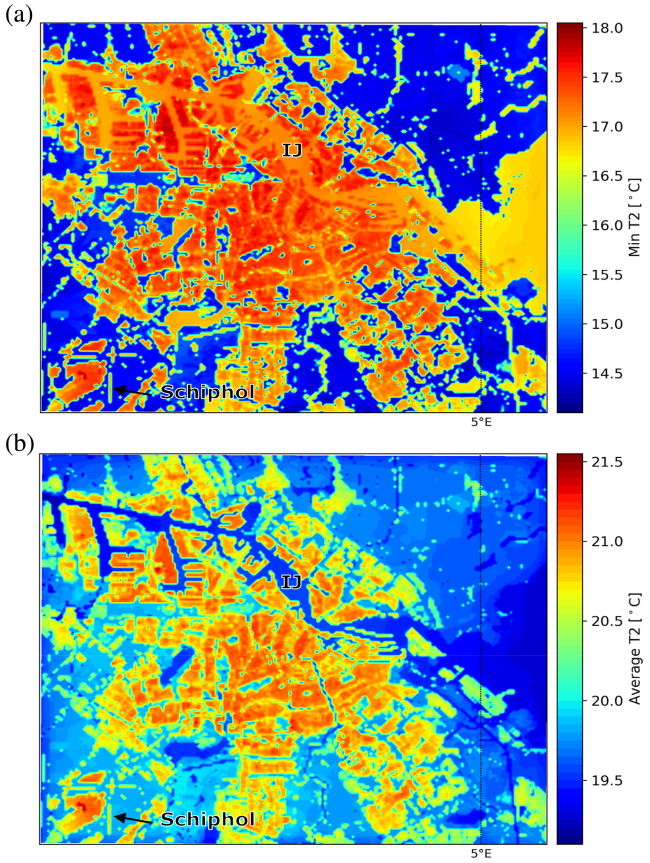

FIGURE 7 (a) Daily minimum temperature, and (b) monthly mean modelled 2-m air temperature (WMO stations data assimilation + radar data assimilation + urban nudging using personal weather stations — URBAN-DA) on hourly intervals for 3–31 July 2014 [Colour figure can be viewed at wileyonlinelibrary.com]

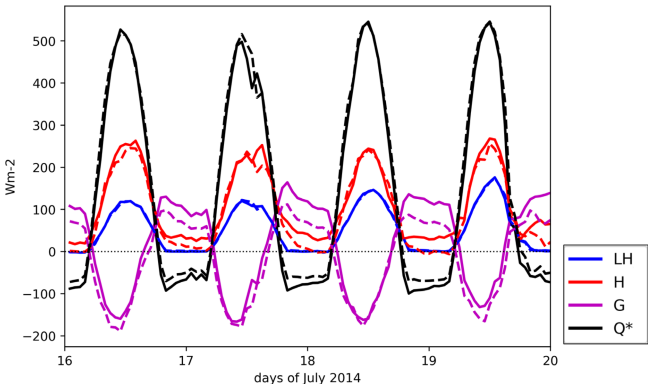

FIGURE 8 Modelled energy balance for WMO stations data assimilation + radar data assimilation + urban nudging using personal weather stations (URBAN-DA) (full line) and WMO stations data assimilation + radar data assimilation (RADAR-DA) (dashed line) for the average of the locations of Amsterdam Atmospheric Monitoring Supersite (AAMS) urban stations, for the four-day warm period (corresponding to Figure 6). *H* denotes sensible heat flux, *LH* denotes latent heat flux, *G* denotes storage heat flux, and *Q** denotes the net radiation [Colour figure can be viewed at wileyonlinelibrary.com]

Here, we illustrate the impact of the urban nudging technique on the modelled surface energy balance during the four-day warm period (Figure 8). The net radiation is calculated as the sum of the modelled energy fluxes. The energy fluxes are typically associated with an urban surface, best seen in the storage heat flux. The maximum storage heat flux in the morning amounts to 150 $W \cdot m^{-2}$, while the soil heat flux for a rural grassland area typically amounts to 30 $W \cdot m^{-2}$ (Cleugh and Oke, 1986).

For URBAN-DA, it appears that increasing the temperature of wall, road and soil layers leads to an increased release of stored heat and an increase in the sensible heat flux, which is most pronounced at night (Figure 8). On clear nights in the warm period, the sensible heat flux is typically 30 $W \cdot m^{-2}$ larger for URBAN-DA compared to RADAR-DA. The storage heat flux is approximately 50% larger. The difference can be explained by the increased skin temperature and an associated increase in outgoing long-wave radiation. During the day, the difference in energy balance between URBAN-DA and RADAR-DA is smaller. At solar noon, around 47% of the net radiation is used for sensible heat flux (RADAR-DA 46% and URBAN-DA 48%). This H/Q^* ratio is close to that of the urban sites of Basel, approximately 51% (Christen and Vogt, 2004), and lies within the range of 75th percentile July energy fluxes of downtown London (Kotthaus and Grimmond, 2014). In a high-rise district of Marseille, the H/Q^* ratios are around 60% (Grimmond *et al.*, 2004), which matches the denser urban sites of Amsterdam with low green fraction (48%–61% for RADAR-DA and 51%–64% for URBAN-DA). On the other hand, Kato and Yamaguchi (2007) shows for the city of Nagoya considerably lower H/Q^* ratios, which is related to low sky-view factors and related obstruction of solar radiation of tall buildings.

The storage heat flux consists of 29% and 25% of the net radiation (G/Q^*) for RADAR-DA and URBAN-DA at solar noon. These results correspond with the observed storage heat flux for the urban districts of Basel (Christen and Vogt, 2004). The maximum G/Q^* ratio is modelled in the morning at 0700-0800 UTC and amounts to 41% for RADAR-DA and 37% for URBAN-DA and these values are in the range of the different urban districts of Basel (Christen and Vogt, 2004), and is close to that for different cities in North America (Grimmond and Oke, 1999). Thus, this means that, in general, the energy partitioning in the urban area is in agreement with typical values in the literature both for URBAN-DA and RADAR-DA.

3.3 | Extreme precipitation

In order to evaluate the modelled extreme precipitation for each of the data assimilation steps, 28 July 2014 was selected as an example of extreme precipitation over Amsterdam, which resulted in precipitation sums over 100 mm in 24 hr and flooding in parts of the city (Figure 9). The synoptic pattern is characterized by little background flow at 850 hPa, and hence the development of showers is strongly bound to a local circulation influenced by land–sea interactions and the diurnal cycle. Below, we evaluate the NO-DA, WMO-DA and RADAR-DA runs (Figure 9) and describe their differences using time series and a skill score designed for precipitation (Figure 10).

In the NO-DA run, a high-pressure area develops over the North Sea, which counteracts warm moist air masses on land. This leads to a weak northerly flow over The Netherlands transporting relatively cool and dry air, which suppresses deep convection. In addition, the air flow is diffluent, which is associated with descending air motions. The WMO-DA run was able to create a convergence zone, leading to heavy showers in the moist unstable air. This is related to better representation in surface pressure (Figure 4). The maximum daily precipitation amounts (127 mm) are close to those using the radar estimates (116 mm) (Figure 9). Unfortunately, the location of this convergence zone and extreme rainfall is approximately 150 km to the south of what was observed by the radar. In the afternoon, the gauge-adjusted radar product had the largest precipitation intensities in the east of the country. WMO-DA is able to generate convective precipitation in this area, although this convective precipitation is more widespread and dispersed with lower intensities. The development of storms in the east seen in the gauge-adjusted radar is not present in WMO-DA either.

The RADAR-DA run is able to simulate the heavy convective precipitation at the approximate location compared to the gauge-adjusted radar. With the assimilated

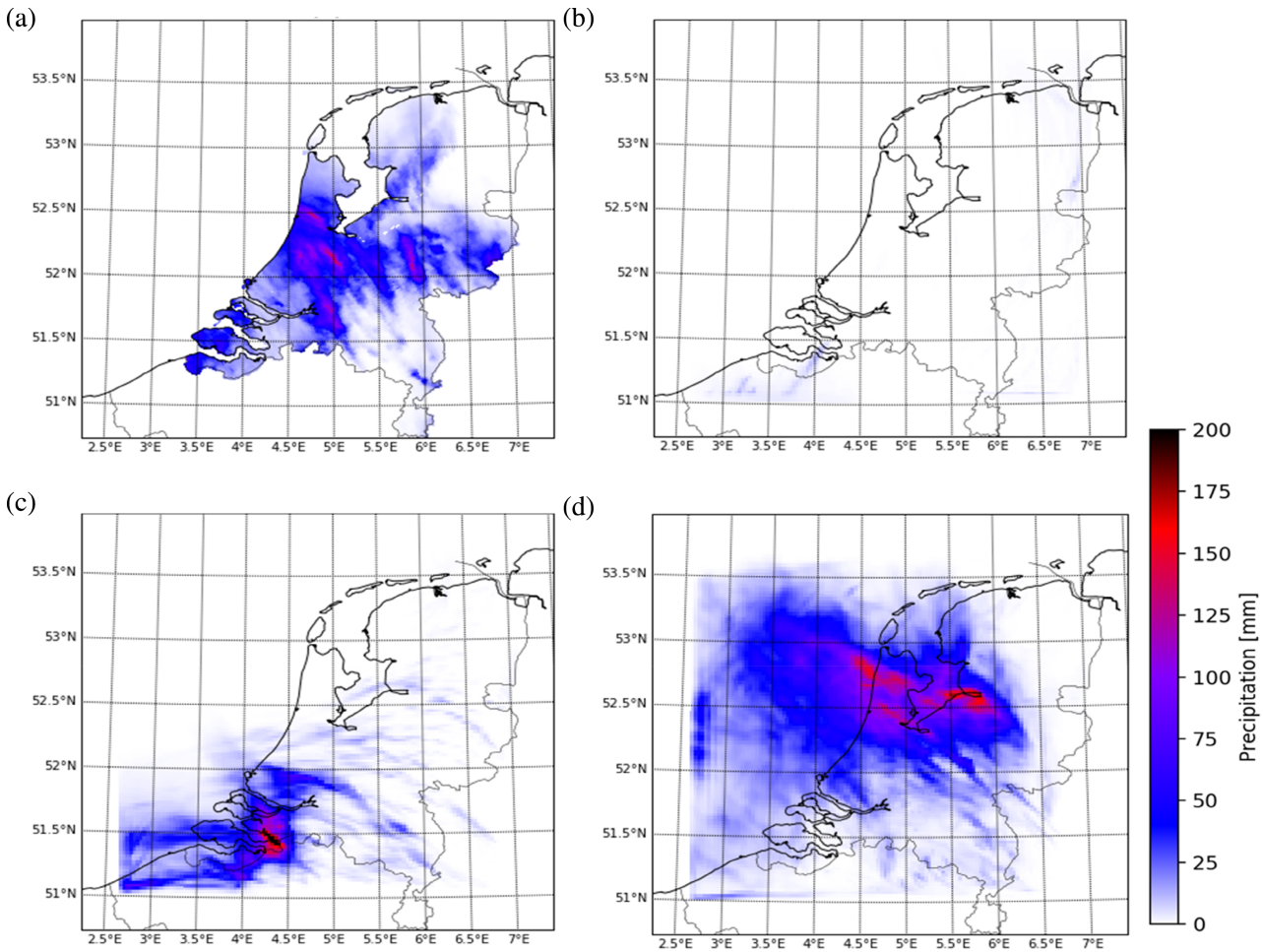

FIGURE 9 Cumulative daily precipitation for 28 July 2014, 0000 UTC — 29 July 2014, 0000 UTC for no data assimilation (NO-DA) (b), WMO stations data assimilation (WMO-DA) (c), and WMO stations data assimilation + radar data assimilation (RADAR-DA) (d), and radar observations (a), for domain 2. The rain observations represent the rain gauge-adjusted radar data [Colour figure can be viewed at wileyonlinelibrary.com]

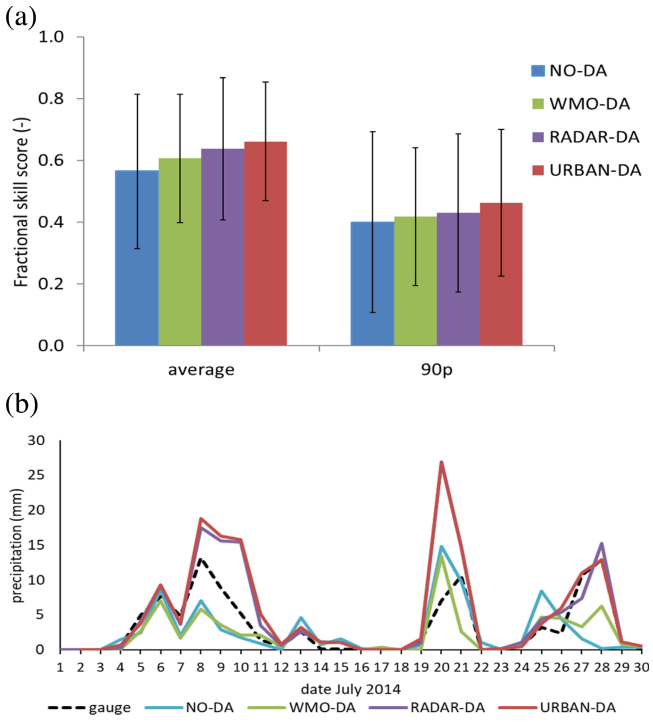

FIGURE 10 (a) Average fractional skill scores for the consecutive data assimilation runs for days with precipitation for domain 2. The bars indicate the standard deviation of the average Fractional Skill Score (FSS) and 90th percentile (90p) FSS of the different days. (b) Daily precipitation sums for the consecutive data assimilation runs and gauge-adjusted radar as reference for domain 2. Precipitation outside the land surfaces of The Netherlands is excluded in the model simulations according the gauge-adjusted radar data in Figure 9a [Colour figure can be viewed at wileyonlinelibrary.com]

convective precipitation, the model is able to create the convergence zone in the middle of the country. Therefore, 28 July 2014 shows a clearly improved performance utilizing radar data assimilation. Until 28 July 2014 1300 UTC, the radar run resembles the gauge-adjusted radar data. The northwest–southeast-oriented convergence line is well represented by the RADAR-DA run (and URBAN-DA). Thereafter, the most active precipitation in the convergence line remains in the northwest in the radar run, while the gauge-adjusted radar data show activation in the east of the country on this convergence line. At the end of the afternoon, the location of convective precipitation diverges between RADAR-DA and the gauge-adjusted radar. The convergence line itself is still present in the RADAR-DA run (and the URBAN-DA run). On other days with convective precipitation, the improvement of consecutive data assimilation steps is less distinct.

The performance of multiple precipitation days is reflected in the FSS (Figure 10a), and daily accumulated precipitation (Figure 10b). The FSS shows better scores for average daily precipitation and extreme precipitation (90p) with each additional data source used for assimilation, although the differences are small (Figure 10a), and there is a large spread between individual days. In the Discussion, Section 4, we further examine which FSS value should be considered as useful for predicting value. Figure 10b presents the daily averaged precipitation amounts for the Netherlands domain. On average, the runs without radar data assimilation (NO-DA and WMO-DA) produce too little precipitation. These runs resemble each other for most days, except for the last decade of July, including 28 July. Occasionally, WMO-DA has substantial impact on the simulated precipitation fields leading to better precipitation sums. RADAR-DA and URBAN-DA, on the contrary, overestimate the precipitation and differ extensively from the runs without data assimilation. From 24 July onwards, the runs with radar data assimilation (RADAR-DA and URBAN-DA) resemble the gauge-adjusted radar much more closely than the runs without (NO-DA and WMO-DA). In the other rainy periods, the magnitude of the biases is similar but reversed between these runs (July 7–12), or the radar data assimilation runs show separately an overestimation of precipitation (July 20–21). The URBAN-DA simulation is very similar to the RADAR-DA run regarding precipitation, since urban data assimilation only impacts urban temperatures at daytime marginally, which therefore has a small impact on the development of convective precipitation.

4 | DISCUSSION

4.1 | Urban data assimilation

The inclusion of observations from personal weather stations in reanalysis products is new and challenging. Other literature rather focuses on operational forecasts with assimilation of citizen observations (4D-Var). For example, recently, a huge number of NetAtmo crowdsourced weather stations were assimilated in a regional weather model (Sgoff *et al.*, 2022). Due to ensemble forecasts (local ensemble transform Kalman filter), an observation error and model error could be derived for the quality control of personal weather stations. Furthermore, the meteorological services in Norway (MET Norway) introduced citizen weather observations in the postprocessing of their forecasts (Nipen *et al.*, 2020). A Kalman filtering approach was applied to spatially interpolate innovations ('innovations' here means the difference between observation and model just before the data assimilation cycling) by the observation to form an analysis, which is the best estimate for current conditions. A remarkable similarity between Kalman filtering and the urban nudging presented in the current work is that the so-called 'innovation'

at a station is less or not spread across meteorological features, such as fronts. In our fit, wind and clouds (typically for fronts) generally cause smaller innovations as the UHI is suppressed during such conditions. Kalman filtering can determine the lack of correlation of innovations across such features in the analysis, preventing single innovations at one observing station affecting areas with different weather. Computational 4D-var data assimilation requires linear approximations used in a tangent linear model (Courtier *et al.*, 1994). In a complex urban boundary layer, this linear model error evolution holds for districts with high aspect ratios like downtown Hong Kong, while other urban geometries exhibit such linear model error evolution to a lesser extent, leading to lower accuracy (Ngan and Lo, 2017). Lastly, the data assimilation of personal weather station observations in operational forecasts is investigated by several European National weather institutes such as the Danish and Finnish Meteorological Institutes and the Deutscher WetterDienst (Hintz *et al.*, 2019).

4.2 | Radar data assimilation

The runs without radar data assimilation appear to have too little precipitation. It is worthwhile also reviewing the Era-Interim dataset, which was used as boundary condition. The near-surface conditions of temperature and humidity are captured well relative to their resolution since there are enough high-quality observations available for the reanalysis. However, the relative and specific humidity is likely too low for the layers between 925 hPa and 775 hPa, due to a known dry bias of radiosonde observations (Vaisala RS90/92) used in the Era-Interim (Simmons *et al.*, 2010). The dry bias is confirmed by comparing these type of radiosonde measurements with reference satellite-based measurement systems (GPSRO and GCOS) (Ladstädter *et al.*, 2015). This dry bias likely has a reduced effect on shower development and precipitation rates. At screen level (~1.5 m), the NO-DA run does not reveal a dry bias looking to the dew-point temperature verification (Figure 5), which is in accordance with Era-Interim showing no structural bias in humidity near the surface (Simmons *et al.*, 2010). So, among other processes (e.g., data assimilation of weather stations), too dry model levels higher up are not forcing a reduction of the humidity near the surface.

Large differences in FSS are observed between individual days displayed in the large standard deviations in Figure 10 (0.19–0.25 for average precipitation and higher for 90p). On many days there is not a clear signal that the model sufficiently adopts the location of precipitation observed by the radar. We hypothesize the model often has a non-supportive atmosphere for shower development. In that case, eventually adopted radar signals of showers die down quickly, so that they are not perceived in the model run. That non-supportive atmosphere is substantiated with the radiosonde-related dry biases (Ladstädter *et al.*, 2015). In addition, the related convection of the showers should be assimilated (via radial wind velocity) at an early stage of development (Bachmann *et al.*, 2019), which is not commonplace every day.

The FSS deviates between 0 (no skill) and 1 (perfect skill). To put the score in a perspective, a random forecast FSS_{random} equates the fraction of threshold exceedances (10% threshold exceedances is a FSS_{random} of 0.1). The average FSS_{random} is 0.31 and 90p FSS_{random} is 0.10 for days with precipitation in July 2014. Achieving this benchmark is very minimal performance. The literature (Roberts and Lean, 2008; Skok, 2015) suggests FSS should be at least higher than 0.5 to be useful, which is the case for our model results. Moreover, Roberts and Lean (2008) propose using $FSS_{uniform}$ as a skill parameter, which is defined as $FSS_{uniform} = 0.5 + f/2$. Herein, f is the fraction threshold exceedances for radar observations. The $FSS_{uniform}$ is 0.66 for the average FSS and 0.55 for the 90p FSS. According to Figure 10a, the $FSS_{uniform}$ criterium for URBAN-DA is just met for the average FSS (0.66) and not met for the 90p FSS (0.46). The improvement of fraction skill scores for the sequential DA runs is not significant since the simulation period is relatively short. In addition, Skok, (2015) obtained the FSS_{useful}, which is an analytical solution corresponding with a rain band, which, displaced, is half the size of a window. Although Skok (2015) obtained an objective benchmark, the FSS_{useful} is less interpretable in situations with multiple rain areas (see e.g., July 28, Figure 9a) that is less than the neighbourhood length apart, which also occur in our study. In addition, rain bands might be interrupted, since lighter precipitation patches within the rainband may not exceed the threshold. This especially applies for the 90p results.

5 | CONCLUSION

In this study, we investigated how WRF simulations can be improved by applying consecutive data assimilation steps (WRF 3D-Var) and nudging techniques for the urban environment in Amsterdam at a 100-m grid spacing. Assimilated observations consist of WMO synoptic weather observations, volume radar data and urban weather observations recorded by personal weather stations. Specifically for the urban canopy, a novel approach was developed to nudge temperatures with quality-controlled personal weather observations. Hindcasts are presented for July 2014 for Amsterdam, which inhibits both warm dry periods and days with extreme precipitation. We have

demonstrated that data assimilation of 2-m air temperature and pressure from WMO stations largely improve the background fields. The assimilation of weather radar observations shows slightly better performance in the location of precipitation fields. In addition, the data assimilation of personal weather station observations reduces the cold bias at night within the urban canopy, and it produces more realistic energy fluxes. The findings and best practices will be used to create a 15-year urban reanalysis for Amsterdam, which will increase knowledge in urban climatology and enable the tracing of trends in thermal comfort and extreme precipitation.

AUTHOR CONTRIBUTIONS

Sytse Koopmans: Formal analysis; investigation; methodology; validation; visualization; writing – original draft. **Ronald van Haren:** Investigation; methodology; software. **Natalie Theeuwes:** Investigation; methodology; writing – original draft; writing – review and editing. **Reinder Ronda:** Funding acquisition; resources; writing – review and editing. **Remko Uijlenhoet:** Writing – review and editing. **Albert A. M. Holtslag:** Funding acquisition; writing – review and editing. **Gert-Jan Steeneveld:** Conceptualization; funding acquisition; project administration; supervision; writing – review and editing.

ACKNOWLEDGEMENTS

The authors acknowledge funding from the NWO-eScience project 027.014.203 (ERA-urban) and from the AMS Institute (grant VIR16002) for supporting the Amsterdam Atmospheric Monitoring Supersite. The authors also acknowledge funding from UK EPSRC Data assimilation for the Resilient City (DARE) (grant EP/P002331/1). In addition, funding from NWO grant 864.14.007 was employed for this research.

ORCID

Sytse Koopmans https://orcid.org/0000-0002-7562-6302
Natalie Theeuwes https://orcid.org/0000-0002-9277-8551
Remko Uijlenhoet https://orcid.org/0000-0001-7418-4445
Albert A. M. Holtslag https://orcid.org/0000-0003-0995-2481
Gert-Jan Steeneveld https://orcid.org/0000-0002-5922-8179

REFERENCES

Adam, S., Behrendt, A., Schwitalla, T., Hammann, E. and Wulfmeyer, V. (2016) First assimilation of temperature lidar data into an NWP model: impact on the simulation of the temperature field, inversion strength and PBL depth. *Quarterly Journal of the Royal Meteorological Society*, 142(700), 2882–2896. https://doi.org/10.1002/qj.2875.

Agentschap N. L. (2011). Voorbeeldwoningen 2011: Bestaande bouw, Publicatie-nr. 2KPWB1034. https://www.rvo.nl/sites/default/files/bijlagen/4.%20Brochure%20Voorbeeldwoningen%202011%20bestaande%20bouw.pdf (in Dutch)

Bachmann, K., Keil, C. and Weissmann, M. (2019) Impact of radar data assimilation and orography on predictability of deep convection. *Quarterly Journal of the Royal Meteorological Society*, 145(718), 117–130. https://doi.org/10.1002/qj.3412.

Barker, D., Huang, X.-Y., Liu, Z., Auligné, T., Zhang, X., Rugg, S., Ajjaji, R., Bourgeois, A., Bray, J., Chen, Y., Demirtas, M., Guo, Y.-R., Henderson, T., Huang, W., Lin, H.-C., Michalakes, J., Rizvi, S., and Zhang, X. (2012) The weather research and forecasting Model's community Variational/ensemble data assimilation system: WRFDA. *Bulletin of the American Meteorological Society*, 93(6), 831–843. https://doi.org/10.1175/bams-d-11-00167.1.

Beekhuis, A., & Holleman, I. (2008). From pulse to product: Highlights of the digital-IF upgrade of the Dutch national radar network. Paper presented at the Proc. Fifth European Conf. of Radar Meteorology and Hydrology, ERAD 5, Helsinki, Finland. https://www.knmi.nl/kennis-en-datacentrum/publicatie/from-pulse-to-product-highlights-of-the-digital-if-upgrade-of-the-dutch-national-radar-network.

Bell, S., Cornford, D. and Bastin, L. (2013) The state of automated amateur weather observations. *Weather*, 68(2), 36–41. https://doi.org/10.1002/wea.1980.

Bell, S., Cornford, D. and Bastin, L. (2015) How good are citizen weather stations? Addressing a biased opinion. *Weather*, 70(3), 75–84. https://doi.org/10.1002/wea.2316.

Boilley, A. and Wald, L. (2015) Comparison between meteorological re-analyses from ERA-interim and MERRA and measurements of daily solar irradiation at surface. *Renewable Energy*, 75, 135–143. https://doi.org/10.1016/j.renene.2014.09.042.

Chen, F., Kusaka, H., Bornstein, R., Ching, J., Grimmond, C.S.B., Grossman-Clarke, S., et al. (2011) The integrated WRF/urban modelling system: development, evaluation, and applications to urban environmental problems. *International Journal of Climatology*, 31(2), 273–288. https://doi.org/10.1002/joc.2158.

Ching, J., Brown, M., Burian, S., Chen, F., Cionco, R., Hanna, A., et al. (2009) National urban database and access portal tool. *Bulletin of the American Meteorological Society*, 90(8), 1157–1168. https://doi.org/10.1175/2009BAMS2675.1.

Christen, A. and Vogt, R. (2004) Energy and radiation balance of a central European city. *International Journal of Climatology*, 24(11), 1395–1421. https://doi.org/10.1002/joc.1074.

Chumchean, S., Sharma, A. and Seed, A. (2003) Radar rainfall error variance and its impact on radar rainfall calibration. *Physics and Chemistry of the Earth, Parts A/B/C*, 28(1), 27–39. https://doi.org/10.1016/S1474-7065(03)00005-6.

Cleugh, H.A. and Oke, T.R. (1986) Suburban–rural energy balance comparisons in summer for Vancouver, B.C. *Boundary-Layer Meteorology*, 36(4), 351–369. https://doi.org/10.1007/BF00118337.

Courtier, P., Thépaut, J.N. and Hollingsworth, A. (1994) A strategy for operational implementation of 4D-Var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, 120(519), 1367–1387. https://doi.org/10.1002/qj.49712051912.

de Vos, L.W., Droste, A.M., Zander, M.J., Overeem, A., Leijnse, H., Heusinkveld, B.G., et al. (2020) Hydrometeorological monitoring using opportunistic sensing networks in the Amsterdam metropolitan area. *Bulletin of the American Meteorological Society*, 101(2), E167–E185. https://doi.org/10.1175/BAMS-D-19-0091.1.

De Waag. (2021). Construction years of buildings. Available at: https://code.waag.org/buildings/#52.38444.917111 [Accessed 12th February 2021].

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011) The ERA-interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656), 553–597. https://doi.org/10.1002/qj.828.

Droste, A.M., Heusinkveld, B.G., Fenner, D. and Steeneveld, G.J. (2020) Assessing the potential and application of crowdsourced urban wind data. *Quarterly Journal of the Royal Meteorological Society*, 146(731), 2671–2688. https://doi.org/10.1002/qj.3811.

DWD (2022). Meteorologische Sensoren und Systeme im Deutschen Wetterdienst. Available at: https://www.dwd.de/DE/derdwd/messnetz/sensorikundsysteme/_functions/Teasergroup/einsatz_sensoren_systeme.pdf?__blob=publicationFile&v=4 [Accessed 6th September 2022].

Ebert, E.E. (2009) Neighborhood verification: a strategy for rewarding close forecasts. *Weather and Forecasting*, 24(6), 1498–1510. https://doi.org/10.1175/2009WAF2222251.1.

Ek, M.B., Mitchell, K.E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., et al. (2003) Implementation of Noah land surface model advances in the National Centers for environmental prediction operational mesoscale eta model. *Journal of Geophysical Research: Atmospheres*, 108(D22), 8851. https://doi.org/10.1029/2002JD003296.

Fenner, D., Meier, F., Bechtel, B., Otto, M. and Scherer, D. (2017) Intra and inter "local climate zone" variability of air temperature as observed by crowdsourced citizen weather stations in Berlin, Germany. *Meteorologische Zeitschrift*, 26(5), 525–547. https://doi.org/10.1127/metz/2017/0861.

Grell, G.A. and Freitas, S.R. (2014) A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. *Atmospheric Chemistry and Physics*, 14(10), 5233–5250. https://doi.org/10.5194/acp-14-5233-2014.

Grimmond, C.S.B., and Oke, T.R. (1999). Heat storage in urban areas: Local-scale observations and evaluation of a simple model. *Journal of Applied Meteorology*, 38(7), 922–940. https://doi.org/10.1175/1520-0450(1999)038<0922:hsiual>2.0.co;2

Grimmond, C.S.B., Salmond, J.A., Oke, T.R., Offerle, B. and Lemonsu, A. (2004) Flux and turbulence measurements at a densely built-up site in Marseille: heat, mass (water and carbon dioxide), and momentum. *Journal of Geophysical Research: Atmospheres*, 109(D24), D24101. https://doi.org/10.1029/2004JD004936.

Grimmond, C.S.B., Blackett, M., Best, M.J., Barlow, J., Baik, J.J., Belcher, S.E., et al. (2010) The international urban energy balance models comparison project: first results from phase 1. *Journal of Applied Meteorology and Climatology*, 49(6), 1268–1292. https://doi.org/10.1175/2010JAMC2354.1.

Heistermann, M., Jacobi, S. and Pfaff, T. (2013) Technical note: an open source library for processing weather radar data (wradlib). *Hydrology and Earth System Sciences*, 17(2), 863–871. https://doi.org/10.5194/hess-17-863-2013.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. https://doi.org/10.1002/qj.3803.

Heusinkveld, B.G., Steeneveld, G.J., van Hove, L.W.A., Jacobs, C.M.J. and Holtslag, A.A.M. (2014) Spatial variability of the Rotterdam urban heat Island as influenced by urban land use. *Journal of Geophysical Research—Atmospheres*, 119(2), 677–692. https://doi.org/10.1002/2012JD019399.

Hintz, K.S., O'Boyle, K., Dance, S.L., Al-Ali, S., Ansper, I., Blaauboer, D., et al. (2019) Collecting and utilising crowdsourced data for numerical weather prediction: propositions from the meeting held in Copenhagen, 4–5 December 2018. *Atmospheric Science Letters*, 20(7), e921. https://doi.org/10.1002/asl.921.

Holleman (2008). Wind observations with Doppler weather radar WMO Technical Conference on Mete orological and Environmental Instruments and Methods of Observation, St. Petersburg, Russian Federation, 27–29 November 2008.

Hong, S.Y., Noh, Y. and Dudhia, J. (2006) A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review*, 134(9), 2318–2341. https://doi.org/10.1175/mwr3199.1.

Huang, X.-Y., Xiao, Q., Barker, D.M., Zhang, X., Michalakes, J., Huang, W., et al. (2009) Four-dimensional Variational data assimilation for WRF: formulation and preliminary results. *Monthly Weather Review*, 137(1), 299–314. https://doi.org/10.1175/2008MWR2577.1.

Imhoff, R.O., Brauer, C.C., Overeem, A., Weerts, A.H. and Uijlenhoet, R. (2020) Spatial and temporal evaluation of radar rainfall Nowcasting techniques on 1,533 events. *Water Resources Research*, 56(8), e2019WR026723. https://doi.org/10.1029/2019WR026723.

Kalverla, P., Steeneveld, G.J., Ronda, R.J. and Holtslag, A.A.M. (2019) Evaluation of three mainstream numerical weather prediction models with observations from meteorological mast IJmuiden at the North *Sea*. *Wind Energy*, 22(1), 34–48. https://doi.org/10.1002/we.2267.

Kanda, M., Kanega, M., Kawai, T., Moriwaki, R. and Sugawara, H. (2007) Roughness lengths for momentum and heat derived from outdoor urban scale models. *Journal of Applied Meteorology and Climatology*, 46(7), 1067–1079. https://doi.org/10.1175/jam2500.1.

Kato, S. and Yamaguchi, Y. (2007) Estimation of storage heat flux in an urban area using ASTER data. *Remote Sensing of Environment*, 110(1), 1–17. https://doi.org/10.1016/j.rse.2007.02.011.

Khain, P., Levi, Y., Shtivelman, A., Vadislavsky, E., Brainin, E. and Stav, N. (2020) Improving the precipitation forecast over the eastern Mediterranean using a smoothed time-lagged ensemble. *Meteorological Applications*, 27(1), e1840. https://doi.org/10.1002/met.1840.

Klok, L., ten Broeke, H., van Harmelen, T., Verhagen, H., Kok, H., & Zwart, S. (2010). Ruimtelijke verdeling van mogelijke oorzaken van het hitte-eiland effect. TNO Rapport (TNO-034-UT-2010-01229_RPT-ML), p.78. Available at: https://edepot.wur.nl/326309 (in Dutch)

Kotthaus, S. and Grimmond, C.S.B. (2014) Energy exchange in a dense urban environment—part I: temporal variability of

long-term observations in Central London. *Urban Climate*, 10, 261–280. https://doi.org/10.1016/j.uclim.2013.10.002.

KNMI (2010). Handbook for the Meteorological observation. Available at: https://cdn.knmi.nl/knmi/pdf/bibliotheek/knmipubDIV/handboekwaarnemingen.pdf [Accessed 15th August 2022].

Kusaka, H., Kondo, H., Kikegawa, Y. and Kimura, F. (2001) A simple single-layer urban canopy model for atmospheric models: comparison with multi-layer and slab models. *Boundary-Layer Meteorology*, 101(3), 329–358. https://doi.org/10.1023/A:1019207923078.

Ladstädter, F., Steiner, A.K., Schwärz, M. and Kirchengast, G. (2015) Climate intercomparison of GPS radio occultation, RS90/92 radiosondes and GRUAN from 2002 to 2013. *Atmospheric Measurement Techniques*, 8(4), 1819–1834. https://doi.org/10.5194/amt-8-1819-2015.

Liang, S. (2001) Narrowband to broadband conversions of land surface albedo I: Algorithms. *Remote Sensing of Environment*, 76(2), 213–238. https://doi.org/10.1016/S0034-4257(00)00205-4.

Lorenz, T. and Barstad, I. (2016) A dynamical downscaling of ERA-interim in the North Sea using WRF with a 3 km grid—for wind resource applications. *Wind Energy*, 19(10), 1945–1959. https://doi.org/10.1002/we.1961.

Loridan, T., Grimmond, C.S.B., Grossman-Clarke, S., Chen, F., Tewari, M., Manning, K., et al. (2010) Trade-offs and responsiveness of the single-layer urban canopy parametrization in WRF: an offline evaluation using the MOSCEM optimization algorithm and field observations. *Quarterly Journal of the Royal Meteorological Society*, 136(649), 997–1019. https://doi.org/10.1002/qj.614.

Martilli, A., Clappier, A. and Rotach, M.W. (2002) An urban surface exchange parameterisation for mesoscale models. *Boundary-Layer Meteorology*, 104(2), 261–304. https://doi.org/10.1023/A:1016099921195.

Masson, V. (2000) A physically-based scheme for the urban energy budget In atmospheric models. *Boundary-Layer Meteorology*, 94(3), 357–397. https://doi.org/10.1023/A:1002463829265.

Meier, F., Fenner, D., Grassmann, T., Otto, M. and Scherer, D. (2017) Crowdsourcing air temperature from citizen weather stations for urban climate research. *Urban Climate*, 19, 170–191. https://doi.org/10.1016/j.uclim.2017.01.006.

Met Office (2018). Iris: A Python library for analysing and visualising meteorological and oceanographic data sets. Exeter, Devon. Available at: https://scitools-iris.readthedocs.io/en/stable/ [Accessed 15th February 2021].

Met Office (2020), Midas Data User Guide for UK Land Observations. Available at: http://cedadocs.ceda.ac.uk/1465/1/MIDAS_User_Guide_for_UK_Land_Observations.pdf [Accessed 15th August 2022].

Miao, S., Chen, F., LeMone, M.A., Tewari, M., Li, Q. and Wang, Y. (2009) An observational and modeling study of characteristics of urban Heat Island and boundary layer structures in Beijing. *Journal of Applied Meteorology and Climatology*, 48(3), 484–501. https://doi.org/10.1175/2008JAMC1909.1.

Mittermaier, M., Roberts, N. and Thompson, S.A. (2013) A long-term assessment of precipitation forecast skill using the fractions skill score. *Meteorological Applications*, 20(2), 176–186. https://doi.org/10.1002/met.296.

Monaghan, A.J., Hu, L., Brunsell, N.A., Barlage, M. and Wilhelmi, O.V. (2014) Evaluating the impact of urban morphology configurations on the accuracy of urban canopy model temperature simulations with MODIS. *Journal of Geophysical Research: Atmospheres*, 119(11), 6376–6392. https://doi.org/10.1002/2013jd021227.

Ngan, K. and Lo, K.W. (2017) Linear error dynamics for turbulent flow in urban street canyons. *Journal of Applied Meteorology and Climatology*, 56(5), 1195–1208. https://doi.org/10.1175/JAMC-D-16-0173.1.

Nipen, T.N., Seierstad, I.A., Lussana, C., Kristiansen, J.r. and Hov, y. (2020) Adopting citizen observations in operational weather prediction. *Bulletin of the American Meteorological Society*, 101(1), E43–E57. https://doi.org/10.1175/BAMS-D-18-0237.1.

Oke, T.R. (2006) *Initial Guidance to Obtain Representative Meteorological Observations at Urban Sites*. IOM Rep. 81. Geneva: WMO.

Overeem, A., Holleman, I. and Buishand, A. (2009a) Derivation of a 10-year radar-based climatology of rainfall. *Journal of Applied Meteorology and Climatology*, 48(7), 1448–1463.

Overeem, A., Buishand, T.A. and Holleman, I. (2009b) Extreme rainfall analysis and estimation of depth-duration-frequency curves using weather radar. *Water Resources Research*, 45(10), W10424. https://doi.org/10.1029/2009WR007869.

Parrish, D.F. and Derber, J.C. (1992) The National Meteorological Center's spectral statistical-interpolation analysis system. *Monthly Weather Review*, 120(8), 1747–1763. https://doi.org/10.1175/1520-0493(1992)120<1747:Tnmcss>2.0.Co;2.

Powers, J.G., Klemp, J.B., Skamarock, W.C., Davis, C.A., Dudhia, J., Gill, D.O., et al. (2017) The weather research and forecasting model: overview, system efforts, and future directions. *Bulletin of the American Meteorological Society*, 98(8), 1717–1737. https://doi.org/10.1175/BAMS-D-15-00308.1.

Roberts, N.M. and Lean, H.W. (2008) Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*, 136(1), 78–97. https://doi.org/10.1175/2007mwr2123.1.

Ronda, R.J., Steeneveld, G.J., Heusinkveld, B.G., Attema, J.J. and Holtslag, A.A.M. (2017) Urban finescale forecasting reveals weather conditions with unprecedented detail. *Bulletin of the American Meteorological Society*, 98, 2675–2688. https://doi.org/10.1175/BAMS-D-16-0297.1.

Sailor, D.J. and Lu, L. (2004) A top–down methodology for developing diurnal and seasonal anthropogenic heating profiles for urban areas. *Atmospheric Environment*, 38(17), 2737–2748. https://doi.org/10.1016/j.atmosenv.2004.01.034.

Salamanca, F.P., Martilli, A., Tewari, M. and Chen, F. (2011) A study of the urban boundary layer using different urban parameterizations and high-resolution urban canopy parameters with WRF. *Journal of Applied Meteorology and Climatology*, 50(5), 1107–1128. https://doi.org/10.1175/2010JAMC2538.1.

Sgoff, C., Acevedo, W., Paschalidi, Z., Ulbrich, S., Bauernschubert, E., Kratzsch, T. and Potthast, R. (2022) Assimilation of crowd-sourced surface observations over Germany in a regional weather prediction system. *Quarterly Journal of the Royal Meteorological Society*, 148, 1–16.

Simmons, A.J., Willett, K.M., Jones, P.D., Thorne, P.W. and Dee, D.P. (2010) Low-frequency variations in surface atmospheric humidity, temperature, and precipitation: inferences from reanalyses and monthly gridded observational data sets. *Journal of*

Geophysical Research: Atmospheres, 115, D01110. https://doi.org/10.1029/2009jd012442.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Wang, W., & Powers, J. G. (2005). A description of the advanced research wrf version 2. Technical Report, DTIC Document (NCAR/TN-468+STR), University Corporation for Atmospheric Research. https://doi.org/10.5065/D6DZ069T

Skok, G. (2015) Analysis of fraction skill score properties for a displaced rainband in a rectangular domain. *Meteorological Applications*, 22(3), 477–484. https://doi.org/10.1002/met.1478.

Skok, G. and Roberts, N. (2016) Analysis of fractions skill score properties for random precipitation fields and ECMWF forecasts. *Quarterly Journal of the Royal Meteorological Society*, 142(700), 2599–2610. https://doi.org/10.1002/qj.2849.

Sun, J., Xue, M., Wilson, J.W., Zawadzki, I., Ballard, S.P., Onvlee-Hooimeyer, J., et al. (2014) Use of NWP for Nowcasting convective precipitation: recent Progress and challenges. *Bulletin of the American Meteorological Society*, 95(3), 409–426. https://doi.org/10.1175/BAMS-D-11-00263.1.

Theeuwes, N.E., Steeneveld, G.J., Ronda, R.J., Heusinkveld, B.G., van Hove, L.W.A. and Holtslag, A.A.M. (2014) Seasonal dependence of the urban heat Island on the street canyon aspect ratio. *Quarterly Journal of the Royal Meteorological Society*, 140(684), 2197–2210. https://doi.org/10.1002/qj.2289.

USGS (2021). Landsat 8 imagery. https://glovis.usgs.gov/app [Accessed 12th February 2021].

Wang, H., Sun, J., Fan, S. and Huang, X.-Y. (2013) Indirect assimilation of radar reflectivity with WRF 3D-Var and its impact on prediction of four summertime convective events. *Journal of Applied Meteorology and Climatology*, 52(4), 889–902. https://doi.org/10.1175/jamc-d-12-0120.1.

Wedi, N.P., Polichtchouk, I., Dueben, P., Anantharaj, V.G., Bauer, P., Boussetta, S., et al. (2020) A baseline for global weather and climate simulations at 1 km resolution. *Journal of Advances in Modeling Earth Systems*, 12(11), e2020MS002192. https://doi.org/10.1029/2020MS002192.

WMO (2018). Guide to Meteorological Instruments and Methods of Observation, Volume 1–Measurement of Meteorological Variables edition WMONo. 8. https://library.wmo.int/doc_num.php?explnum_id=10616 [Accessed 15th August 2022]

Weather Underground (2022). Wundermap. https://www.wunderground.com/wundermap [Accessed 5th September 2022].

Xiao, Q. and Sun, J. (2007) Multiple-radar data assimilation and short-range quantitative precipitation forecasting of a squall line observed during IHOP_2002. *Monthly Weather Review*, 135(10), 3381–3404. https://doi.org/10.1175/MWR3471.1.

Yang, J., Duan, K., Wu, J., Qin, X., Shi, P., Liu, H., Xie, X., Zhang, X. and Sun, J. (2015) Effect of data assimilation using WRF-3DVAR for heavy rain prediction on the northeastern edge of the Tibetan plateau. *Hindawi Publishing Corporation*, 2015, 294589.

Zhang, X., Steeneveld, G.J., Zhou, D., Ronda, R.J., Duan, C., Koopmans, S. and Holtslag, A.A.M. (2020) Modelling urban meteorology with increasing refinements for the complex morphology of a typical Chinese city (Xi'an). *Building and Environment*, 182, 107109. https://doi.org/10.1016/j.buildenv.2020.107109.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Koopmans, S., van Haren, R., Theeuwes, N., Ronda, R., Uijlenhoet, R., Holtslag, A.A.M. *et al*. (2022) The set-up and evaluation of fine-scale data assimilation for the urban climate of Amsterdam. *Quarterly Journal of the Royal Meteorological Society*, 1–21. Available from: https://doi.org/10.1002/qj.4401

APPENDIX A

A1 Derivation of wall and road temperature increments

This appendix presents the derivation of temperature increment of wall and road layers based on differences in modelled–observed urban canopy temperature ΔT_c. Our point of departure is the formulation of the sensible heat flux from walls, H_w, and roads, H_G (km·s^{-1}) in the SLUCM model (Kusaka *et al.*, 2001) and visualized in Figure A1.

$$H_w = C_w (T_w - T_c) \tag{A1}$$

$$H_G = C_G (T_G - T_c) \tag{A2}$$

Herein T_w and T_G are the wall and road temperature (K), T_c is the urban canopy temperature (K), and C_G and C_w are the heat transfer coefficients (m·s^{-1}) that have been parameterized as follows:

$$C_w = C_G \begin{cases} 7.51 U_c^{0.78} & (U_c > 5 \text{ m} \cdot \text{s}^{-1}) \\ 6.15 + 4.18 U_c & (U_c \leq 5 \text{ m} \cdot \text{s}^{-1}) \end{cases} \tag{A3}$$

Herein U_c is the wind speed at the canopy level of T_c. In addition, the sensible heat exchange through the canyon top towards the overlying atmosphere H_a (km·s^{-1}) is parameterized using classical Monin–Obukhov theory:

$$H_a = \frac{\kappa u_*}{\psi_h} (T_c - T_a) \tag{A4}$$

in which κ is the Von Karman constant (dimensionless), u_* the surface friction velocity (m·s^{-1}), ψ_h the stability correction (dimensionless) and T_a the air temperature (K) at the first model layer.

In Figure A1, w represents the road width and h the building height (m). Conservation of these energy fluxes

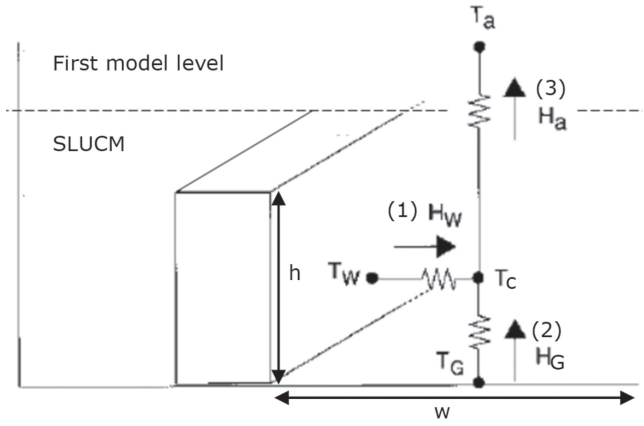

FIGURE A1 Schematic of single-layer urban canopy model (SLUCM) with the relevant energy fluxes adapted for urban nudging. Reused and adapted from Kusaka *et al.* (2001) with permission

(with no storage in the urban canopy) requires that:

$$wH_a = 2hH_w + wH_G \tag{A5}$$

The ratio $2h$ for walls compared to w for roads arises because there are walls on both sides of one road in an urban canopy.

Inserting Equations (A1), (A2), (A4) in Equation (A5) results in:

$$w\frac{\kappa u_*}{\psi_h}(T_c - T_a) = 2hC_w(T_w - T_c) + 2hH_W + wC_G(T_G - T_c) \tag{A6}$$

Bring T_a in front, because later on we can set changes in T_a to 0 between two data assimilation cycles:

$$w\frac{\kappa u_*}{\psi_h}T_a = w\frac{ku_*}{\psi_h}T_c - 2hC_wT_w + 2hC_wT_c - wC_GT_G + wC_GT_c$$

$$T_a = T_c - \frac{\psi_h 2hC_wT_w}{w\kappa u_*} + \frac{\psi_h 2hC_wT_c}{w\kappa u_*} - \frac{\psi_h wC_GT_G}{w\kappa u_*} + \frac{\psi_h wC_GT_c}{w\kappa u_*}$$

$$T_a = T_c\left(1 + \frac{\psi_h 2hC_w}{w\kappa u_*} + \frac{\psi_h wC_G}{w\kappa u_*}\right) - \frac{\psi_h 2hC_wT_w}{w\kappa u_*} - \frac{\psi_h wC_GT_G}{w\kappa u_*}$$

Write the equation as differences Δ denoting the changes in canopy and wall and road temperatures. Assume $\Delta T_a = 0$, since the first model layer is not part of SLUCM and serves as boundary condition:

$$\frac{\psi_h 2hC_w\Delta T_w}{w\kappa u_*} + \frac{\psi_h wC_G\Delta T_G}{w\kappa u_*} = \Delta T_c\left(1 + \frac{\psi_h 2hC_w}{w\kappa u_*} + \frac{\psi_h wC_G}{w\kappa u_*}\right)$$

$$2hC_w\Delta T_w + wC_G\Delta T_G = \frac{\Delta T_c\left(1 + \frac{\psi_h 2hC_w}{wku_*} + \frac{\psi_h wC_G}{wku_*}\right)}{\frac{\psi_h}{wku_*}} = \Delta T_c\left(\frac{w\kappa u_*}{\psi_h} + 2hC_w + wC_G\right)$$

Subsequently, assume $\Delta T_g = \Delta T_w$ since $C_w = C_G$.

$$(2h + w)C_w\Delta T_w = ..$$

$$C_w\Delta T_w = \frac{\Delta T_c\left(\frac{w\kappa u_*}{\psi_h} + 2hC_w + wC_w\right)}{2h + w} = \Delta T_c\frac{w\kappa u_*}{\psi_h(2h + w)} + \frac{\Delta T_cC_w(2h + w)}{(2h + w)}$$

$$C_w\Delta T_w = C_w\Delta T_c\left(1 + \frac{w\kappa u_*}{\psi_hC_w(2h + w)}\right)$$

The final result (A7) matches Equation (5) of the main paper with the full notation $C_{ah} = \frac{\kappa u_*}{\psi_h}$

$$\Delta T_{w,G} = \Delta T_c\left(1 + \frac{w\kappa u_*}{\psi_hC_w(2h + w)}\right) = \Delta T_c\left(1 + \frac{w}{(2h + w)}\frac{\kappa u_*}{\psi_hC_w}\right) = \Delta T_c\left(1 + \frac{w}{(2h + w)}\frac{C_{ah}}{C_w}\right) \tag{A7}$$