Article:
Salman, N, Khan, A orcid.org/0000-0002-7521-5458, Kemp, AH et al. (1 more author) (2022) Indoor Temperature Forecast based on the Lattice Boltzmann method and Data Assimilation. Building and Environment, 210. 108654. p. 108654. ISSN 0007-3628

https://doi.org/10.1016/j.buildenv.2021.108654

Indoor Temperature Forecast based on the Lattice Boltzmann method and Data Assimilation

N. Salman[a], A. Khan[a], A. H. Kemp[b], C. J. Noakes[a]

[a]*School of Civil Engineering, University of Leeds, Woodhouse Ln., Leeds, LS2 9DY, United Kingdom*
[b]*School of Electronic and Electrical Engineering, University of Leeds, Woodhouse Ln., Leeds, LS2 9DY, United Kingdom*

Abstract

Control of the indoor environment in buildings is reactive and usually based on sensor data at a single location in a room without any consideration of the air flows within the space. The ability to accurately simulate indoor conditions would enable predictive control offering improved environmental conditions and better energy efficiency. This study investigates the temporal prediction of temperature and its coupling with sensor data for improving thermal comfort forecasting in an indoor environment. We present a real-time implementation of a computational fluid dynamics (CFD) Lattice Boltzmann method (LBM) model coupled with data assimilation (DA) to make periodic updates to the state of the model. *Variational* and *sequential* DA techniques are evaluated and two novel methods for real-time accurate temperature prediction are presented. The models are demonstrated for prediction of temperature in an idealised room, with simulated temperature sensor readings used to update LBM-based flow predictions. The LBM-DA approach overcomes the need for accurate boundary conditions while considerably improving the transient and spatial prediction of temperature in terms of root mean squares error, thereby avoiding large deviation from the room's true flow state. The accuracy of prediction is shown to depend on number of sensors with poor prediction using just a single sensor. Results also depend on quality of sensor data, with highly variable data yielding poorer results. The approach has potential for application in indoor environments to provide more accurate and faster response of control systems to changing environmental conditions.

Keywords: Thermal comfort, Lattice Boltzmann method, Data Assimilation, Ensemble Kalman filter, temperature sensors.

1. Introduction

Monitoring and predicting indoor air quality (IAQ) and thermal comfort is essential given that we spend about 90% of the time living indoors [21]. Poor indoor environments are widely recognised to have a detrimental impact on human health, comfort, cognitive performance, and creativity. Control of the indoor environment also has a

Email addresses: n.salman1@leeds.ac.uk (N. Salman), a.khan@leeds.ac.uk (A. Khan), a.h.kemp@leeds.ac.uk (A. H. Kemp), c.j.noakes@leeds.ac.uk (C. J. Noakes)

substantial impact on energy performance, with buildings representing around 40% of EU energy use [41] and 36% of carbon emissions [9]. Strategies to control temperature, humidity and IAQ to manage pollutant exposure, thermal comfort and balance the quality of the environment against energy use range from simple thermostatic control of radiators for heating, to more complex demand controlled ventilation that responds to sensing of temperature and IAQ parameters such as CO_2 [15]. However, these approaches are reactive rather than predictive and are generally based on a simple single point measure assuming a well-mixed indoor airflow rather than considering the detail of the indoor flow [28]. The ability to make real-time or faster than real-time prediction of indoor conditions that considers that complexity of airflow, temperature and contaminant distributions would enable more effective and localised control of the environmental conditions, improving both individual experience and the performance of the building. This is particularly important for critical environments such as healthcare, where comfort needs vary between patients and IAQ includes control of infection which can be transmitted through the air [27].

Computation simulation of indoor environments

In recent years, computational fluid dynamics (CFD) based simulations have been widely used by architects and engineers to assess various aspects of building design [38]. For example, outdoor airflow around the buildings, analysis of thermal environments [25], ventilation design [2], and IAQ evaluation [39]. Conventional CFD methods, although successfully applied to simple steady-state conditions have limited application in complex indoor environments, due to their transient nature resulting from variable temperature sources or human mobility. Transient CFD models suffer from excess computational complexity resulting in unacceptable processing time (several hours to months) [18], [40]. To address the excessive time issue, high-end supercomputers are used [5], which is not a financially viable solution in many cases. Other attempts to speed up the computational process include multi-zone network models [35], zonal models [26] and fast fluid dynamics (FFD) methods. The multi-zone network and zonal models although faster than conventional CFD, suffer from simplistic assumptions and loss of accuracy. FFD on the other hand provides better accuracy and much higher speeds (50 times faster than CFD) but still suffer from loss of spatial and temporal accuracy.

Non-traditional CFD methodologies, such as Lattice Boltzmann Method (LBM), have recently been shown to give comparable results to unsteady CFD methods like large eddy simulation (LES) without the substantial computational overhead or excessive solution times [19], [8]. LBM is a microscopically inspired method designed to solve macroscopic fluid dynamics problems. It originates from the lattice gas automata (LGA) method and can be regarded as an explicit discretisation of the Boltzmann equa-

tion. The LBM has several advantages over traditional Navier-Stokes equations (NSE), such as its numerical stability and accuracy, the capacity to efficiently handle complex geometries, and the data-parallel nature of its algorithm. Thus, LBM is an explicit numerical scheme with only local operations. The LBM describes the fluid by particle distribution functions (PDFs) located at grid points (lattice), while the macroscopic properties of the fluid are recovered from moments of these PDFs. It has the advantage of being easy to implement and is especially well suited for massively parallel architectures like graphics processing units (GPU). The approach is gaining popularity and usefulness in modelling of the built environment including exploration of indoor [19] and outdoor [20] flows.

Application of data assimilation approaches

Regardless of the method used, CFD predictions are likely to diverge from reality over extended periods of time due to unknown/unresolved boundary conditions as well as changes in the dynamic environment. To circumvent this, data assimilation (DA) [1] can be used to *correct* the prediction of flow models using data from real-world measurements. DA techniques have been widely used in weather prediction (e.g. ECMWF, Met Office, Meteo-France, etc.) and oceanography, however, their utility has not yet been exploited for indoor environments. DA is a tool for incorporating observation into the state model to update (correct) the state of the model. DA techniques can generally be classed into statistical DA [11] and variational DA [30], both have their benefits and disadvantages. Statistical DA methods incorporate observation to the model sequentially in time and include approaches like the Kalman filter (KF) [36], extended Kalman filter (EKF), and Ensemble Kalman filters (EnKF) [7]. Variational DA on the other hand includes the 3DVAR and 4D-VAR methods [24]. 3DVAR is a classical DA implementation and is based on minimising a cost function that contains information on the background errors with their respective spatial correlation (and also cross-correlation between different physical parameters) and the observations with their errors. DA techniques can be modified to cater for dynamic indoor environment needs, this can be, for example, achieved by accurately modeling the interaction within the physical environment (the background) that encompasses both spatial and inter-variable correlations. Simplistic approximations to the correlations can be made such as assuming the correlations to be Gaussian with an exponentially decaying profile, and thus simplifying the correlation structure by separating the correlation matrices along each physical dimension [23], [22]. In present work, the background covariance matrix is modeled following a similar approach, where it is assumed that the simulated sensor measurement of temperature in an indoor environment is correlated to the surrounding cells via a Gaussian function. The correlations indeed die out after a certain distance, which can in practice be determined with an exhaustive offline measurement

campaign.

Study objectives

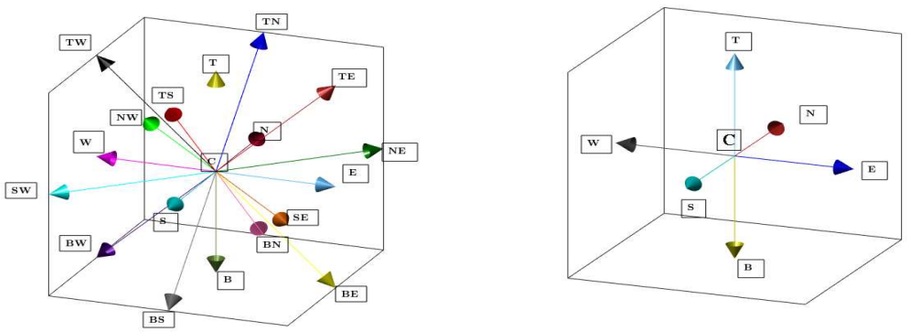
The aim of this study is to develop an approach to real-time visualization and prediction of spatially and transiently varying temperature within the built environment through a data assimilation approach to enable a coupled sensor-airflow predictive modelling approach for thermal comfort forecast. For accurate temperature prediction in real-time, two novel implementations are presented that consider both variational and statistical DA methods, i) LBM-3DVAR, ii) LBM-EnKF. In both approaches, the airflow in an indoor environment is simulated using the LBM which is coupled with energy equations to model the temperature distribution in space and time within the space. This is used to produce a simulated ground truth model to compare to the DA methods. The prediction from the LBM is then corrected with simulated sensor data using data assimilation cycles to account for changing temperature boundary conditions. Simulation and analysis for both methods are presented for an idealised office environment setup with various boundary conditions. The root mean squares error (RMSE) is obtained for both methods and compared to the case where the boundary conditions change but no DA is performed. It is demonstrated that the LBM coupled with DA performs significantly better than without DA.

The rest of the paper is organized as follows, Section 2 presents the LBM approach while Section 3 explains the 3DVAR and EnKF techniques. Section 4 presents the simulation results which are followed by a discussion section. Finally the conclusions are presented.

2. Lattice Boltzmann Method

The LBM discretises the Boltzmann equations in phase space, involving space, time, and velocities. The flow domain, in this case the indoor environment, is represented fully by a regular grid or lattice of points. The fluid is represented by the PDFs which resides at these lattice points and indicates a probability of the presence of particles with a given velocity. Macroscopic properties of the fluid, e.g., density and velocity are recovered from these PDFs. However, the PDFs can only propagate in set directions depending on the underlying model. A convention for these models is the DdQq name scheme, where d denotes the number of dimensions and q represents the discrete directions. In this study, D3Q19 is used for the velocity of the flow while a smaller D3Q7 is used for heat flow. The rationale behind the choice of D3Q19 lattice for velocity field is a compromise between computational cost and the minimum number velocity directions required to faithfully reproduce the hydrodynamics in the continuum limit. Similarly, the D3Q7 has been shown to be the minimum number which reproduces the advection diffusion equations adequately in 3D. More details in

[19, 10]. Two steps are involved in the LBM simulation, i) a streaming operation in which the PDFs are streamed or propagated to the neighboring grid points and depend on the lattice geometry and ii) a collision step, in which the PDFs are relaxed towards an equilibrium depending on the chosen relaxation scheme.



(a) D3Q19 (b) D3Q7

Figure 1: The discrete velocities directions for D3Q19 and D3Q7. C (center), E (east), N (north), W (west), S (south), NE (north east), NW (north west), SW (south west) and SE (south east). T (top), B (bottom),TE (top east), TN (top north), TW (top west), TS (top south), BE (bottom east), BN (bottom north), BW (bottom west) and BS (bottom south).

2.1. The D3Q19 model for velocities

The D3Q19 model, shown in Fig. 1a, maintains sufficient isotropy while having the minimum number of velocity directions. The collision operation is based on the Bhatnagar-Gross-Krook (BGK) scheme [6]. The BGK approach states that the PDFs $\mathbf{f} = \{f_i\}_{i\in\{0,\dots,18\}}$ relaxe to the equilibrium PDFs $\mathbf{f}^{eq} = \{f_i^{eq}\}_{i\in\{0,\dots,18\}}$ over a relaxation time τ_f. The relationship between the PDFs using the BGK collision approximation is given by the lattice Boltzmann equation

$$f_i\left(\mathbf{x} + c\mathbf{e}_i\Delta t, t + \Delta t\right) = f_i\left(\mathbf{x}, t\right) - \frac{1}{\tau_f}\left(f_i\left(\mathbf{x}, t\right) - f_i^{eq}\left(\mathbf{x}, t\right)\right), \tag{1}$$

where $c = \frac{\Delta x}{\Delta t}$ is the lattice speed while Δx is the space between lattice points and Δt is the time step. The macro fluid density ρ and velocity $\boldsymbol{v}$ are recovered from the PDFs by

$$\rho = \sum_{i=0}^{18} f_i, \qquad \boldsymbol{v} = \frac{1}{\rho}\sum_{i=0}^{18} c\mathbf{e}_i f_i,$$

where $\mathbf{e}_i$ represent the velocity directions, i.e.,

$$\mathbf{e}_i = \begin{cases} (0,0,0) & i = 0 \quad (C)\,, \\ (\pm 1,0,0) & i = 1,2 \quad (E,W)\,, \\ (0,\pm 1,0) & i = 3,4 \quad (N,S)\,, \\ (0,0,\pm 1) & i = 5,6 \quad (T,B)\,, \\ (\pm 1,\pm 1,0) & i = 7,8,9,10 \quad (NE,NW,SE,SW)\,, \\ (\pm 1,0,\pm 1) & i = 11,12,13,14 \quad (TE,TW,BE,BW)\,, \\ (0,\pm 1,\pm 1) & i = 15,16,17,18 \quad (TN,TS,BN,BS)\,. \end{cases}$$

With these new computed values, the local equilibrium is obtained by

$$f_i^{eq} = \rho w_i \left(1 + 3\frac{\mathbf{e}_i \boldsymbol{v}}{c} + 4.5\frac{(\mathbf{e}_i \boldsymbol{v})^2}{c^2} - 1.5\frac{\boldsymbol{v}^2}{c^2}\right), \tag{2}$$

where w_i represent the weight coefficients of individual velocity direction $\mathbf{e}_i$ i.e., $w_0 = \frac{1}{3}, w_{1,\ldots,6} = \frac{1}{18}, w_{7,\ldots,18} = \frac{1}{36}$.

The NSE can be recovered from BGK model through the Chapman-Enskog expansion [16] with the kinematic viscosity of the fluid ν relating to τ_f by

$$\nu = \frac{(2\tau_f - 1)\,\Delta x^2}{6\Delta t}. \tag{3}$$

2.2. The D3Q7 model for temperature

In this study, the temperature is chosen as a metric for IAQ evaluation. Other IAQ parameters such as CO_2, humidity, etc. can be modeled similarly. To include the flow of heat, a coupled temperature model in which the temperature is treated as a scalar quantity advected by the fluid and is solved on a smaller D3Q7 lattice, as shown in Fig. 1b is used [10]. This smaller lattice is sufficient to represent heat while maintaining low computational complexity. The two-step streaming and collision expression for temperature is given by

$$T_i\left(\mathbf{x} + c\mathbf{e}_i\Delta t, t + \Delta t\right) = T_i\left(\mathbf{x}, t\right) - \frac{1}{\tau_T}\left(T_i\left(\mathbf{x}, t\right) - T_i^{eq}\left(\mathbf{x}, t\right)\right), \tag{4}$$

where T_i is the temperature PDF along the direction $\mathbf{e}_i$, where now $i \in \{1, ..., 7\}$ and τ_T is the relaxation time of temperature flow towards equilibrium T_i^{eq}. τ_T when related to the thermal diffusivity D is given by $\tau_T = \frac{3D\Delta t}{\Delta x^2} + 0.5$ and the equilibrium PDF is given by

$$T_i^{eq} = \frac{T}{7}\left(1 + 2\frac{\mathbf{e}_i \boldsymbol{v}}{c}\right). \tag{5}$$

The macro temperature is computed from the micros by the summation $T = \sum_{i=0}^{6} T_i$ [14]. In this work, it is assumed that the temperature has little effect on the flow except for the buoyancy term [14]. The buoyancy effects of temperature are addressed

by adding a forcing term to the NSE, also known as the Boussinesq approximation [34], however other models exist [4], [29]. For the coupling of the temperature with the flow, the Boussinesq forcing term $\mathbf{F}_B = -\mathbf{g}\beta\left(T - T_0\right)$ is added to the LBM equation, where $\mathbf{g}$ is the acceleration due to gravity and β is the coefficient of thermal expansion. In this study, however, a simplified approximation of the forcing term that is aligned with the direction of gravity is used, i.e.,

$$F_i = \frac{\mathbf{e}_i.\mathbf{F}_B}{2} = \frac{\pm\mathbf{g}\beta\left(T - T_0\right)}{2}. \tag{6}$$

Finally the coupled LBM equation is given by

$$f_i\left(\mathbf{x} + c\mathbf{e}_i\Delta t, t + \Delta t\right) = f_i\left(\mathbf{x}, t\right) - \frac{1}{\tau_f}\left(f_i\left(\mathbf{x}, t\right) - f_i^{eq}\left(\mathbf{x}, t\right)\right) + F_i\Delta t. \tag{7}$$

3. Data Assimilation Approaches

Data assimilation (DA) is the approximation of the true state of some physical system at a given time by combining time-distributed observations with a dynamic model in an optimal way [1]. Forecast models like the one used in numerical weather prediction (NWP), or the LBM approach in this study for temperature prediction start with erroneous initial conditions, or more precisely with initial conditions that do not represent the true state of the physical space. Moreover, these errors accumulate over time rendering the forecast meaningless in many cases. The predictive models can be steered back to the true state by the periodic injection of physical observations. Observations, such as sensor readings for IAQ measurements are themselves not error-free. DA is the art of dealing with these uncertainties in models and observations. DA forms the link between a free-flowing model and spatiotemporally sparse observations. DA has been applied in many different fields and has been referred to by as many names [33], [37]. However, the underlying mechanism is the same i.e., most modern DA methods are based on Bayes theorem [32]. Broadly, DA techniques are classed into two categories i) *Varitaional DA* ii) *Statistical DA* . Variational methods including 3DVAR or 4DVAR obtain the optimized compromise between the background (that is the flow model) and observation by minimizing a cost function. These assume linearity in the model as in the case of 3DVAR or goes through a linearization process as in 4DVAR.

Algorithm 1: LBM-DA algorithm

Required: LBM parameters initialization listed in table 2 and time step (ts)=0

while $ts \leq$ *Total time* **do**

1: Operate LBM equation for flow PDF
$f_i\left(\mathbf{x}+c\boldsymbol{e}_i\Delta t, t+\Delta t\right) = f_i\left(\mathbf{x},t\right) - \frac{1}{\tau_f}\left(f_i\left(\mathbf{x},t\right) - f_i^{eq}\left(\mathbf{x},t\right)\right)$

2a: Operate the LBM streaming step for temperature
$T_i\left(\mathbf{x}+c\mathbf{e}_i\Delta t, t+\Delta t\right) = T_i\left(\mathbf{x},t\right)$

2b: Go to step 3 if ts=DA cycle otherwise go to step 4.

3: Go to 3DVAR or EnKF to perform DA and update macro temperature
$T = T^a$

4: Operate the LBM collision step for temperature with updated marco T
$T_i^{eq} = \frac{T}{6}\left(1+2\frac{\mathbf{e}_i\boldsymbol{v}}{c}\right)$

5: Operate the coupled LBM equations
$f_i\left(\mathbf{x}+c\mathbf{e}_i\Delta t, t+\Delta t\right) = f_i\left(\mathbf{x},t\right) - \frac{1}{\tau_f}\left(f_i\left(\mathbf{x},t\right) - f_i^{eq}\left(\mathbf{x},t\right)\right) + F_i\Delta t$
$ts = ts+1$

end

Statistical DA includes the ensemble Kalman filter (EnKF) and its derivatives. These obtain the background statistics from an ensemble (multiple runs of the flow forecast) and is not limited to linearized models. The variational DA assumes a pre-calculated background covariance matrix while the sequential DA uses a Monte Carlo approach to estimate the covariance with a multiple (ensemble) simulation runs. A third sub-category can be the *Monte-Carlo methods (particle filters)* which deal with non-Gaussian errors in the flow and observation models. In this study, linearity in the observation model and Gaussian errors are assumed, therefore the 3DVAR and the EnKF techniques are studied. The LBM-DA algorithm is presented in Algorithm 1.

3.1. 3DVAR

3DVAR is a classical DA implementation and is based on minimizing a cost function that contains information on the background errors with their respective spatial correlation (and also cross-correlation between different physical parameters) and the observations with their errors. The cost function of 3DVAR in its incremental form [3] is given as follows:

$$J\left(\delta\boldsymbol{x}\right) = \frac{1}{2}\left(\delta\boldsymbol{x}-\delta\boldsymbol{x}^b\right)^T\mathbf{B}^{-1}\left(\delta\boldsymbol{x}-\delta\boldsymbol{x}^b\right) + \frac{1}{2}\left(\boldsymbol{y}-\mathbf{H}\left(\boldsymbol{x}^g+\delta\boldsymbol{x}\right)\right)^T\mathbf{R}^{-1}\left(\boldsymbol{y}-\mathbf{H}\left(\boldsymbol{x}^g+\delta\boldsymbol{x}\right)\right), \tag{8}$$

where $\delta\boldsymbol{x}$ is the N by 1 vector given by $\delta\boldsymbol{x} = \left(\boldsymbol{x}-\boldsymbol{x}^g\right)$, with N being the number of states, $\delta\boldsymbol{x}^b$ is the background increment given by $\delta\boldsymbol{x}^b = \boldsymbol{x}^b - \boldsymbol{x}^g$ with $\boldsymbol{x}^g$ being a known reference state. $\boldsymbol{x}$ is the vector of the model state and $\boldsymbol{x}^b$ is the background state, which is the state of the forecast model just before the DA cycle. $\boldsymbol{y}$ is the O by 1 measurement vector with the number of measurements $O << N$. The operator $\mathbf{H}$ maps the variables from the state space into the observation space. $\mathbf{R}$ is the covariance matrix of the observation and $\mathbf{B}$ is the background error covariance. $\mathbf{R}$ and its inversion is easy to deal with as its size only depends on the number of observations. If the observations

are obtained from sparsely located sensors, they can be assumed to be uncorrelated, hence in most cases, $\mathbf{R}$ is a diagonal matrix. The matrix $\mathbf{B}$ encapsulates the correlation between grid points and hence is responsible for the spread of measurement (sensor) information to the surrounding cells and is key to the success of modern variational DA techniques [3]. In general, explicit determination of the $\mathbf{B}$ is not feasible, this is partly because of its large size (and rank) $> 10^7$ and also the fact that the true state is unknown. Thus $\mathbf{B}$ is commonly modeled and is assumed to be static. The literature on modeling $\mathbf{B}$, almost entirely has been focused on using the dynamical properties of the atmosphere and assumes homogeneity and horizontal isotropy in the background errors. The success of NWP in terms of accurate analysis and forecast relies intimately on the accurate modeling of $\mathbf{B}$. In this study, an approximation of the correlations is made by assuming the correlations to be Gaussian with an exponentially decaying profile. This assumption was made in [22], [23] for oceanography and outdoor pollution dispersion. It is assumed in the current work, that temperature (or any other air quality parameter) will follow a similar exponential decay profile with distance. Thus matrix $\mathbf{B}$ is factorized i.e.,

$$\mathbf{B} = \mathbf{SCS}^T, \tag{9}$$

where $\mathbf{S}$ is a diagonal matrix of the standard deviations of individual cell points while $\mathbf{C}$ is the correlation matrix. Also

$$\mathbf{C} = \mathbf{C}_L * (\mathbf{C}_L)^T, \tag{10}$$

where $\mathbf{C}_L$ is the lower triangle of the Cholesky decomposition of the matrix $\mathbf{C}$. With this modification, the variable transformed representation of the cost function is given as

$J(\delta \boldsymbol{z}) =$

$$\frac{1}{2}\left(\delta \boldsymbol{z} - \delta \boldsymbol{z}^b\right)^T \left(\delta \boldsymbol{z} - \delta \boldsymbol{z}^b\right) + \frac{1}{2}\left(\boldsymbol{y} - \mathbf{H}\left(\boldsymbol{x}^g + \mathbf{SC}_L \delta \boldsymbol{z}\right)\right)^T \mathbf{R}^{-1}\left(\boldsymbol{y} - \mathbf{H}\left(\boldsymbol{x}^g + \mathbf{SC}_L \delta \boldsymbol{z}\right)\right). \tag{11}$$

Where $\delta \boldsymbol{z}$ is the transformed variable given by $\delta \boldsymbol{z} = \mathbf{S}^{-1}\mathbf{C}_L^{-1}\delta \boldsymbol{x}$. If the elements of $\mathbf{C}$ are a product of a Gaussian function i.e., $corr(i,j) = \exp\left(-\frac{(i-j)^2}{L}\right)$, with L being the decorrelation distance, then $\mathbf{C}_L$ can be further decomposed as

$$\mathbf{C}_L = \mathbf{C}_{L_x} \otimes \mathbf{C}_{L_y} \otimes \mathbf{C}_{L_z}. \tag{12}$$

Where $\mathbf{C}_{L_x}$,$\mathbf{C}_{L_y}$,$\mathbf{C}_{L_z}$ are the correlation matrices along the x, y, and z directions, and $\otimes$ is the Kronecker product. This simplification greatly reduces the computation complexity of the 3DVAR cost function as the correlation matrices along the individual axis have significantly reduced dimensions. These correlation matrices can be determined

and stored before the DA cycles. Furthermore, this also enables the DA to incorporate anisotropic correlations. For any indoor environment, the decorrelation distance can be determined with an offline measurement campaign. The 3DVAR methodology is shown in Algorithm 2.

Algorithm 2: 3DVAR algorithm

Required: Macro $T = \boldsymbol{x}^b$, $\mathbf{B}$ and 3DVAR parameters initialization given in table 3.

$k = 0$,

while $|J_k - J_{k-1}| > \epsilon$ **do**

1: Compute $J(\delta \boldsymbol{z})$ from Eq. (11)

2: Compute $\nabla \delta z$

$$\nabla \delta z = \left(\delta \boldsymbol{z} - \delta \boldsymbol{z}^b\right) - \left(\mathbf{C}_L \mathbf{S}\right)^T \mathbf{H}^T \mathbf{R}^{-1} \left(\boldsymbol{y} - \mathbf{H}\left(\boldsymbol{x}^g + \mathbf{S}\mathbf{C}_L \delta \boldsymbol{z}\right)\right)$$

3: Update δz_{k+1} with gradient descent $k = k + 1$

end

4: $\boldsymbol{x}^a = \boldsymbol{x}^g + \mathbf{S}\mathbf{C}_L \delta \boldsymbol{z} = T^a$

3.2. Ensemble Kalman filter

The EnKF mimics the Kalman filter by approximating the calculation of complex covariance matrices with a limited number of ensemble members. The analysis step is given by

$$\boldsymbol{x}_i^a = \boldsymbol{x}_i^b + \mathbf{K}\left(\boldsymbol{y} - \mathbf{H}\left(\boldsymbol{x}_i^b\right)\right), \tag{13}$$

where i=1,...,M is the ensemble index, $\boldsymbol{x}_i^b$ is the forecast state (or the background) i.e., the state just before the assimilation cycle. The superscript b is used here instead of the more conventional f to be consistent with the 3DVAR formulation in the previous section. $\mathbf{K}$ is the so-called Kalman gain given by

$$\mathbf{K} = \mathbf{P}^b \mathbf{H}^T \left(\mathbf{H}\mathbf{P}^b \mathbf{H}^T + \mathbf{R}\right)^{-1}, \tag{14}$$

where $\mathbf{P}^b$ is the background error covariance matrix and is approximated by

$$\mathbf{P}^b = \frac{1}{M-1} \sum_{i=1}^{M} \left(\boldsymbol{x}_i^b - \bar{\boldsymbol{x}}^b\right)\left(\boldsymbol{x}_i^b - \bar{\boldsymbol{x}}^b\right)^T, \quad \text{where } \bar{\boldsymbol{x}}^b = \frac{1}{M} \sum_{i=1}^{M} \boldsymbol{x}_i^b.$$

The analysis covariance matrix $\mathbf{P}^a$ of the analysis ensemble $\boldsymbol{x}^a$ is given by

$$\mathbf{P}^a = \left(\mathbf{I} - \mathbf{K}\mathbf{H}\right) \mathbf{P}^b \left(\mathbf{I} - \mathbf{K}\mathbf{H}\right)^T. \tag{15}$$

However in order achieve the best linear unbiased estimator (BLUE) analysis of the Kalman filter, one should obtain [1] $\mathbf{P}^a$ as follows

$$\mathbf{P}^a = \left(\mathbf{I} - \mathbf{K}\mathbf{H}\right) \mathbf{P}^b. \tag{16}$$

One way to obtain (16) is by perturbing the observation with errors drawn from a Gaussian distribution, i.e., $\hat{\boldsymbol{y}} = \boldsymbol{y} + \boldsymbol{n}$ where $\boldsymbol{n} \sim \mathcal{N}(0, \mathbf{R})$. However, this step adds an additional source of sampling error into the observation. The implementation of EnKF requires some additional treatment which is discussed in the next subsection.

Implementation of the EnKF

The EnKF attempts to estimate a large background covariance matrix with only a limited number of ensemble members, however, in practice this is accompanied by large sampling error rendering the covariance matrix rank deficient. This results in spurious correlations at longer distances that have no relevance to the ground truth. This almost always results in the divergence of the filter. Fortunately, some tricks can be incorporated into the algorithm to avoid the divergence of the filter. Here *localisation* and *inflation* will be discussed as measures to avoid filter divergence.

Localisation. Localisation is based on the notion that cell points that are close to each other have a higher correlation while distant cell points have lower correlations. In this work, the low-rank background covariance matrix $\mathbf{P}^b$ is regularised by removing the long-range (and meaningless) correlations by a point-wise multiplication of a correlation matrix $\mathbf{L} \in \mathcal{R}^{N \times N}$, i.e.,

$$\left[\mathbf{L} \circ \mathbf{P}^b\right]_{i,j} = \left[\mathbf{P}^b\right]_{i,j} \left[\mathbf{L}\right]_{i,j}$$

where $\circ$ is the Schur product. [17] shows that the $\mathbf{L} \circ \mathbf{P}^b$ is positive definite if $\mathbf{P}^b$ and $\mathbf{L}$ are positive definite. Thus, this requires the building of a positive definitive covariance matrix $\mathbf{L}$ from a correlation function. A candidate for such functions is the Gaspari-Chon function [13], which cuts off the correlations at longer distances. However, to stay consistent with the 3DVAR formulation, the correlation function as used in the 3DVAR algorithm i.e., $\mathbf{L}_{i,j} = \exp\left(-\frac{(i-j)^2}{L}\right)$ will be used for localisation. In order to obtain the full rank matrix $\mathbf{L} \circ \mathbf{P}^b$, one performs this regularisation in the Kalman gain step (14), it is noted here that the computation of $\left(\mathbf{L} \circ \mathbf{P}^b\right) \mathbf{H}^T$ and $\mathbf{H}\left(\mathbf{L} \circ \mathbf{P}^{\boldsymbol{b}}\right)\mathbf{H}^T + \mathbf{R}$ becomes unfeasible for moderate to large cell numbers. To avoid the multiplication of large matrices, the computational complexity is reduced by using localistion in the observation space which generally has a much smaller dimension than the state space. Thus

$$\left(\mathbf{L} \circ \mathbf{P}^b\right) \mathbf{H}^T \rightarrow \mathbf{L}_c \circ \left(\mathbf{P}^b \mathbf{H}^T\right)$$

and

$$\mathbf{H}\left(\mathbf{L} \circ \mathbf{P}^{\boldsymbol{b}}\right)\mathbf{H}^T + \mathbf{R} \longrightarrow \mathbf{L}_s \circ \left(\mathbf{H}\mathbf{P}^{\boldsymbol{b}}\mathbf{H}^T\right) + \mathbf{R},$$

where $\mathbf{L}_c \in \mathcal{R}^{N \times O}$ and represents the correlations of the sensor readings with the surrounding cells, while $\mathbf{L}_s \in \mathcal{R}^{O \times O}$ represents the correlations between the simulated sensor readings.

The modelling of $\mathbf{L}_c$ requires some care, in particular, if the mesh grid of the area is given by $\mathbf{G} = \left[\boldsymbol{g}_x, \boldsymbol{g}_y, \boldsymbol{g}_z\right] \in \mathcal{R}^{N\times 3}$ and vector of the index of the cells where sensors are located is given by $\boldsymbol{j}$, then $\mathbf{L}_c$ can be determined as

$$\left[\mathbf{L}_c\right]_{i,j} = \exp\left\{-\left[\frac{\left(\boldsymbol{g}_x\left(i\right) - \boldsymbol{j}\left(j\right)\right)^2}{L_x} + \frac{\left(\boldsymbol{g}_y\left(i\right) - \boldsymbol{j}\left(j\right)\right)^2}{L_y} + \frac{\left(\boldsymbol{g}_z\left(i\right) - \boldsymbol{j}\left(j\right)\right)^2}{L_z}\right]\right\} \tag{17}$$

for $i = 1, ..., N$ and $j = 1, ..., O$. One can then easily determine the correlations along the x, y and z axis with the decorrelation distances L_x, L_y and L_z respectively.

$\mathbf{L}_s$ on the other hand is given by $\mathbf{L}_s = \exp\left\{-\left[\frac{(i-j)^2}{L_s}\right]\right\}$, however in this study, the simulated sensor measurements are assumed to be uncorrelated with each other thus $\mathbf{L}_s = \mathbf{I}_O$.

Inflation. For a limited number of ensemble members the sampling error will accumulate over time resulting in the sampled covariance matrix $\mathbf{P}^b$ to be a poor estimate of the true covariance, this inevitably will lead to filter divergence. One way counter this to inflate the covariance matrix by a factor λ i.e., $\mathbf{P}^b = \lambda\mathbf{P}^b$ where the value of λ is greater than 1. The EnKF steps are presented in Algorithm 3.

Algorithm 3: EnKF algorithm

Required: Ensemble $\boldsymbol{x}^b_{i,ts} = T_{i,ts}$, and initiate EnKF parameters listed in table 5

1: Perturb observation $\hat{\boldsymbol{y}}_{i,ts} = \boldsymbol{y}_{ts} + \boldsymbol{n}_i, \quad \boldsymbol{n}_i \sim \mathcal{N}\left(0, \mathbf{R}\right), \quad i = 1, ..., M$

2: Calculate ensemble means $\bar{\boldsymbol{x}}^b_{ts} = \frac{1}{M}\sum_{i=1}^{M}\boldsymbol{x}^b_{i,ts}$

3: Compute the normalised anamolies $\left[\mathbf{X}^b\right]_{i,ts} = \frac{\boldsymbol{x}^b_{i,ts} - \bar{\boldsymbol{x}}^b_{ts}}{\sqrt{M-1}}$

4: Compute the background covariance $\mathbf{P}^b_{ts} = \mathbf{X}^b_{i,ts}\left(\mathbf{X}^b_{i,ts}\right)^T$

5: Perform inflation $\mathbf{P}^b_{ts} = \lambda\mathbf{P}^b_{ts}$

6: Compute the Kalman gain with localisation

$\mathbf{K}_{ts} = \mathbf{L}_c \circ \left(\mathbf{P}^b_{ts}\mathbf{H}^T\right)\left(\mathbf{L}_s \circ \left(\mathbf{H}\mathbf{P}^b_{ts}\mathbf{H}^T\right) + \mathbf{R}\right)^{-1}$

7: Updating ensemble $\boldsymbol{x}^a_{i,ts} = \boldsymbol{x}^b_{i,ts} + \mathbf{K}_{ts}\left(\hat{\boldsymbol{y}}_{i,ts} - \mathbf{H}\left(\boldsymbol{x}^b_{i,ts}\right)\right)$

8: $T^a_{ts} = \boldsymbol{x}^a_{i,ts}$,

Go to be step 4 in algorithm 1.

4. Model Application and Simulation Results

In this section, the performance of LBM-DA and its comparison with LBM without DA is presented. All simulations are carried out using our in-house code written in MATLAB R2017a. The choice of MATLAB is purely to demonstrate the viability of our proposed methodology. An office environment with $n_x = 40, n_y = 20, n_z = 20$ is represented by a room of non-dimensional lattice units for the purpose of testing. The geometry of the office environment is shown in Fig. 2.

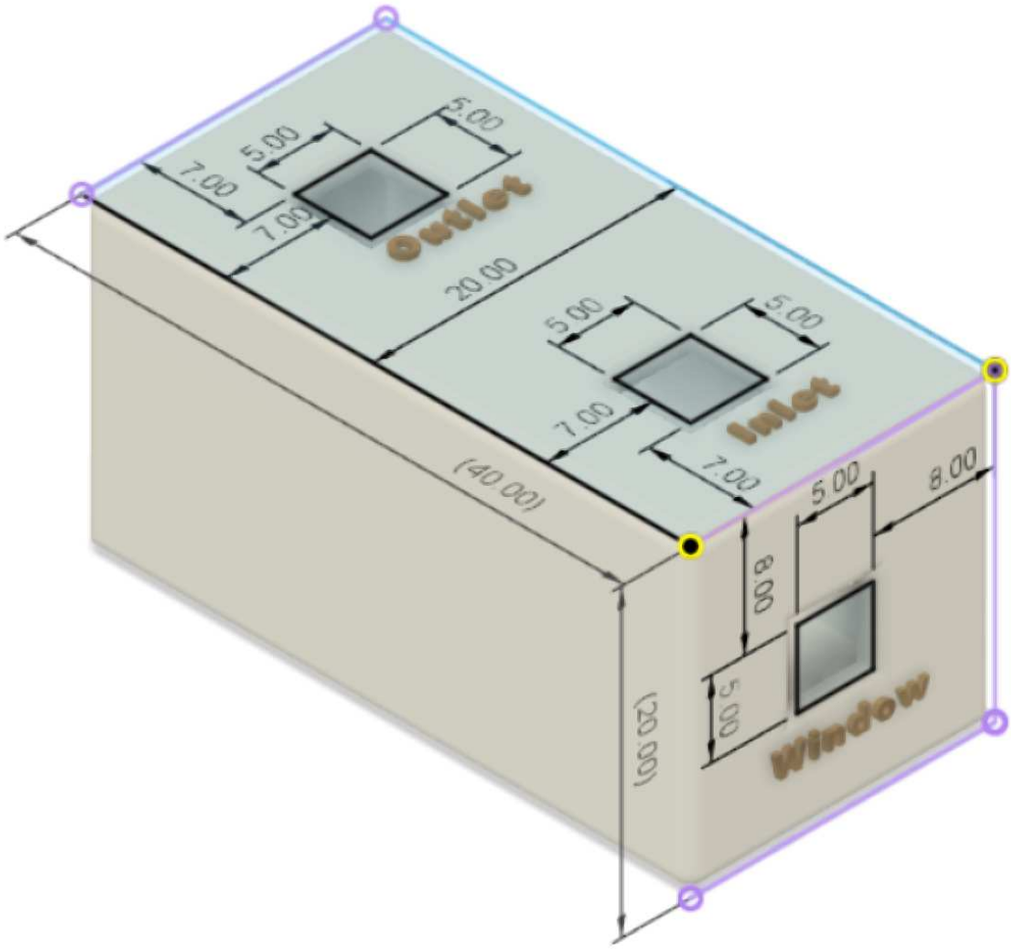


Figure 2: Simulated office geometry, with all dimensions in non-dimensional lattice units

The walls of the room are insulated while the floor is heated to a particular temperature to emulate human activity and office electronics. An inlet and an outlet with known fixed temperature and airflow velocity is simulated in the ceiling. A window is also simulated on one of the walls, its temperature changes emulating variable outdoor temperature. The inlet, outlet and window velocities are 0.1 $\frac{\Delta x}{\Delta t}$ and are assumed to be known. The Rayleigh number for the simulation is 11,691 which is calculated by

$$R_a = \frac{\rho\beta\Delta T L^3 \mathbf{g}}{\nu D}, \tag{18}$$

where $\Delta T = 20$ (difference between reference temperature T_0 and maximum temperature in the simulation environment), the values of the other parameters are given in table 2. The simulation aims to demonstrate the efficacy of the proposed method under circumstances with dynamic boundary conditions (BC). The initial room and floor temperature is assumed to be not known accurately. The window temperature is variable and also not known. The LBM-DA algorithm is operated to update the state based on simulated sensor measurements within the room. The changing BC is given in table 1.

An initial LBM simulation with the correct changing boundary conditions in Table 1 is carried out to simulate the correct temperature conditions in the model room which

Boundary condition (BC)	Window	Floor	Inlet
1	25 °C	15 °C	25 °C
2	20 °C	15 °C	25 °C
3	15 °C	15 °C	25 °C
4	5 °C	15 °C	25 °C
5	-5 °C	15 °C	25 °C

Table 1: Variable boundary conditions

we will refer to as the ground truth solution. The change in the window temperature is purely to highlight the efficacy of the proposed method, and is not based on real data. This is an exaggerated temperature change to show that if the boundary conditions change drastically then the proposed methods are able to steer back the simulation to reality. To evaluate the robustness of the proposed DA methods, the LBM-DA simulation is initiated with an incorrect initial boundary condition and is allowed to approximate the ground truth with four simulated sensor temperature measurements taken from the ground truth solution. The flow field at each boundary condition is shown by a cross section at $y = 10$ in Fig. 3. The flow fields show that as the window temperature drops the air mixing in the room changes. As expected at a higher temperature, there is evidence of short circuiting in the flow, with the warm inlet air being extracted from the room without significant mixing in the space. At lower window temperatures, the inlet air is cooled creating a greater degree of mixing into the lower zone of the room.

Parameter	LBM (Lattice Units)
Reference temperature T_0	5 °C
Wall temperature	10 °C
Temperature density (ground truth) T	10 °C
Temperature density (DA)	8 °C
Temperature density (without DA)	8 °C
Fluid density ρ	1
Prandtl number Pr	0.7
Kinematic viscosity ν	0.08
Thermal diffusivity D	0.114
Coefficient of thermal expansion β	0.000034
Box dimension Length (L) × Width (W) × Room Height (H)	$40 \times 20 \times 20$
Reynolds number	100
Number of grid points	16000
Lattice space Δx	1
Time step Δt	1
Acceleration due to gravity $\mathbf{g}$	(0, 0, -9.8)

Table 2: LBM parameters

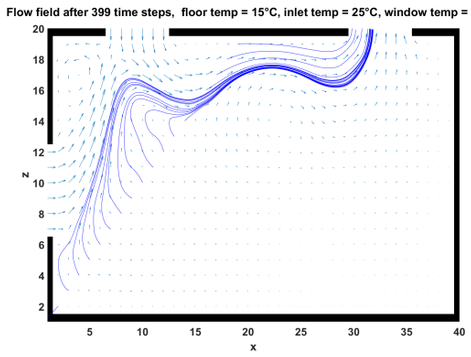
The performance of LBM-DA relies heavily on the accurate modeling of the background covariance matrix from the sensor data, which in this study has been modeled with the Gaussian correlation function. In real indoor scenarios, these can be obtained with an offline measurement campaign and appropriate modeling based on real data.

For this study, the temperature is assumed to be highly correlated on the x and y axis, while low correlation is assumed on the third axis. Fig. 4 shows the performance of three LBM simulations: i) the ground truth with correct boundary conditions, ii) LBM-3DVAR algorithm with an initial incorrect boundary condition and iii) LBM without implementation of DA and an initial incorrect boundary condition.

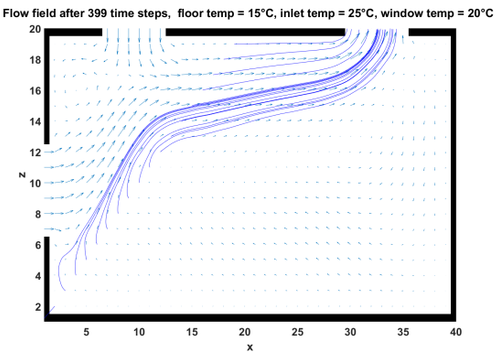
The parameters for the LBM are given in table 2 and the parameters for the LBM-3DVAR are given in table 3. The simulation begins with BC 1 (table 1) with the temperature density of the ground truth in Fig. 4a at 10 °C, the floor at 15 °C and inlet temperature at 25 °C. The simulation is run for 400-time steps before the BC changes to BC 2 and then allowed to run for another 400 time steps until the BC changes to BC 3 and so on. Incorrect initial conditions are given to the LBM-3DVAR and LBM without DA with temperature density, floor temperature at 8 °C and window temperature at 10 °C. Four simulated sensors are placed at locations [(20,10,19);(39,10,10);(20,2,10);(20,10,2)], these are represented by circles in the 2nd column of Fig. 4. When the BC changes from BC 1 to BC 2, the LBM-3DVAR and LBM without DA algorithms are ignorant to it. The LBM without DA continues to assume the wrong BC. The LBM-3DVAR model on the other hand aims to compensate for this change via sensor measurements from within the domain. The simulated sensor measurement in Fig. 4, 2nd column are taken from corresponding cell points in the ground truth with a small Gaussian noise added to represent sensor measurement error. Fig. 4 also presents the root mean squares error (RMSE) of the LBM-3DVAR and LBM without DA. The RMSE is obtained at the end of each of the 5 BC periods. The RMSE of the LBM-3DVAR is obtained as

$$\mathbf{RMSE}_{LBM-DA} = \sqrt{\left(\sum_{t=1}^{5}\sum_{i=1}^{n_x}\sum_{j=1}^{n_y}\sum_{k=1}^{n_z}\left(T^a_{i,j,k,t} - T_{i,j,k,t}\right)^2\right)/5}, \tag{19}$$

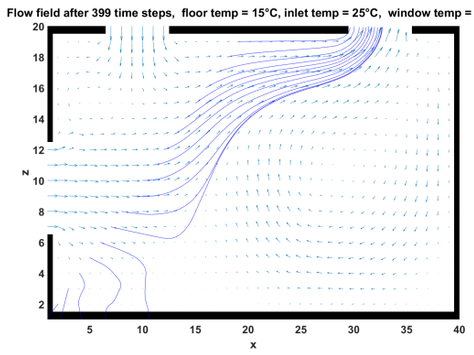
where T^a is the macro temperature of the LBM-DA and T is the macro temperature of the ground truth. The RMSE for LBM without DA is similar but with macro T^a replaced by the macro temperature with no DA. It can be seen from Fig. 4 that the LBM-3DVAR approximates the ground truth with changing BC. This can also be observed from the RMSE comparison, where the RMSE initially with DA is not significantly less than when no DA is performed; this is because both simulations started with the same incorrect initial condition and DA did not perform enough cycles to approximate the ground truth. However as can be seen in Fig. 4 with more DA cycles the RMSE drops more rapidly than in the case without DA, which also coincides with the window temperature dropping. For the simulation in Fig. 4, 5 DA cycles are performed before each BC change at time step 100, 150, 200, 250 and 300.



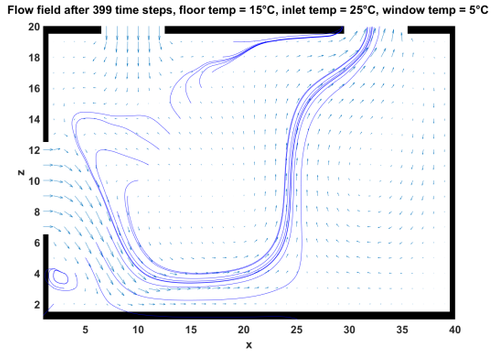
(a) Flow field of LBM with window temperature of 25 °C



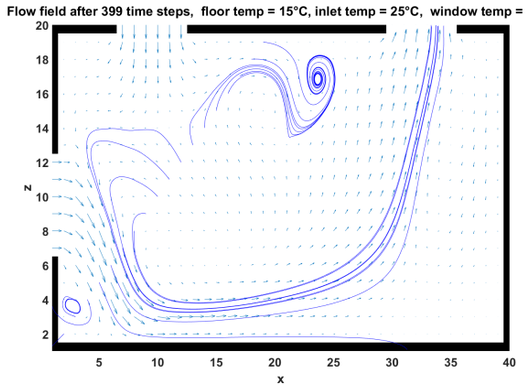
(b) Flow field of LBM with window temperature of 20 °C



(c) Flow field of LBM with window temperature of 15 °C



(d) Flow field of LBM with window temperature of 5 °C



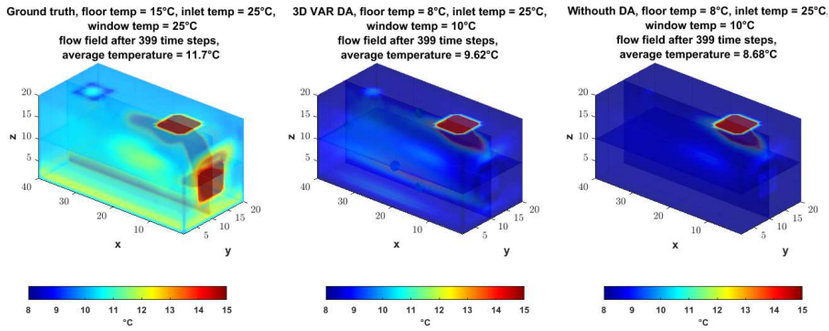
(e) Flow field of LBM with window temperature of -5 °C

Figure 3: Flow field on cross section at $y = 10$ showing velocity vectors as temperature boundary conditions change for the simulated ground truth

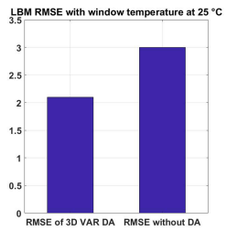
Parameter	value
$\mathbf{C}_{L_x}$	$\exp\left(-\frac{(i-j)^2}{n_x^2/4}\right)$
$\mathbf{C}_{L_y}$	$\exp\left(-\frac{(i-j)^2}{n_y^2/4}\right)$
$\mathbf{C}_{L_z}$	$\exp\left(-\frac{(i-j)^2}{3}\right)$
$\mathbf{S}$	$\mathbf{I}_N$
$\mathbf{R}$	$0.1\mathbf{I}_O$
ϵ	5

Table 3: 3DVAR parameters

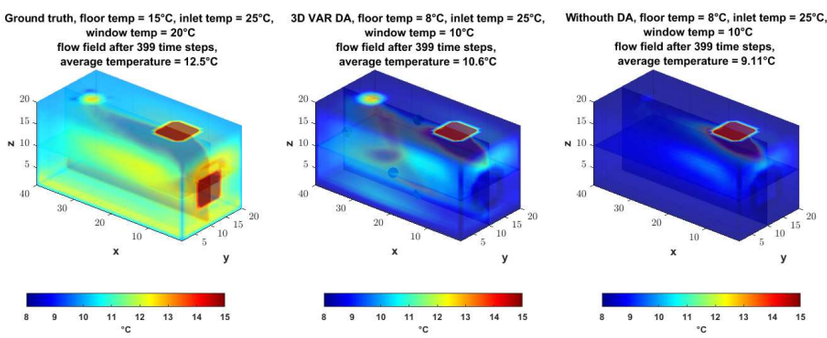
The RMSE shown in Fig. 5a to Fig. 5c is the average over BC1 to BC5 and unless mentioned, all parameters are kept same as for Fig. 4 and given in table 2 and table 3. Fig. 5a compares the RMSE of the LBM-3DVAR algorithm with an increasing



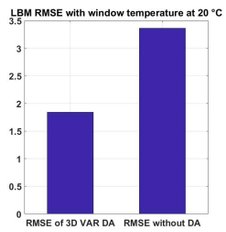
(a) Temperature field of LBM with window temperature of 25 °C



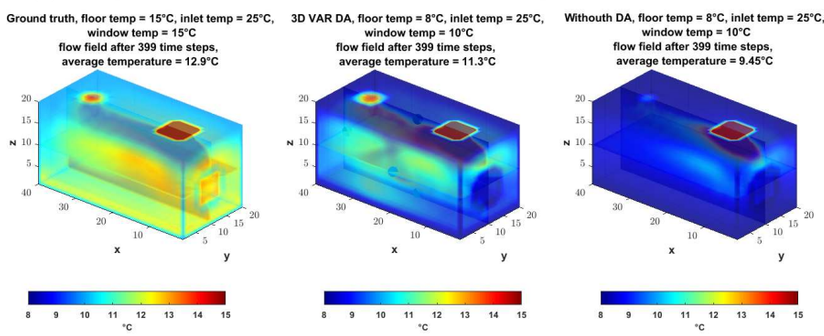
(b) RMSE



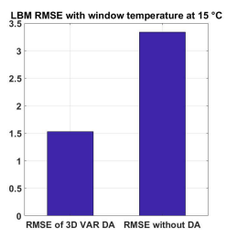
(c) Temperature field of LBM with window temperature of 20 °C



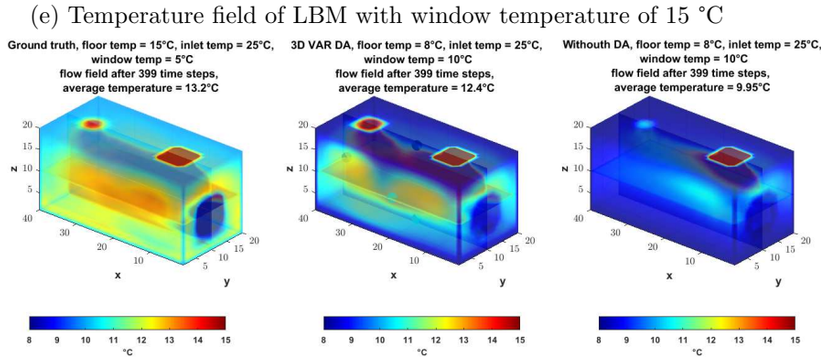
(d) RMSE



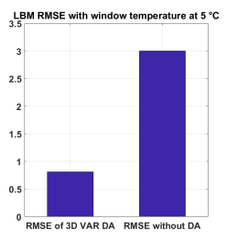
(e) Temperature field of LBM with window temperature of 15 °C



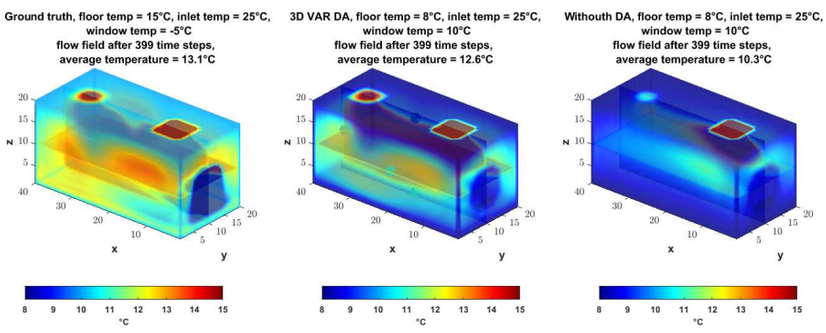
(f) RMSE



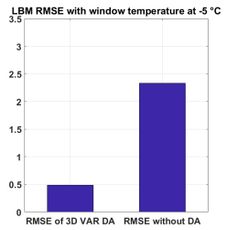
(g) Temperature field of LBM with window temperature of 5 °C



(h) RMSE



(i) Temperature field of LBM with window temperature of -5 °C



(j) RMSE

Figure 4: Temperature field on central planes showing results as temperature boundary conditions change for ground truth, with 3DVAR DA, and without DA

number of DA cycles within each BC. It is observed that with more DA cycles, the RMSE decreases due to more information being assimilated. However, this comes at a higher computational cost.

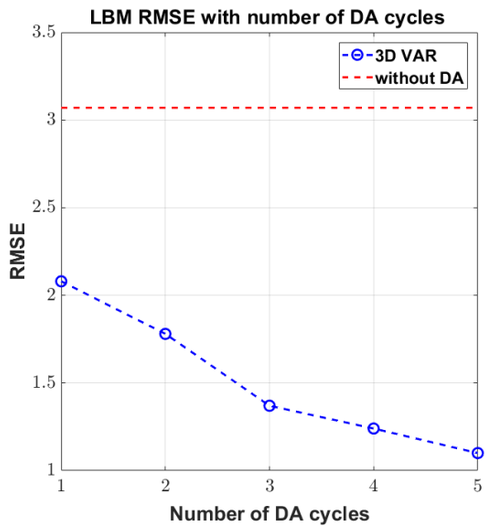
Number of sensors	location (x, y, z)
1	[(20,10,19)]
2	[(20, 10, 19); (39, 10, 10)]
3	[(20,10,19); (39,10,10); (20,2,10)]
4	[(20,10,19); (39,10,10); (20,2,10); (20,10,2)]
5	[(20,10,19); (39,10,10); (20,2,10); (20,10,2); (20,10,10)]

Table 4: Simulated sensor location in Fig. 4

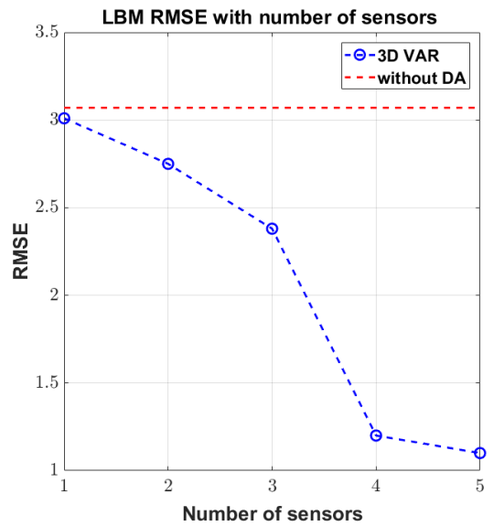
Fig. 5b compares the RMSE of the LBM-3DVAR with an increasing number of simulated sensor measurements. The location of the sensor used in the simulation is given in the table 4. It is noted that the location of sensors has a profound impact on the performance of LBM-DA. Although the optimal location of sensors for an indoor scenario is not the focus of this work, to capture the variability in the environment it is noted that sensors must be placed where they are able to cover the entire domain. In Fig. 5b, the simulated sensor locations that show the best performance in terms of lower RMSE are selected; this is achieved by trial and error. A few random combinations of sensor locations were tested and the locations generating the lowest RMSE were selected. These are by no means optimal sensor location, which could for example be achieved by using an optimisation algorithm such as simulation annealing or genetic algorithm. It is noted that no significant difference in RMSE is observed with using 4 and 5 sensors which could be due to the small dimension of the room. The results suggest that the transient nature of any indoor environment can be represented by a minimum number of sensors. This minimum number can for example be determined via simulating all sensor locations and comparing the RMSE of individual setup. It is also noted that using just one sensor has negligible performance improvement and in fact from our testing we observed that at certain sensors locations the DA performance is worse than when no DA is performed.

The performance of DA also depends on the quality of the sensor readings. With highly erroneous measurements the RMSE of the LBM-3DVAR increases. This is shown in Fig. 5c where the RMSE is plotted against the standard deviation (Std) in the simulated sensor measurements. The simulations are run 10 times independently for every Std value and the average RMSE is obtained. It can be seen that with highly erroneous measurements the performance of the LBM-3DVAR degrades.

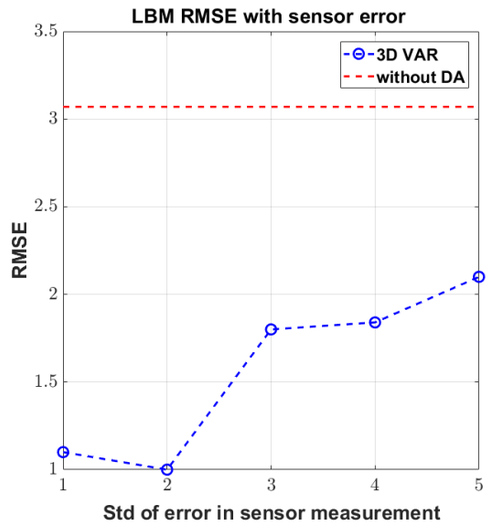
For the LBM-EnKF, the same LBM parameters as set in table 2 are used. The parameters for the EnKF algorithm are given in table 5. Fig. 6 shows the flow evolu-



(a) RMSE of 3DVAR DA with number of DA cycles



(b) RMSE of 3DVAR DA with number sensors



(c) RMSE of 3DVAR DA with sensor error

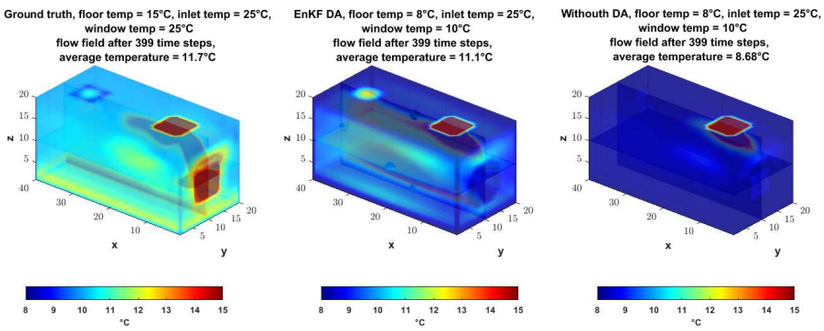
Figure 5: 3DVAR RMSE analysis

tion with changing BC and the corresponding flow of the LBM-EnKF, LBM without DA, and the RMSEs. Similar to the LBM-3DVAR algorithm, the LBM-EnKF also approximates the ground truth and performs considerably better than without DA.

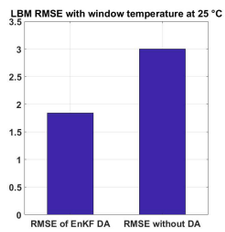
Parameter	value
Number of ensemble members (M)	10
L_x	$n_x^2/4$
L_y	$n_y^2/4$
L_z	3
L_s	$\mathbf{I}_O$
$\mathbf{R}$	$0.1\mathbf{I}_O$
λ	1.5

Table 5: EnKF parameters

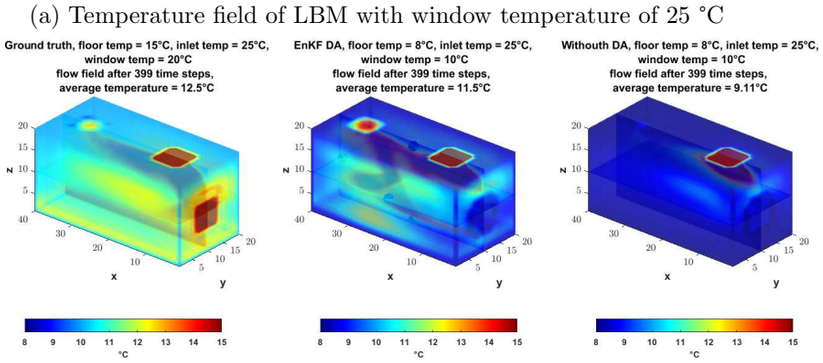
The RMSE results shown in Fig. 7a to Fig. 7d are the average over BC1 to BC5. Four sensors are used in this simulation (except for Fig. 7b) with the rest of the parameters kept same as for Fig. 6 and given in table 2 and table 5. Fig. 7a presents the performance of the LBM-EnKF with a different number of ensemble members. As



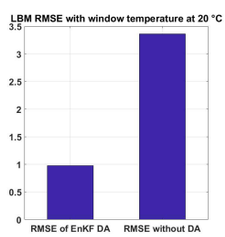
(a) Temperature field of LBM with window temperature of 25 °C



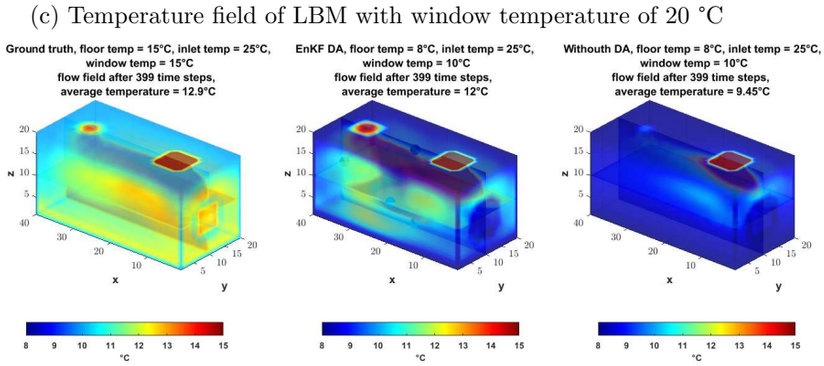
(b) RMSE



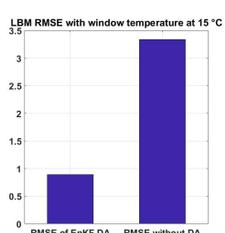
(c) Temperature field of LBM with window temperature of 20 °C



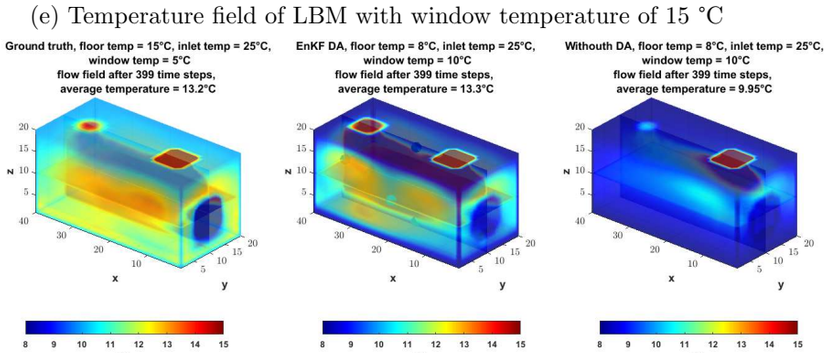
(d) RMSE



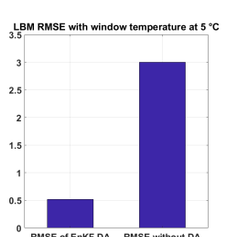
(e) Temperature field of LBM with window temperature of 15 °C



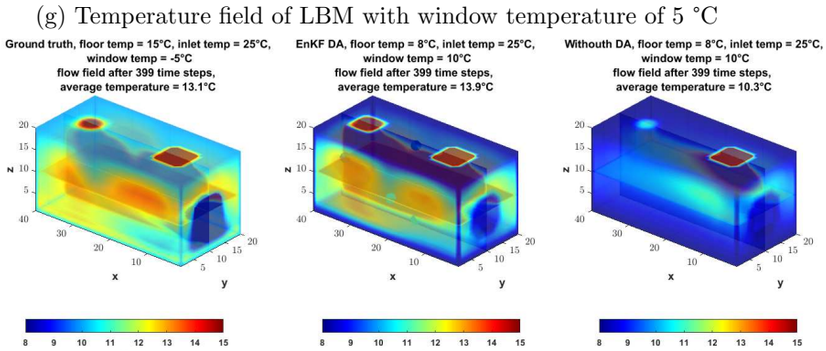
(f) RMSE



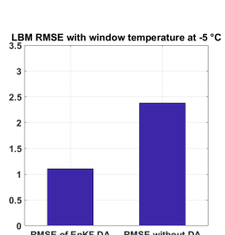
(g) Temperature field of LBM with window temperature of 5 °C



(h) RMSE



(i) Temperature field of LBM with window temperature of -5 °C



(j) RMSE

Figure 6: Temperature field on central planes showing results as temperature boundary conditions change for ground truth, with EnKF DA, and without DA

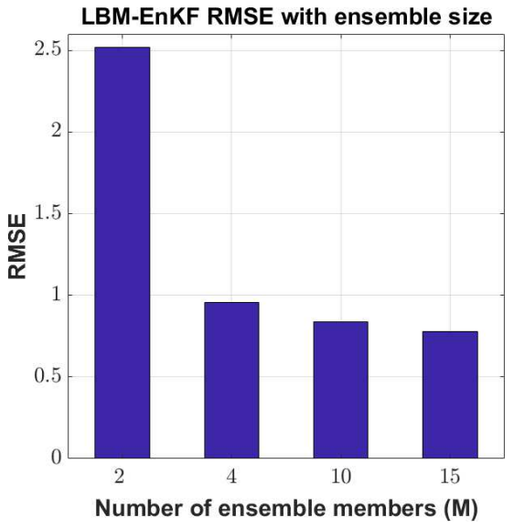
expected the performance improves with ensemble size. Yet again, this comes with an additional computational cost.

Fig. 7b shows the performance of the EnKF DA approach with a different number of simulated sensor measurements; the location of the sensors is shown in table 4. Like the LBM-3DVAR, the LBM-EnKF performs better with additional sensors, and again degraded performance is observed with only one sensor. Interestingly, using a single sensor in the LMB-EnKF case shows worse performance than with no DA. Fig. 7c shows the LBM-EnKF performance improvement with the number of DA cycles. It is noted that even with 1 DA cycle the RMSE is lower than with no DA. Also, it is noted that the performance does not improve significantly after 3 DA cycles. This is expected as the flow does not have enough time to evolve between DA cycles to assimilate any new information from the simulated sensor measurements. Fig. 7d shows the performance of the LBM-EnkF against the standard deviation (Std) in the sensor measurements. The simulations are run 10 times independently for every Std value and the average RMSE is obtained. The RMSE follows a similar trend as in the LBM-3DVAR case with highly erroneous measurements resulting poor performance.

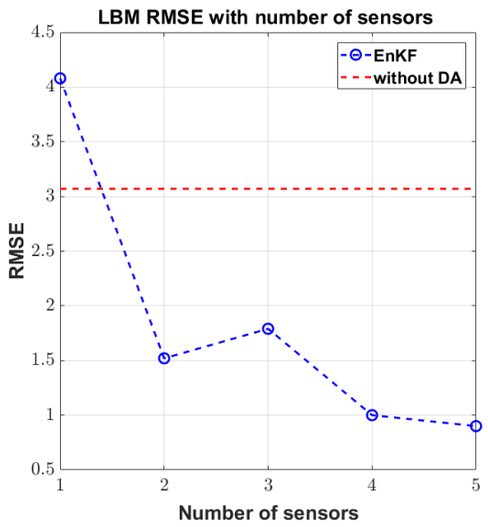
5. Discussion

Predicting accurate thermal conditions and IAQ is becoming exceedingly important for commercial and domestic buildings especially in light of the COVID-19 pandemic. Information on how the indoor temperatures and IAQ will change with changes in ventilation, thermal conditions, or with occupancy will be vital for building owners and designers. The proposed methods in this study provide accurate real time temperature forecast due to faster simulation times of the LBM over conventional CFD and also periodic corrections of the DA cycles mitigate the errors due to miss-specified boundary conditions in transient indoor environments. The choice between LBM-3DVAR and LBM-EnKF is a trade-off between accuracy and computational cost; with the LBM-EnKF outperforming LBM-3DVAR with large ensemble size. There is also a trade-off between grid size, dimension of the indoor environment and computational cost, with much finer grid or larger rooms requiring higher computational overhead both for the LBM and DA algorithms; in such cases, the LBM-3DVAR is preferred.

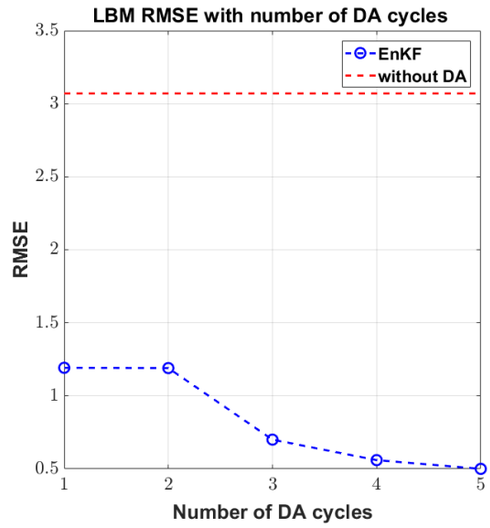
To demonstrate the viability of the LBM-DA method and the performance of the two DA algorithms we have focused this study on computational simulation only. This allows an evaluation of how the airflow parameters, DA method and simulated sensor locations affect the outcomes of the model without the complexity added by noisy sensor measurements and more uncertain flow fields that would be present in a real environment. We have also compared the DA models to a simulated ground truth



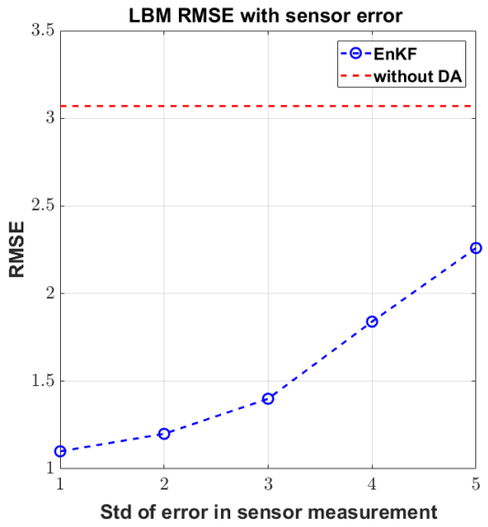
(a) RMSE of EnKF with number of ensemble members



(b) RMSE of EnKF DA with number of sensors



(c) RMSE of EnKF DA with number of DA cycles



(d) RMSE of EnKF DA with sensor error

Figure 7: EnKF DA RMSE analysis

that is generated using the LBM model. While this isn't validated against a real-world environment, we have confidence from previous studies that the LBM results are realistic. We have compared LBM to a traditional large eddy simulation (LES) methodology for a small room flow similar in size to the case in the current paper, and shown that the LBM model can reliably simulate transient indoor airflows and capture both velocity and temperature characteristics [19]. We have also compared the LBM approach with both LES simulations and experimental data for an urban natural ventilation flow case, and again shown the LBM model to be accurate with significantly faster computational times [20].

Several open research questions still need to be addressed to develop the approaches presented here to be used in real-world applications. Most importantly, the design of the background covariance matrix which in this work has been based on a Gaussian function. Other similar models such as the Gaspari-Cohn correlation function [13] could also be used. To capture the true correlations between cell points information

regarding the room's ventilation flow pattern should be incorporated. Moreover, in this study, only a single parameter i.e., the temperature is used, for multiple parameter formulation, the correlations between these need to be modeled. A simplistic approach would assume these parameters are independent however extensive data collection in buildings could conclusively determine these dependencies. If the aim is the real-time prediction of IAQ parameters, then attention needs to be given to the trade-off between performance and speed. In particular, how many sensors would suffice to accurately update the flow prediction with DA. It was shown in this work that even for a simple room, one sensor performs almost as poorly (or worse than) as no DA, and at the same time, the addition of more sensors after an optimum number of the sensor does not provide any significant improvement.

Then there is the issue of sensor placement which is another important and interesting topic that needs exploration [12]. In real buildings, there are physical constraints on where these sensors can be placed. Analytical solutions may be intractable however brute force methods or optimization algorithms such as the genetic algorithm could be used to arrive at the optimal sensor location. Furthermore, it has been shown that highly erroneous measurements degrade the DA performance, which may mean that the approach is better suited to some parameters over others. For example temperature and humidity changes in buildings tend to be gradual, while parameters relating to pollutants such as CO_2 or particulate matter may fluctuate more significantly. Although very high errors in the measurement are unrealistic, cheap low-quality sensors may potentially divert the flow prediction away from reality. Due the lack of real data, this study is limited to simulated sensor measurement only and thus has the limitation that it does not present the error in sensor measurements and its impact on the effectiveness of the proposed algorithms.

A final trade-off that needs to be considered is the number of DA cycles vs performance. For large dimensions, the DA cycles are computationally expensive, which may limit application in larger or more complex spaces. To determine, how often one requires to update the flow prediction will depend on how transience of the indoor environment and also on the accuracy/speed requirement.

Although this study performs the simulations for a simple office environment, the proposed methods can potentially be extended for more complex indoor scenarios. Moreover, the parameter used in this study is temperature, however other IAQ parameters such as CO_2 and relative humidity can be simulated similarly. High CO_2 prediction which is driven by sensor measurements can indicate a high occupancy in the building relative to the ventilation rate, and hence warrant preventive measures to mitigate adverse effects such as the risk of infection transmission. In this regard, sensors such as developed by us in [31] could be used to wirelessly transmit IAQ data and

update the LBM algorithm. Other applications of the proposed work include demand controlled ventilation, early detection of accidents such as gas leaks or fire, and energy conservation.

6. Conclusion

This study presents a new data assimilation approach for near real-time prediction of temperature coupled with sensor data for improving predictive accuracy in indoor environmental conditions. The LBM was used for flow and temperature prediction and coupled with two DA algorithms. Key conclusions from the study are:

- Both the LBM-3DVAR and LBM-EnKF approaches are able to effectively use sensor data to correct the flow simulation as boundary conditions change, and showed superior performance compared to stand-alone LBM with variable BC;
- The accuracy of both models improves with an increased number of sensors, however for the case presented here there was little benefit in using more than 4 sensor data points;
- The DA simulations with a single sensor data point are not accurate and could potentially lead to worse predicted conditions than the LBM simulation without DA;
- The accuracy of DA simulations are sensitive to sensor error and are less accurate with highly varying sensor data;
- The model performance is dependent on the design of the background covariance matrix and therefore understanding of correlations between sensor data in an indoor environment is important to enable accurate model predictions.
- The results obtained in this study have to be seen in light of some limitations. First, simulated sensor measurements were assumed instead of real sensor data. Real data may impact the performance of the proposed methods. Second, a simple office environment is considered instead of a more complex indoor scenario with more convoluted flow patterns.

The model represents a first step in developing accurate sensor driven real-time predictive control for indoor environments. Future work aims to employ the proposed techniques to indoor scenarios with real sensor data for a wider range of IAQ parameters.

Acknowledgment.

This work is funded by the EPSRC HECOIRA Project EP/P023312/1. The MATLAB code for the simulation is available at: https://github.com/Nav-101/LBM-DA-matlabcode.

References

[1] M. Asch, M. Bocquet, and M. Nodet. *Data assimilation: methods, algorithms, and applications*. Society for Industrial and Applied Mathematics, 12 2016. ISBN 978-1-611974-53-9.

[2] O. Asfour and M. Gadi. A comparison between cfd and network models for predicting wind-driven ventilation in buildings. *Building and Environment*, 42:4079–4085, 12 2007. doi: 10.1016/j.buildenv.2006.11.021.

[3] R. N. Bannister. A review of forecast error covariance statistics in atmospheric variational data assimilation. i: Characteristics and measurements of forecast error covariances. *Quarterly Journal of the Royal Meteorological Society*, 134(637):1951–1970, 2008. doi: 10.1002/qj.339. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.339.

[4] A. Bartoloni, C. Battista, S. Cabasino, P. S. Paolucci, J. Pech, R. Sarno, G. M. Todesco, M. Torelli, W. Tross, P. Vicini, R. Benzi, N. Cabibbo, F. Massaioli, and R. Tripiccione. Lbe simulations of rayleigh-bÉnard convection on the ape100 parallel processor. *International Journal of Modern Physics C*, 04(05):993–1006, 1993. doi: 10.1142/S012918319300077X. URL https://doi.org/10.1142/S012918319300077X.

[5] C. Beghein, Y. Jiang, and Q. Chen. Using large eddy simulation to study particle motions in a room. *Indoor air*, 15:281–90, 09 2005. doi: 10.1111/j.1600-0668.2005.00373.x.

[6] P. L. Bhatnagar, E. P. Gross, and M. Krook. A model for collision processes in gases. i. small amplitude processes in charged and neutral one-component systems. *Phys. Rev.*, 94:511–525, May 1954. doi: 10.1103/PhysRev.94.511. URL https://link.aps.org/doi/10.1103/PhysRev.94.511.

[7] G. Burgers, P. Jan van Leeuwen, and G. Evensen. Analysis scheme in the ensemble kalman filter. *Mon. Wea. Rev.*, 126(6):1719–1724, June 1998. ISSN 0027-0644. doi: 10.1175/1520-0493(1998)126⟨1719:ASITEK⟩2.0.CO;2. URL https://doi.org/10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2.

[8] S. Chen and G. D. Doolen. Lattice boltzmann method for fluid flows. *Annual Review of Fluid Mechanics*, 30(1):329–364, 1998. doi: 10.1146/annurev.fluid.30.1.329. URL https://doi.org/10.1146/annurev.fluid.30.1.329.

[9] E. Commision. Energy. 2019. URL https://ec.europa.eu/easme/en/news/sustainable-buildings-europe-s-climate-neutral-future.

[10] N. Delbosc, J. Summers, A. Khan, N. Kapur, and C. Noakes. Optimized implementation of the lattice boltzmann method on a graphics processing unit towards real-time fluid simulation. *Computers & Mathematics with Applications*, 67(2):462–475, 2014. ISSN 0898-1221. doi: https://doi.org/10.1016/j.camwa.2013.10.002. URL https://www.sciencedirect.com/science/article/pii/S0898122113006068. Mesoscopic Methods for Engineering and Science (Proceedings of ICMMES-2012, Taipei, Taiwan, 23–27 July 2012).

[11] G. Evensen. The ensemble kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4):343–367, 2003. ISSN 1616-7228. doi: 10.1007/s10236-003-0036-9. URL https://doi.org/10.1007/s10236-003-0036-9.

[12] A. D. Fontanini, U. Vaidya, and B. Ganapathysubramanian. A methodology for optimal placement of sensors in enclosed environments: A dynamical systems approach. *Building and Environment*, 100:145–161, 2016. ISSN 0360-1323. doi: https://doi.org/10.1016/j.buildenv.2016.02.003. URL https://www.sciencedirect.com/science/article/pii/S0360132316300403.

[13] G. Gaspari and S. E. Cohn. Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125(554):723–757, 1999. doi: 10.1002/qj.49712555417. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49712555417.

[14] Z. Guo, B. Shi, and C. Zheng. A coupled lattice bgk model for the boussinesq equations. *International Journal for Numerical Methods in Fluids*, 39(4):325–342, 2002. doi: 10.1002/fld.337. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/fld.337.

[15] G. Guyot, M. H. Sherman, and I. S. Walker. Smart ventilation energy and indoor air quality performance in residential buildings: A review. *Energy and Buildings*, 165:416–430, 2018. ISSN 0378-7788. doi: https://doi.org/10.1016/j.enbuild.2017.12.051. URL https://www.sciencedirect.com/science/article/pii/S0378778817324027.

[16] X. He and L.-S. Luo. Lattice boltzmann model for the incompressible navier-stokes equation. *Journal of Statistical Physics*, 88(3):927–944, 1997. ISSN 1572-9613. doi: 10.1023/B:JOSS.0000015179.12689.e4. URL https://doi.org/10.1023/B:JOSS.0000015179.12689.e4.

[17] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 2 edition, 2012. doi: 10.1017/9781139020411.

[18] M. Jin, W. Zuo, and Q. Chen. Improvements of fast fluid dynamics for simulating air flow in buildings. *Numerical Heat Transfer, Part B: Fundamentals*, 62(6):

419–438, 2012. doi: 10.1080/10407790.2012.724988. URL https://doi.org/10.1080/10407790.2012.724988.

[19] M. A. I. Khan, N. Delbosc, C. J. Noakes, and J. Summers. Real-time flow simulation of indoor environments using lattice boltzmann method. *Building Simulation*, 8(4):405–414, 2015. doi: 10.1007/s12273-015-0232-9.

[20] M.-F. King, A. Khan, N. Delbosc, H. L. Gough, C. Halios, J. F. Barlow, and C. J. Noakes. Modelling urban airflow and natural ventilation using a gpu-based lattice-boltzmann method. *Building and Environment*, 125:273–284, 2017. ISSN 0360-1323. doi: https://doi.org/10.1016/j.buildenv.2017.08.048. URL https://www.sciencedirect.com/science/article/pii/S036013231730402X.

[21] N. Klepeis, W. Nelson, W. Ott, J. Robinson, A. Tsang, P. Switzer, J. Behar, S. Hern, and W. Engelmann. The national human activity pattern survey (nhaps): a resource for assessing exposure to environmental pollutants. *Journal of exposure analysis and environmental epidemiology*, 11(3):231—252, 2001. ISSN 1053-4245. doi: 10.1038/sj.jea.7500165. URL https://doi.org/10.1038/sj.jea.7500165.

[22] Z. Li, Y. Chao, J. C. McWilliams, and K. Ide. A Three-Dimensional Variational Data Assimilation Scheme for the Regional Ocean Modeling System. *Journal of Atmospheric and Oceanic Technology*, 25(11):2074–2090, 11 2008. ISSN 0739-0572. doi: 10.1175/2008JTECHO594.1. URL https://doi.org/10.1175/2008JTECHO594.1.

[23] Z. Li, X. Cheng, W. I. Gustafson Jr., and A. M. Vogelmann. Spectral characteristics of background error covariance and multiscale data assimilation. *International Journal for Numerical Methods in Fluids*, 82(12):1035–1048, 2016. doi: 10.1002/fld.4253. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/fld.4253.

[24] A. C. Lorenc. Modelling of error covariances by 4d-var data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 129(595):3167–3182, 2003. doi: 10.1256/qj.02.131. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1256/qj.02.131.

[25] V. C. Mariani and A. da Silva. Natural convection: Analysis of partially open enclosures with an internal heated source. *Numerical Heat Transfer, Part A: Applications*, 52(7):595–619, 2007. doi: 10.1080/10407780701338423. URL https://doi.org/10.1080/10407780701338423.

[26] A. Megri and F. Haghighat. Zonal modeling for simulating indoor environment of buildings: Review, recent developments, and applications. *HVAC&R Research*, 13:887–905, 11 2007. doi: 10.1080/10789669.2007.10391461.

[27] L. Morawska, J. W. Tang, W. Bahnfleth, P. M. Bluyssen, A. Boerstra, G. Buonanno, J. Cao, S. Dancer, A. Floto, F. Franchimon, C. Haworth, J. Hogeling, C. Isaxon, J. L. Jimenez, J. Kurnitski, Y. Li, M. Loomans, G. Marks, L. C. Marr, L. Mazzarella, A. K. Melikov, S. Miller, D. K. Milton, W. Nazaroff, P. V. Nielsen, C. Noakes, J. Peccia, X. Querol, C. Sekhar, O. Seppänen, S. ichi Tanabe, R. Tellier, K. W. Tham, P. Wargocki, A. Wierzbicka, and M. Yao. How can airborne transmission of covid-19 indoors be minimised? *Environment International*, 142:105832, 2020. ISSN 0160-4120. doi: https://doi.org/10.1016/j.envint.2020.105832. URL https://www.sciencedirect.com/science/article/pii/S0160412020317876.

[28] T. R. Nielsen and C. Drivsholm. Energy efficient demand controlled ventilation in single family houses. *Energy and Buildings*, 42(11):1995–1998, 2010. ISSN 0378-7788. doi: https://doi.org/10.1016/j.enbuild.2010.06.006. URL https://www.sciencedirect.com/science/article/pii/S0378778810001994.

[29] B. J. Palmer and D. R. Rector. Lattice boltzmann algorithm for simulating thermal flow in compressible fluids. *Journal of Computational Physics*, 161(1):1 – 20, 2000. ISSN 0021-9991. doi: https://doi.org/10.1006/jcph.2000.6425. URL http://www.sciencedirect.com/science/article/pii/S0021999100964258.

[30] F. Rabier. Overview of global data assimilation developments in numerical weather-prediction centres. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3215–3233, 2005. doi: 10.1256/qj.05.129. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1256/qj.05.129.

[31] N. Salman, A. H. Kemp, A. Khan, and C. Noakes. Real time wireless sensor network (wsn) based indoor air quality monitoring system. *IFAC-PapersOnLine*, 52(24):324 – 327, 2019. ISSN 2405-8963. doi: https://doi.org/10.1016/j.ifacol.2019.12.430. URL http://www.sciencedirect.com/science/article/pii/S2405896319323389. 5th IFAC Symposium on Telematics Applications TA 2019.

[32] J. Stone. *Bayes' Rule: A Tutorial Introduction to Bayesian Analysis*. Sebtel Press, 06 2013. ISBN 978-0956372840. doi: 10.13140/2.1.1371.6801.

[33] A. Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*, volume xii. Society for Industrial and Applied Mathematics, 01 2005. ISBN 978-0-89871-572-9. doi: 10.1137/1.9780898717921.

[34] D. J. Tritton. *Physical Fluid Dynamics*. Springer, Dordrecht, 1977. ISBN 978-94-009-9992-3.

[35] L. Wang and Q. Chen. Validation of a coupled multizone-cfd program for building airflow and contaminant transport simulations. *HVAC& R Research*, 13 (2):267–281, 2007. doi: 10.1080/10789669.2007.10390954. URL https://www.tandfonline.com/doi/abs/10.1080/10789669.2007.10390954.

[36] G. Welch and G. Bishop. An introduction to the kalman filter. Technical Report 95-041, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, 1995. URL http://www.cs.unc.edu/~welch/kalman/kalmanIntro.html.

[37] C. Wunsch. *Discrete Inverse and State Estimation Problems: With Geophysical Fluid Applications*. Cambridge University Press, 2006. doi: 10.1017/CBO9780511535949.

[38] Z. Zhai. Application of computational fluid dynamics in building design: Aspects and trends. *Indoor and Built Environment*, 15(4):305–313, 2006. doi: 10.1177/1420326X06067336. URL https://doi.org/10.1177/1420326X06067336.

[39] B. Zhao and P. Guan. Modeling particle dispersion in personalized ventilated room. *Building and Environment*, 42(3):1099 – 1109, 2007. ISSN 0360-1323. doi: https://doi.org/10.1016/j.buildenv.2005.11.009. URL http://www.sciencedirect.com/science/article/pii/S0360132305004816.

[40] W. Zuo and Q. Chen. Fast and informative flow simulations in a building by using fast fluid dynamics model on graphics processing unit. *Building and Environment*, 45(3):747 – 757, 2010. ISSN 0360-1323. doi: https://doi.org/10.1016/j.buildenv.2009.08.008. URL http://www.sciencedirect.com/science/article/pii/S036013230900208X.

[41] N. Šajn. Energy efficiency of buildings a nearly zero-energy future. *Briefing, EPRS — European Parliamentary Research Service*, May 2016. URL https://www.europarl.europa.eu/thinktank/en/document.html?reference$=$EPRS_BRI(2016)582022.