Model-based Bayesian inference of the ventilation distribution in patients with Cystic Fibrosis from multiple breath washout, with comparison to ventilation MRI

Carl A. Whitfield*[1,2], Alexander Horsley[1], Oliver E. Jensen[2], Felix C. Horn[3], Guilhem J. Collier[3], Laurie J. Smith[3], Jim M. Wild[3]
[1]Division of Infection, Immunity and Respiratory Medicine, University of Manchester, Manchester, UK
[2]Department of Mathematics, University of Manchester, Manchester, UK
[3]POLARIS Imaging, Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield, Sheffield, UK

Running head: Bayesian inference of ventilation distribution from MBW

*Corresponding author: Carl A. Whitfield
E-mail: carl.whitfield@manchester.ac.uk
Address: Education and Research Centre
Division of Infection, Immunity & Respiratory Medicine,
The University Hospital of South Manchester
Wythenshawe
Manchester
M23 9LT

Abstract

Background: Indices of ventilation heterogeneity (VH) from multiple breath washout (MBW) have been shown to correlate well with VH indices derived from hyperpolarised gas ventilation MRI. Here we report the prediction of ventilation distributions from MBW data using a mathematical model, and the comparison of these predictions with imaging data.

Methods: We developed computer simulations of the ventilation distribution in the lungs to model MBW measurement with 3 parameters: $\sigma_{V,}$ determining the extent of VH; V_0, the lung volume; and V_D, the dead-space volume. These were inferred for each individual from supine MBW data recorded from 25 patients with cystic fibrosis (CF) using approximate Bayesian computation. The fitted models were used to predict the distribution of gas imaged by ^{3}He ventilation MRI measurements collected from the same visit.

Results: The MRI indices measured ($I_{1/3}$, the fraction of pixels below one-third of the mean intensity and I_{CV}, the coefficient of variation of pixel intensity) correlated strongly with those predicted by the MBW model fits ($r = 0.93, 0.87$ respectively). There was also good agreement between predicted and measured MRI indices (mean bias $\pm$ limits of agreement: $I_{1/3}$: 0.002 ± 0.112 and I_{CV}: -0.001 ± 0.293). Fitted model parameters were robust to truncation of MBW data.

Conclusion: We have shown that the ventilation distribution in the lung can be inferred from an MBW signal, and verified this using ventilation MRI. The Bayesian method employed extracts this information with fewer breath cycles than required for LCI, reducing acquisition time required, and gives uncertainty bounds, which are important for clinical decision making.

New and Noteworthy

This paper demonstrates that the ventilation distribution observed by ventilation MRI in cystic fibrosis patients can be inferred using multiple breath washout data. The Bayesian method used quantifies prediction uncertainty. This has the potential to be used in the analysis of washout data in the clinic to give greater physiological insight more efficiently. The predictions also remained robust to truncation of the washout dataset, meaning that data-capture time can be significantly reduced using this method.

1. Introduction

Ventilation heterogeneity (VH) refers to the unevenness of inspired air distribution in different lung regions during breathing. It is an early and prominent feature of lung diseases such as cystic fibrosis (CF), bronchiectasis, asthma and COPD[1–4]. Clinical assessment of VH in the lung is performed by the multiple breath washout test (MBW), from which the most commonly used primary outcome is the lung clearance index (LCI)[5]. LCI is now well established, particularly in CF where it is a sensitive, robust measure of early disease, and responsive to clinical status[6]. However, LCI utilises a relatively small proportion of the gas washout data collected: the alveolar gas concentrations, and specifically those preceding the first and last washout breaths. Alternative VH indices from MBW, such as moment ratios[7] and phase-III slopes[8], have been developed but are less commonly used as they are more sensitive than LCI to other factors besides ventilation heterogeneity[5,9], including variations in

tidal volume and gas diffusivity. One current limitation to widespread clinical use of LCI as a pulmonary function test is the time required for data collection[5]. Although LCI appears robust to earlier test thresholds to some extent, this may be at the cost of reduced sensitivity and the practice is not widespread[10–13].

Hyperpolarised gas ventilation MRI allows for visualisation of the ventilation distribution in the lung[14]. Patients inhale a bolus of hyperpolarised tracer gas (^{129}Xe or ^{3}He) which is imaged during a short breath hold (as in this study), dynamically during the respiratory cycle[15] or over several cycles[16]. This enables the identification of small ventilation defects that are not detected by spirometry (FEV_1) or LCI, and the method is therefore highly sensitive to early lung disease progression[1,17,18]. Ventilation MRI also provides a 3-dimensional representation of the distribution of lung disease, allowing regional changes to be identified and tracked. Furthermore, MRI can identify lung regions where tidal flow is entirely obstructed, which would not normally contribute to the MBW signal. This powerful technique is therefore potentially more informative about the nature and severity of lung disease. We have previously reported a good correlation between MRI markers of airway obstruction and LCI in patients with CF[19]. The availability of this paired data, previously published and analysed in [17,19], presents a unique opportunity to link clinically usable measures of gas washout with detailed lung imaging in order to better inform understanding of MBW and to develop more sophisticated washout metrics.

Mathematical models have been used previously to predict the lung ventilation distribution from inert gas washout tests. Continuous models[20–22] used functional fitting of end-tidal nitrogen, but were constrained by the requirement for equal breath sizes during the test. The widely-employed method of Lewis et al.[22] involving 50-compartments with specific-ventilation equally spaced on a log-scale. The volume of each compartment is fitted using a LASSO-type (least absolute shrinkage and selection operator) approach, with an empirical smoothing term between neighbouring compartments. An initial estimate is calculated by linear least-squares fit, then a local gradient descent method is deployed with the constraint that no compartments have negative volume. The multiple inert gas elimination technique[23] (MIGET) involves an estimation of the ventilation distribution using similar methods to those originally proposed by Lewis et al. but with several measurement gases. Similarly, Prisk et al. used the Lewis method to measure the effects of micro-gravity on VH[24]. Two-compartment models have also been developed[25], which allow all parameters to be identifiable without smoothing assumptions, but lack the resolution to recover a realistic ventilation distribution.

More recent algorithms have refined these approaches through a variety of techniques, including systematic variation in the number of compartments and model assumptions[26], and the introduction of dead-space compartments and acinar-mixing effects[27]. A more detailed compartmental model of ventilation has also been developed that accounts for CO_2 and N_2 exchange[28] to better estimate unventilated space. However, this particular model is currently too computationally intensive for implementation with Bayesian fitting methods (and thus lacks robust uncertainty bounds for the parameters), and was developed for a bespoke high-precision device. Furthermore, Bayesian inference has shown promise in extracting more useful indices from MBW data using only the end-tidal concentrations[29]. While several studies have shown that washout curves can be predicted from structure-function measurements including imaging[30–32], and the model of Lewis et al. has been refined through comparison with benchtop experiments[33], no studies have yet validated predictions

of the ventilation distribution in a clinical setting using direct imaging measurement in the same patients.

The software we have developed addresses the identified shortcomings. First, the algorithm can be used on directly-measured tracer gas and flow data from the Innocor™ SF_6 device (Innovision, Glamsbjerg, Denmark)[34], which has already been used in several studies, and can applied to measurements from other devices. Secondly, we have used Bayesian parameter identification to fit the compartmental model to data for individual patients, allowing for robust prediction estimates of uncertainty based on approximating the probability of all plausible solutions. Finally, we have also used independent measurements of VH from ventilation MRI to directly test the validity of the inferred parameters. This makes the physiological assumption that the heterogeneity of the hyperpolarised gas distribution is ventilation-dominated, and only negligibly affected by collateral diffusive mixing.

The aim of this study was to develop computational software to predict the ventilation distribution in individual subjects using raw MBW data and to use this to produce robust indices of VH that directly correspond to those measured directly with ventilation MRI. A secondary aim was to test whether the measures derived using this new method were robust at shorter test times.

2. Methods

Tables 1, 2, and 3 introduce and the mathematical symbols that are used in this section prior to their first use in the text. Table 1 shows the subscript indices used in the modelling and data analysis and so represent integer quantities. Table 2 shows quantities associated with the MBW and MRI data processing, while Table 3 shows simulation parameters and variables of the compartmental lung model,

Index	Use
i	Lung units and dead space compartments in the compartmental lung model.
j	Volume elements in the compartmental lung model.
k	Breath number in processed MBW data.
n	Time-step in processed MBW data.
s	Voxel in MRI data (simulated or observed).

Table 1: A list of subscript indices used in the text, and what they represent.

2.1 Study design and recruitment

As part of a longitudinal observational study, children and adults with CF were recruited from three UK specialist centres[18,19]. At recruitment, patients had to be over the age of five years, be clinically stable for four weeks prior to their visit and achieve an FEV_1 >30% predicted within the previous six months. This study was approved by the Yorkshire and Humber - Leeds West Research Ethics Committee (REC reference: 16/YH/0339). Parents/guardians of children and all adult patients provided written informed consent. For this analysis we have used data from a single visit when patients completed both MRI and MBW, including supine MBW (since this corresponds to the same position as the MRI procedure, and so minimises the effect of changes due to body position or gravity[35,36]).

All of the experimental data in this paper is from a single visit of the studies[17,19], and therefore an extensive analysis of the MBW and MRI data has been published previously.

Symbol	Description
$I_{1/3}$	Proportion of the masked He^3 MRI image with voxel brightness less than 1/3 of the mean.
I_{CV}	Coefficient of variation in brightness of masked He^3 MRI image.
$\mathrm{LCI_z}$	Lung clearance index measured by MBW (mean from 3 test repeats), $z = 2.5, 5, 10, 20, 40$ is the termination threshold used in % of equilibrium SF_6 concentration.
V_{FRC}	Functional residual capacity (L) measured by MBW (mean from 3 test repeats).
V_{FDS}	Fowler dead-space (L) measured from CO_2 curve during MBW (mean from 3 test repeats – each test is the median value of all washout breaths).
$\mathrm{V}_{\mathrm{FRC}}^{(\mathrm{pleth})}$	Functional residual capacity (L) measured by plethysmography
$\tilde{V}_0$, $\tilde{V}_D$	Initial guess for model parameters V_0 and V_D, used to set model priors.
$V_k^{(\mathrm{inh})}$, $V_k^{(\mathrm{exh})}$	Volume of inhalation and exhalation respectively for breath k of an MBW test.
N_{II}, N_{III}	Number of data-points used to represent phase-II and phase-III parts of each breath ($N_{II} = N_{III} = 4$ used throughout).

Table 2: Symbols used for quantities derived from or used in the processing of experimental data.

2.2 Ventilation MRI

Ventilation MRI was performed on a 1.5T GE HDx scanner (GE, Milwaukee, WI, USA) using hyperpolarised helium-3 (^{3}He) using a transmit-receive vest coil (CMRS, Milwaukee, WI, USA) and a three-dimensional (3D) steady state free precession ventilation imaging sequence as described previously[16]. Images used in this study were acquired during a breath-hold at end-inspiratory tidal volume following the inhalation of a predetermined fixed volume of test gas from their resting functional residual capacity (FRC). The volume of gas was titrated based on the subject's height and consisted of scaled doses of ^{3}He balanced with nitrogen[19].

Contiguous ventilation MR images of the coronal plane were acquired with slice thickness of 5mm and pixel size ranging from 2.73 x 2.73mm to 3.28 x 3.28 mm (depending on patient lung size). For each slice, a mask was manually derived from ^{1}H images (acquired during the same breath hold as the ^{3}He MRI) to determine which pixels corresponded to positions inside the lung cavity (excluding visible airways). The masked images were eroded by one pixel to avoid edge effects. The intensity (brightness) of each pixel is taken as a relative measure of the gas concentration at that point. This is used as a proxy for the local ventilation rate, and is used to characterise VH in two ways (see figure 1):

- Measuring the 'poorly ventilated' fraction of the lung by $I_{1/3}$. That is, the fraction of pixels with intensity less than one-third of the mean (as used previously to define ventilation defects[37]).
- Calculating the total coefficient of variation of the normalised pixel intensity, I_{CV}.

Note that both of these indices are measures of global ventilation heterogeneity, and we do not use more common indices such as CV_{mean}[38] or ΔR[39] that quantify spatial heterogeneity, or any of the information regarding regional ventilation patterns. This is because the aim of this study is to recover measures of global VH from MBW data, and to use MR to verify these.

Since the MBW data is collected at the mouth, it is a mixed signal from all of the lung regions, and hence distinguishing regional contributions is practically impossible unless washouts are performed in a range of postures[40]. Extending this method to account for posture changes and regional ventilation differences is a promising area of future research.

2.3 Pulmonary function tests

MBW was performed using a modified open-circuit Innocor device using 0.2% sulphur hexafluoride (SF_6) trace gas in air. MBW tests were collected in triplicate in the supine position. Spirometry and body plethysmography were performed to international standards[41,42] using a PFT Pro (Vyaire, Basingstoke, UK). All tests were performed on the same day. Either MBW or MRI was performed first, followed by the other. Spirometry was always performed last to minimise the influence of VH redistribution due to a forced manoeuvre.

MBW data were analysed using a software package for Igor Pro v6 (Wavemetrics Inc., Lake Oswego, OR, USA) as previously described[10]. We extracted from this analysis the following parameters for each patient: $LCI_{2.5}$, lung volume at FRC (V_{FRC}), and Fowler dead-space from CO_2 curves (V_{FDS})[43]. We also re-computed LCI for different termination thresholds (5%, 10%, 20%, 40% of initial concentration).

Symbols	**Description**
$\{V_0, V_D, \sigma_v\}$	Compartmental lung model parameters setting the FRC, dead-space volume and VH respectively.
$\{v_{ij}, c_{ij}^{(p)}, c_{ij}^{(d)}, g_{ij}\}$	Properties of the volume element j in dead-space compartment i: gas volume, inert gas concentrations (distal end and proximal end), and inert gas volumes respectively.
$\{C_i, V_i, G_i\}$	Properties of the lung unit i: inert gas concentration, gas volume and inert gas volume respectively.
N_c	Number of lung units in the model ($N_c = 50$ used throughout)
$v_{\max}$	Re-discretisation volume used for mixing-step ($v_{\max} =$ used throughout)

Table 3: Symbols used to represent parameters and quantities in the compartmental lung model.

2.4 Compartmental lung model

Figure 2 provides schematic overview of the model-fitting process, which this section describes in detail.

2.4.1 Processing of MBW data for model fitting: The MBW data we have used was collected using the Innocor SF_6 device (Innovision, Odense, Denmark) and converted to plain-text format. Processing of the raw data was carried out using a custom-built C++ program in line with recommendations[5]. Corrections for re-breathed SF_6 were not applied as these are simulated in the model. In summary:

1. Gas traces from each MBW are corrected for the flow-gas delay (as measured during device calibration), which is taken to be the same for the CO_2 and SF_6 datasets.
2. The data are separated into inhalations and exhalations, with a filtering step to ensure that flow fluctuations near to zero are not counted as separate breaths.
3. Exhaled volumes are corrected to body temperature, pressure, water vapour saturated (BTPS) by multiplying the measured flow by 1.016. Then, inhaled volumes are all scaled by a constant factor to give unitary respiratory quotient for the whole test.
4. Fowler dead space (FDS) volume was measured using the CO_2 traces, and functional residual capacity (FRC) approximated as outlined in Robinson et al. 2013[5]. The median FDS over all valid breaths in each test was used as the test average, then $\widetilde{V}_D$ was taken as the mean of this over all the test repeats of an individual. Similarly, $\widetilde{V}_0$ was the mean FRC measured in this step over all test repeats.
5. Finally, exhalations are down-sampled to 10 points per breath: these were the start and end points, 4 evenly spaced points in phase II (defined as exhaled volume between $\widetilde{V}_D/2$ and $3\,\widetilde{V}_D/2$), and 4 in phase III (defined as between $3\,\widetilde{V}_D/2$ and $0.95\,V_k^{(exh)}$), where $V_k^{(exh)}$ is the total exhalation volume of breath k. This down-sampling was achieved by linearly interpolating from the nearest concentration measurements in the dataset. Inhalations were down-sampled to a single step between inhalations, as the gas trace is approximately zero during washout and therefore not used in the fitting process. We label the time-steps with index n and the associated volume change as $\Delta V_n{}'$.

2.4.2 Compartmental lung model: As in Bates and Peters[27], the computational lung model comprises $N_c = 50$ lung units of equal size with total volume V_0 at FRC, with each compartment connected to independent dead-space compartments of equal size with total volume V_D. As in Mountain et al.[28], the relative inflation rate x of each compartment is drawn from a lognormal probability distribution with unit mean

$$P(x) = \frac{1}{x\sigma_V\sqrt{2\pi}}\exp\left(-\frac{(ln\,x-\mu)^2}{2\sigma_V^2}\right), \qquad where \qquad \mu = -\frac{\sigma_V^2}{2}.$$

where σ_V is the VH parameter. Once the x values have been drawn ($x_i \sim P(x)$ for $i = 1,..,N_c$), they are normalised so that the mean is $\bar{x} = 1$. The model simulates advection of the inhaled/exhaled volumes measured from MBW through the dead-space and into/out of the lung units via the transport of discrete volume elements. These volume elements therefore effectively form a Lagrangian grid for the 1D network of dead-space components. Each element j is indexed sequentially from distal to proximal end in each dead-space

compartment $i = 0, \ldots N_c$ (where $i = 0$ is the common-dead space and $i = 1, .., N_c$ the private dead-space compartments corresponding to the acinar units with the same index). An element has volume v_{ij}, and concentration values $c_{ij}^{(p)}$ and $c_{ij}^{(d)}$ defined at its proximal (mouth) and distal (acinar) ends respectively. The elements are assumed linear in concentration and so have inert gas volume $g_{ij} = v_{ij}(c_{ij}^{(p)} + c_{ij}^{(d)})/2$. Each acinar unit i exhibits instantaneous gas-mixing, so the gas concentration in the unit at time-step n is $C_i(n) = G_i(n)/V_i(n)$, where G_i is the volume of inert gas in the unit, and V_i is the total gas volume of the unit. Synchronous ventilation of the units is also assumed.

At each time-step , the volume flux is perturbed by random measurement noise such that $\Delta V_n = \Delta V_n{}'(1 + \varepsilon)$ where ε is drawn from a Gaussian distribution with mean 0 and standard deviation $\sigma_f = 0.01$ (based on upper estimates of the signal-to-noise ratio from Horsley et al. 2007[6]). As shown in figure 3, an inhalation or exhalation is then simply computing by shunting volume elements through the dead-space. A detailed description of this process is given in Appendix A.

For each time-step n of the simulation, concentration measurements are taken at the distal end of this dead-space $\gamma_n = c_{0j'}(n) + \eta$, where j' is the distal-most element of the common dead-space, and $\eta \sim N(0, \sigma_c)$ is measurement noise where $\sigma_c = 2 \times 10^{-4}\%$ SF_6 based on Horsley et al. 2007[6]. Three model parameters are fitted which determine the ventilation distribution and resulting washout signal (V_0, V_D, σ_V).

Following MBW simulations, the same model is used to simulate hyperpolarised gas MRI measurement. A single inhalation step is simulated of size V_{bag} (corresponding to the volume of the bag of ^3He and N_2 mixture used in the MRI imaging) with unit concentration. We assume that the majority of dead-space has been successfully excluded by the airway imaging mask, and take a volume-weighted sample of the concentration in the lung units for 1000 representative datapoints, such that $I_s = C_{i_s} + \delta$ where i_s satisfies $\sum_{i=1}^{i_s-1} V_i \leq r_s \sum_{i=1}^{N_c} V_i < \sum_{i=1}^{i_s} V_i$ and $r_s \sim U(0,1)$ for $s = 1, \ldots, 1000$. This means that average number of samples taken of unit s is proportional to the fraction of volume this unit takes up at end-inspiration. The measurement noise $\delta \sim N(0, \sigma_I)$ is chosen to equate to a signal-to-noise ratio of 50 such that $\sigma_I = 0.02 \sum_{i=1}^{N_c} G_i / \sum_{i=1}^{N_c} V_i$. Finally, the measurements are normalised to have unit mean (since absolute pixel intensity values are meaningless), so $\bar{I}_s = I_s / \sum_s I_s$.

This model has a number of shared features with previous methods for similar applications, as cited above. It is a stochastic model so simulations with the same distribution parameter (σ_v), will have slightly different ventilation distributions, as in Mountain et al.[28]. The number of compartments in the model (N_c) then dictates how similar an individual realisation of the ventilation distribution is to the lognormal distribution used to generate it. As N_c is decreased, extra uncertainty around the parameter σ_v is introduced, since the correlation between σ_v and the generated ventilation distribution is weakened, and this can result in greater numbers of simulations being rejected during the ABC algorithm, making the fitting process less efficient. Furthermore, reducing the number of compartments also results in washout curves that are less smooth, as the concentration jumps up as gas from different compartments reach the mouth, which can result in worse model fits to the data.

2.4.3 Model parameter estimation: Adaptive Bayesian Computation Sequential Monte Carlo (ABC-SMC)[44] was used for each individual's data. The flow data from MBW is taken as a model input, and the SF_6 concentration is the output to be fitted against. Parameters sets are drawn and evaluated to build a (posterior) probability distribution for their actual values through an iterative refinement process (identical to the procedure in Toni et al. 2009[44] (algorithm S)). The algorithm is given in Appendix B.

The parameter prior distributions we assumed uniform and independent as:

$$V_0 \sim \mathcal{U}(\tilde{V}_0/2,\, 2\tilde{V}_0), \quad V_D \sim \mathcal{U}(\tilde{V}_D/2,\, 3\tilde{V}_D), \quad \sigma_V \sim \mathcal{U}(0,4).$$

The range bounds are defined using their estimated quantities from MBW processing (see 2.4.1) where $\tilde{V}_0$ is the estimated FRC (mean of three tests) and $\tilde{V}_D$ the test median Fowler dead-space (mean of three tests), whereas σ_V is given a broad plausible range.

The posterior parameter sets are used to predict the alveolar ventilation distribution measured by imaging (as detailed in the previous section) and compared directly to those measurements in the same patient. The final number of iterations required in the ABC-SMC algorithm was set adaptively (see Appendix B for full details).

2.5 Statistical Analysis

Data were analysed in Matlab (v. R2017a)[45] . The Pearson correlation r was used to quantify correlation between measured and predicted values of the same quantities. Agreement of these values was measured by Bland-Altman analysis. Variability in MBW derived indices (V_{FRC} and LCI) was quantified by the standard deviation over 3 tests. To compare measures of different properties we have used Spearman's rank correlation coefficient ρ, these are labelled as such in the figures. A p-value of <0.05 was considered statistically significant.

Receiver operator characteristic (AUC_{ROC}) analysis was used to compare the reproducibility of different indices in clustering/grouping patients. Since this dataset does not contain healthy volunteers, we split the patients into sub-groups based on severity of VH as measured by the imaging indices $I_{1/3}$ and I_{CV}. This was performed using a k-means clustering algorithm on the two indices. The optimum number of clusters was found using the gap statistic.

3. Results

3.1 Parameter identifiability and convergence of the fitting model

We varied the discretisation parameters N_{II}, N_{III} and N_c to test for convergence of the inferred parameters to those used to generate artificial data from a single model realisation. The results are given in Supplementary Figures S1 and S2 (https://doi.org/10.5281/zenodo.5497647). Supplementary Figure S1 (https://doi.org/10.5281/zenodo.5497647) shows that the inferred parameters are effectively converged for $N_{II} = N_{III} = 4$. Since the simulation run-time is linearly proportional to the number of time steps, this represents the optimum balance between accuracy and performance. The choice of model compartment count N_c is more nuanced than the time-step discretisation, since this affects the variability between individual realisations of simulations with the same parameters. Supplementary figure S2

(https://doi.org/10.5281/zenodo.5497647) shows the results of fitting models with varying N_c to a model with $N_c = 500$. We see that $N_c = 50$ recovers the parameters well from data generated using the model with 500 compartments. We also found that model run-time scales linearly with number of compartments.

Using the values of N_{II}, N_{III} and N_c from tables 2 and 3, we generated and processed artificial data using models with a range of parameters. We then used the ABC-SMC to recover the model parameters, as show in figure 4. We see that, for the full range of σ_V tested (up to 1.5), this heterogeneity parameter is recovered well (within the uncertainty limits). However, we observe that the volume parameters are poorly estimated at very high VH. This is primarily due to a large portion of the ventilation distribution not contributing much to the washout (because it rarely clears the dead-space). This leads to greater uncertainty because, above $\sigma_V \sim 1$ solutions with larger FRC and larger VH, or smaller FRC and smaller VH, can have similar fitness. In gradient-descent fitting methods, the chosen fit is therefore likely to be sensitive to measurement error and initial guess for these cases, a major benefit of the Bayesian method here is that this uncertainty is quantified.

3.2 Patient population

Paired MRI and LCI data were available for 25 children and adults with CF, with FEV_1 z-scores ranging from -5.32 to 1.10. Table 4 shows the patient demographics and lung function of all subjects. In the original study, the patient population was sub-divided by clinical status, which included LCI as a determinant. In this study, our aim is to compare how well the model recovers MRI observed VH, and compare its predictive value to LCI. Therefore, we construct patient clusters based on MRI measurements alone, in order to compare $\mathrm{AUC_{ROC}}$ measures for both the model fits and LCI. Thus, patients were divided into 2 clusters based on $I_{1/3}$ and I_{CV} as outlined in section 2.5, the results are shown in Table 1. Note that there is also a strongly significant difference in the LCI and FEV_1 scores between the two groups, indicating that $I_{1/3}$ and I_{CV} are consistent with these indicators of VH.

Table 4: Patient characteristics of the dataset. Age, height and weight are median values with the range in brackets, the lung function indices below are given as mean ± standard deviation. The p-value corresponds to a two-tailed t-test between cluster 1 and cluster 2. FEV_1: forced expiratory volume in 1 second; FRC_{pleth}: functional residual capacity measured by plethysmography; $LCI_{2.5}$: lung clearance index measured at the conventional endpoint of expired gas concentration 2.5% of the starting concentration; V_{FRC}: lung volume at functional residual capacity; $I_{1/3}$: fraction of pixels on MRI with low ventilation (below 1/3 of the mean); I_{CV}: coefficient of variation of inhaled helium measure by image intensity.

	All subjects	**Cluster 1 (mild to moderate VH)**	**Cluster 2 (severe VH)**	**p-value**
Subjects (female)	25(10)	16(6)	9(4)	
Age years	17.5 (8.9—43.7)	14.7 (8.9—25.8)	32.5 (15.5—43.7)	< 0.001
Height cm	162 (133—189)	159 (133—189)	168 (153—185)	0.300
Weight kg	55.0 (27.5—95.0)	50.5 (27.5—95.0)	61 (44—83.5)	0.117
FEV_1 % pred.	76.5 ± 23.0	90.9 ± 12.1	50.8 ± 12.0	< 0.001
FEV_1 z-score	-1.89 ± 1.81	-0.77 ± 1.01	-3.89 ± 0.94	< 0.001
FRC_{Pleth} (L)	2.76 ± 1.08	2.26 ± 0.68	3.65 ± 1.11	< 0.001
$LCI_{2.5}$ supine	10.5 ± 3.2	8.3 ± 1.1	14.5 ± 1.3	< 0.001
V_{FRC} supine (L)	1.50 ± 0.47	1.37 ± 0.45	1.72 ± 0.45	0.07

$I_{1/3}$	0.146 ± 0.140	0.057 ± 0.045	0.320 ± 0.078	< 0.001
I_{CV}	0.569 ± 0.210	0.439 ± 0.069	0.827 ± 0.122	< 0.001

3.3 Prediction of physiological parameters and MRI measured ventilation distribution from MBW data

Figures 5(a) and (c) compare the measured $I_{1/3}$ and I_{CV} values with those predicted from simulations fitted to the MBW data. There is a strong correlation for both measures ($r = 0.87$ and $r = 0.93$ respectively). The Bland-Altman analysis of the predictions vs. measurements in figures 5(b) and (d) shows small mean biases of the two measures, in favour of higher values for the predictions: 0.002 ± 0.112 and -0.001 ± 0.293 respectively ($\pm$ 2 S.D.). The uncertainty approximation appears to explain most but not all of the prediction error: for I_{CV}, 18 (72%) measured values fell within the predicted 95% prediction intervals, whilst for $I_{1/3}$ it was 20 (80%).

Supplementary figure S3 (https://doi.org/10.5281/zenodo.5497647) compares the posterior model parameters to their MBW counterparts. Figures S3(a) and (b) show that the parameters V_0 and V_D agree well with the MBW measured values V_{FRC} and V_{FDS} in the majority of cases. Figure S2(c) shows that there is strong correlation between the VH parameter σ_V and LCI measured by MBW. In figure S2(a), it is also shown that model inferred FRC values (V_0) are better correlated and agree more consistently with plethysmography measurement of FRC than MBW measured FRC volumes (V_{FRC}) do (r-value = 0.93 vs. 0.83 and mean bias -1.21 ± 0.93L vs. -1.26 ± 1.47L, respectively).

Furthermore, the fitted parameters of the ventilation model show some strong interdependence, as shown in Supplementary figure S4 (https://doi.org/10.5281/zenodo.5497647). Figure S4(a) shows that increased ventilation heterogeneity σ_V results in an increase of the fitted FRC volume (V_0) relative to the value directly computed by plethysmography, while the exact opposite effect is seen for MBW vs. plethysmography measured FRC. Similarly, figure S4(b) shows that the model overestimates dead-space for large VH cases. Supplementary figure S4(c) shows that the uncertainty in predictions increases with VH. All of the effects are inter-related, as explained in the discussion.

Nonetheless, in the majority of cases, the imaged ventilation distributions show good agreement with those predicted by the model (Figure 6). These fits are quantified in Supplementary Table S1 (https://doi.org/10.5281/zenodo.5497647), where it is shown that they fit the ventilation distributions better than a single parameter distribution fit directly to the imaging data.

3.4 Impact of shorter washout

To measure the effect of shortening the MBW test time, we computed the LCI and model fits for increasing MBW termination thresholds (retaining the same model priors for V_0 and V_D). Figure 7 shows that LCI values correlate much less strongly with $\mathrm{LCI}_{2.5}$ values as the threshold is increased, however the model parameter σ_{V} remains practically unchanged up to the 20% threshold, and still correlates strongly with its initial values even at 40% termination threshold (r=0.92 for σ_{V} vs. r=0.69 for LCI).

Table 5 shows that both the LCI and model predictions are good at classifying subjects into the 2 clusters as MBW test threshold limits are changed. At low threshold values, $LCI_{2.5}$ was an ideal classifier for this dataset ($AUC_{ROC} = 1$). There was a similar decline in the AUC_{ROC} for all measures of VH as the termination threshold was raised. In this case, the model parameter σ_V was the best classifier when the termination threshold was raised to 20% or above.

Table 5: AUC_{ROC} values for the classification of subjects by different VH measures. The reference classification is given by the two clusters (see Table 4) based on measured I_{CV} and $I_{1/3}$ values. The washout duration per test is measured for an individual by taking the mean duration of the washout period (from time of first inhalation of room air to the end of 2nd exhalation following the termination threshold) over three tests. The median of the washout duration per test for all subjects is given with the range in brackets.

MBW test termination threshold	Median washout duration per test (s)	LCI (AUC_{ROC})	Model predictions (AUC_{ROC})		
			σ_V	$I_{1/3}$	I_{CV}
2.5%	97.0 [57.3 – 215.2]	1	0.979	0.979	0.972
5%	73.9 [41.6 – 137.1]	1	0.979	0.979	0.972
10%	47.0 [32.5 – 93.6]	1	0.972	0.972	0.958
20%	27.9 [23.1 – 58.8]	0.938	0.965	0.965	0.944
40%	19.2 [11.0 – 38.2]	0.854	0.938	0.924	0.889

4 Discussion

We have developed a method to predict the ventilation distribution in the lungs of people with CF from MBW data. The results show good agreement with the distribution measured by hyper-polarised gas ventilation MRI in the same cohort (figure 5). The shape of the ventilation distribution was also well characterised in the majority of cases (figure 6 and Supplementary table S1 https://doi.org/10.5281/zenodo.5497647). The parameters generated are consistent with those measured directly by MBW and body plethysmography (Supplementary figure S3 https://doi.org/10.5281/zenodo.5497647), reinforcing that the algorithms work as expected. The simulation method introduced is an efficient and flexible representation of ventilation heterogeneity, that is faster than similar models[28]. This has enabled the use of Approximate Bayesian Computation, which has in turn allowed us to generate confidence intervals around the predicted outcomes, as well more consistent and robust inferences of the model parameters.

This model has used to infer the distribution of ventilation from the MBW data of CF patients. While similar models have been used previously to infer the distribution of ventilation for inert-gas washout, they have not been compared directly to VH measurements from imaging. The measure we label I_{CV} has been used to quantify VH in a number of previous studies, while $I_{1/3}$ represents the poorly ventilated regions of the lung. This method therefore makes the MBW and MRI results more directly comparable.

In addition, we have shown that the model predictions of VH remained consistent as the test-data was truncated, remaining more strongly correlated with the predictions from the full dataset than LCI (figure 7). This demonstrates that we can extract sufficient information about VH from the early breaths of washout curves to characterise the full ventilation distribution. It should be noted that the prior distributions of the model parameters were still

generated using estimated FRC from the full washout (V_{FRC}), so the model predictions are not entirely independent of classic MBW indices. However, closed-circuit wash-in can be used to calculate more accurate priors for FRC during wash-in[46]. Thus, combining our model with closed-circuit wash-in could greatly reduce overall test time while retaining the sensitivity to VH. Even in the case of nitrogen washout, the shortened washout time would also allow a quicker recovery between tests.

Measured LCI remained relatively robust despite increasing the end point threshold (table 5). Other studies have shown that LCI specificity and sensitivity decreases with increasing termination threshold[10]. The LCI sensitivity and specificity are much higher in this cohort since the clustering is based imaging measurements of VH rather than clinical status. Another reason is that, for a threshold of 2.5%, there is a large gap between the largest LCI value in cluster 1 and the smallest LCI value in cluster 2 (even though the clusters were based on $I_{1/3}$ and I_{CV}). This is an idiosyncrasy of this small dataset that means that even though LCI_{40} is only moderately correlated with $I_{1/3}$ and I_{CV} (r =0.66, r=0.63 respectively), the $\mathrm{AUC}_{\mathrm{ROC}}$ remains high (0.854). The model predicted indices correlate better at the 40% threshold (r=0.85 and r=0.75 for model predicted vs. MRI-measured values of $I_{1/3}$ and I_{CV} respectively) and have better $\mathrm{AUC}_{\mathrm{ROC}}$ values (see table 5). Therefore, although it appears that the model predictions are no more sensitive than LCI for a full washout, they retain their predictive ability better when washout is truncated. Furthermore, this method incorporates the interdependence of the inferred FRC, dead-space and ventilation heterogeneity, resulting in improved FRC volume predictions from the model vs. MBW (Supplementary figure S3(a) https://doi.org/10.5281/zenodo.5497647). Further testing in larger datasets and refinements of the model may be able to improve performance and provide additional useful insights.

The main innovation in the compartmental model used here is its flexibility and efficiency. Unlike older fitting methods[21,22,25], we use multiple points per breath to characterise the washout curve, meaning we can accurately fit dead-space volume as well as ventilation heterogeneity. However, unlike previous models[27,28], we discretise exhalations into large time-steps (10 per exhalation), to increase the model efficiency. Also, unlike most older methods, but similar to Mountain et al.[28] we use the actual flow rates measured from MBW to run the simulations, which means that the model is accurate even when tidal volumes fluctuate between breaths (which is common in young subjects or those who have trouble regulating their breath volume). Finally, the method is implemented in a modular fashion using object-oriented C++. This means that adapted models can be built on top of the existing structure with relative ease. Examples of this are given in the Github repository[47].

The model also has some limitations. First, it assumes that gas transport to the lung units occurs in parallel, which is not entirely true in the branching airway network. This was chosen because we and others have found that a "common" dead-space parameter was not always identifiable in this model[28]. Second, diffusion in the alveolar region was modelled here as instantaneous. Part of the mechanism to generate phase-III slopes[48], particularly for gases with high molecular weight such as SF_6, was therefore missed. In particular, S_{acin} effects appear to impact on the model performance, with patients with higher S_{acin}values[4] showing worse fits to both the MBW trace and the MRI distributions (Supplementary Figure S5 https://doi.org/10.5281/zenodo.5497647). Future improvements of the model will be aimed at addressing these issues. Third, it is implicit in the model that the ventilation distribution is the same on inhalation and exhalation, regardless of breath volume or flow rate. These factors both affect airway closure and reopening, which may explain some of the discrepancy between predictions and outcomes. Fourth, the model in its current form is

designed to model inert exogenous gases, such as SF_6, extensions will be required to simulate exchange of N_2 or CO_2 as measured by certain MBW devices. Finally, an inherent difficulty of parameter identifiability when VH is large enough to mean that some lung regions have very low specific ventilation, and it appears that the actual size of this unventilated region is occasionally poorly accounted for in this model (see Supplementary figure S7 https://doi.org/10.5281/zenodo.5497647 for a detailed example). This leads to poorer estimates of the lung volumes (both FRC and dead-space) and greater parameter uncertainty when VH is large, as seen in figure 4. Related to this is the assumption of an underlying continuous distribution that is used to generate the discrete ventilation distribution in the model (in this case lognormal), which places a restrictive prior on the shapes of distribution that can be predicted (e.g. multi-modal distributions are much less likely to be predicted).

Notwithstanding these limitations, the concept of this approach has been justified. Future refinement will be required to address the issues raised around including more realistic acinar mixing effects and lung mechanics. To achieve this in a computationally efficient way, we may need to employ new approaches to dimensionality reduction[49] and homogenisation[50] of acinar and airway transport simulations. The Bayesian framework laid out here will also aid in these developments, since Bayesian model selection[44] (which the algorithm[47] is programmed to perform) can be used to compare the ability of different models to adequately explain the data independently of external verification.

In summary, we have developed a tool that uses an individual's MBW test data to predict ventilation heterogeneity in their lungs and for the first time have been able to corroborate these distributions with regionally resolved ventilation imaging. This will enable clearer interpretation of clinical data, more direct comparison between ventilation imaging and MBW data, and help to enable reductions in test time that are required to improve clinical practicality. Furthermore, it is the first example of a physiological model fitted to patient washout data using Bayesian parameter estimation, which will provide clinicians with all important estimates of uncertainty in physiological inferences.

Acknowledgements

This report is independent research supported by the National Institute for Health Research and Health Education England and also the Medical Research Council. This work was supported by the NIHR grants; ICA-CDRF-2015-01-027 and NIHR-RP-R3-12-027 and MRC grant MR/M008894/1. CAW was supported by MRC grant MR/R024944/1. AH was supported by the NIHR Manchester Biomedical Research Centre and by NIHR grant NIHRCS12-013. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research, Health Education England or the Department of Health.

We would also like to acknowledge Thomas House (University of Manchester) for instrumental discussions around Bayesian methods.

Appendix A: Description of algorithm for simulation a single time-step in the compartmental lung model

An inhalation time-step ($\Delta V_n > 0$) proceeds as follows (sketched in figure 3(a)):

1. A volume element with $v_{00} = \Delta V_n$ is added to the top of the common dead-space, with $c_{00}^{(p)} = c_{00}^{(d)} = 0$.

2. Volume elements are removed from the common dead-space, starting with the distal-most element j_{end} up to, but not including, j' such that $\sum_{j=j'}^{j_{end}} v_{0j} > \Delta V_n$ and $V_{rem} \equiv \sum_{j=j'+1}^{j_{end}} v_{0j} \leq \Delta V_n$.

3. The new distal-most element of the common dead-space j' is split into two elements j_{upper} and j_{lower}, such that $v_{0j_{lower}} = \Delta V_n - V_{rem}$ and $v_{0j_{upper}} = v_{0j'} - v_{0j_{lower}}$. The interface is assigned concentration $c_{0j_{upper}}^{(d)} = c_{0j_{lower}}^{(p)} = (c_{0j'}^{(p)} v_{j_{lower}} + c_{0j'}^{(d)} v_{j_{upper}})/v_{j'}$. Then, the element j_{lower} is removed from the common dead-space.

4. All elements removed from the common dead-space in the previous two steps are replicated, the volumes scaled by x_i/N_c for each lung unit i, and added sequentially to the proximal end of each corresponding private dead-space.

5. Elements with total volume $\Delta V_n x_i/N_c$ are removed from the distal end of each private dead-space i using the same method as for the common dead-space in steps 2 and 3.

6. The volume V_i of each lung unit is updated to $V_i(n) = V_i(n-1) + \Delta V_n x_i/N_c$ and the total inert gas volume in it updated to $G_i(n) = G_i(n-1) + \sum_j g_{ij}$ where the sum is over the elements removed from private dead-space i.

An exhalation timestep ($\Delta V_n < 0$) proceeds as follows (sketched in figure 3(b)):

1. The volume V_i of each lung unit is updated to $V_i(n) = V_i(n-1) + \Delta V_n x_i/N_c$ and the total inert gas volume in it is updated to $G_i(n) = G_i(n-1) + C_i(n-1)\Delta V_n x_i/N_c$.

2. A volume element j with volume $v_{ij} = -\Delta V x_i/N_c$ and concentrations $c_{ij}^{(p)} = c_{ij}^{(d)} = C_i(n-1)$ is added to the distal end of each private dead-space i.

3. Volume elements are removed from the private dead-space, starting with the proximal-most element $j = 0$ down to, but not including, j' such that $\sum_{j=0}^{j'} v_{ij} > -\Delta V_n x_i/N_c$ and $V_{rem} \equiv \sum_{j=0}^{j'-1} v_{ij} \leq -\Delta V_n x_i/N_c$.

4. The new proximal-most element j' of each private dead-space i is split into two elements j_{upper} and j_{lower}, such that $v_{ij_{upper}} = -\Delta V_n x_i/N_c - V_{rem}$ and $v_{ij_{lower}} = v_{ij'} - v_{ij_{upper}}$. The interface is assigned concentration $c_{ij_{upper}}^{(d)} = c_{ij_{lower}}^{(p)} = (c_{ij'}^{(p)} v_{ij_{lower}} + c_{ij'}^{(d)} v_{ij_{upper}})/v_{ij'}$. Then, the element j_{upper} is removed from the private dead-space i.

5. The elements removed from the private dead-spaces in steps 3 and 4 are combined and re-discretised as follows.

    a. There are $M \equiv \lfloor -\Delta V_n/v_{max} \rfloor$ new elements created with volume v_{max} and a final element j^* with volume $v_{ij^*} = -\Delta V_n - M v_{max}$.

b. For each new element ($j = j^* - M, \dots, j^*$), the concentration is interpolated from the elements removed from the private dead-spaces in steps 3 and 4 such that $c_{ij}^{(d)} = c_{ij+1}^{(p)} = \sum_i x_i\, c_i^{(interp)}(j v_{max})/N_c$ and $c_i^{(interp)}(x) = \left[\left(x - \sum_{j=0}^{j_1-1} v_{ij}\right)c_{ij_1}^{(d)} + \left(\sum_{j=0}^{j_1} v_{ij} - x\right)c_{ij_1}^{(p)}\right]/v_{ij1}$ where $j = 0$ is the index of the proximal-most element ejected from dead-space i and j_1 is the element such that $\sum_{j=0}^{j_1-1} v_{ij} \leq x$ and $\sum_{j=0}^{j_1} v_{ij} > x$.

6. The new elements created in step 5 are sequentially added to the distal end of the common dead-space.
7. Elements with total volume $-\Delta V_n$ are removed from the proximal end of the common dead-space using the same method as for the private dead-spaces in steps 3 and 4 and discarded.

Appendix B: ABC-SMC algorithm

We define t as the current `generation' index, where a generation of simulations in one iteration of the simulation acceptance algorithm, the maximum number of generations before the algorithm terminates is N_t. Within a generation, the accepted simulations are indexed m, and once N_m have been accepted, that generation of simulations is completed. We use the shorthand $\theta = \{V_0, V_D, \sigma_V\}$ to represent a parameter set, with $\theta_t^{(m)}$ representing the mth accepted parameter set in generation t. The model prior probability distributions (henceforth "priors") are labelled $\pi(\theta)$, while $K_t(\theta|\theta^*)$ is the perturbation kernel, $d(\mathbf{a}, \mathbf{b})$ is the distance function, and ϵ_t the distance threshold for generation t.

The algorithm proceeds as follows:

A1. Initialise $t = 0$.
A2. Initialise $m = 1$.
A3. Set ϵ_t. If $t = 0$, go to (a), otherwise go to (b)
(a) Draw a parameter set θ^{**} independently from the model priors $\pi(\theta)$.
(b) Draw a parameter set θ^* from the previous generation of accepted simulations $\{\theta_{t-1}^{(m)}\}$ with weights $\{w_{t-1}^{(m)}\}$. Next, perturb θ^* to get $\theta^{**} \sim K_t(\theta|\theta^*)$.
A4. If $\pi(\theta^{**}) = 0$, return to step A3, else continue to A6.
A5. Simulate candidate dataset $\mathbf{c}^* = f(\theta^{**})$. If $d(\mathbf{c}^*, \mathbf{c}) > \epsilon_t$ return to step A3, else continue to A6.
A6. Set $\theta_t^{(m)} = \theta^{**}$ and calculate weight $w_t^{(m)}$ given by

$$w_t^{(m)} = \frac{\pi(\theta_t^{(m)})}{\sum_{m'} w_{t-1}^{(m')} K_t(\theta_t^{(m)}|\theta_{t-1}^{(m\prime)})}.$$

A7. If $m < N_m$, set $m = m + 1$ and return to step A3.
A8. If $t < N_t$ set $t = t + 1$ and return to step A2.

The number of accepted simulations required per generation was set to $N_m = 1120$, and the number of iterative generations N_t was determined adaptively, such that the algorithm terminated when at least $56{,}000$ simulations (i.e. 1 in 50 acceptance rate) were required in the last generation. This termination threshold was implemented after finding, in a test

dataset, that different MBW series required a wide range in numbers of generations to converge on a posterior distribution. The distance function $d(\mathbf{a}, \mathbf{b})$ is simply taken as the root-mean-squared error between the vectors **a** and **b**. We also chose the distance threshold ϵ_t adaptively, such that ϵ_t is the 60th percentile of the distances associated with the accepted parameter sets from the previous generation. Finally, we use a Gaussian kernel $K_t(\theta|\theta^*) = \mathcal{N}(\theta; \mu, \sigma)$ where $\mu = \theta^*$, and $\sigma^2 = \frac{2\sum_m \left(\theta_{t-1}^{(m)} - \bar{\theta}_{t-1}\right)^2}{N_c}$, where $\bar{\theta}_{t-1} = \frac{\sum_m \theta_{t-1}^{(m)}}{N_c}$.

References

1. Horsley, A. & Wild, J. M. Ventilation heterogeneity and the benefits and challenges of multiple breath washout testing in patients with cystic fibrosis. *Paediatr. Respir. Rev.* **16**, 15–18 (2015).
2. Downie, S. R. *et al.* Ventilation heterogeneity is a major determinant of airway hyperresponsiveness in asthma, independent of airway inflammation. *Thorax* **62**, 684–689 (2007).
3. Gonem, S. *et al.* Lung clearance index in adults with non-cystic fibrosis bronchiectasis. *Respir. Res.* **15**, 1–9 (2014).
4. Verbanck, S. *et al.* Conductive and Acinar Lung-zone Contributions to Ventilation Inhomogeneity in COPD. *Am. J. Respir. Crit. Care Med.* **157**, 1573–1577 (1998).
5. Robinson, P. D. *et al.* Consensus statement for inert gas washout measurement using multiple- and single- breath tests. *Eur. Respir. J.* **41**, 507–522 (2013).
6. Horsley, A. R. *et al.* Lung clearance index is a sensitive, repeatable and practical measure of airways disease in adults with cystic fibrosis. *Thorax* **63**, 135–140 (2007).
7. Saidel, G. M., Salmon, R. B. & Chester, E. H. Moment analysis of multibreath lung washout. *J. Appl. Physiol.* **38**, 328–34 (1975).
8. Verbanck, S. & Paiva, M. Model simulations of gas mixing and ventilation distribution in the human lung. *J. Appl. Physiol.* **69**, 2269–79 (1990).
9. Horsley, A. Lung clearance index in the assessment of airways disease. *Respir. Med.* **103**, 793–799 (2009).
10. Hannon, D. *et al.* Shortened Lung Clearance Index is a repeatable and sensitive test in children and adults with cystic fibrosis. *BMJ Open Respir. Res.* **1**, e000031 (2014).
11. Egger, B. *et al.* Lung clearance index and moment ratios at different cut-off values in infant multiple-breath washout measurements. *Pediatr. Pulmonol.* **51**, 1373–1381 (2016).
12. Green, R., Toner, J. & Vitelli, V. The geometry of threshold-less active flow in nematic microfluidics. *arXiv Prepr.* 1–17 (2016).
13. Shaw, M. *et al.* The utility of moment ratios and abbreviated endpoints of the multiple breath washout test in preschool children with cystic fibrosis. *Pediatr. Pulmonol.* **55**, 649–653 (2020).
14. Deninger, A. J. *et al.* Quantitative measurement of regional lung ventilation using 3He MRI. *Magn. Reson. Med.* **48**, 223–232 (2002).
15. Wild, J. M. *et al.* Dynamic radial projection MRI of inhaled hyperpolarized 3He gas. *Magn. Reson. Med.* **49**, 991–997 (2003).
16. Horn, F. C., Deppe, M. H., Marshall, H., Parra-Robles, J. & Wild, J. M. Quantification of regional fractional ventilation in human subjects by measurement of hyperpolarized 3He washout with 2D and 3D MRI. *J. Appl. Physiol.* **116**, 129–139 (2014).
17. Smith, L. *et al.* Longitudinal Assessment of Children with Mild CF Using Hyperpolarised Gas Lung MRI and LCI. *Am. J. Respir. Crit. Care Med.* rccm.201705-

0894LE (2017). doi:10.1164/rccm.201705-0894LE
18. Smith, L. J. *et al.* The assessment of short and long term changes in lung function in CF using 129Xe MRI. *Eur. Respir. J.* (2020).
19. Smith, L. J. *et al.* Patterns of regional lung physiology in cystic fibrosis using ventilation magnetic resonance imaging and multiple-breath washout. *Eur. Respir. J.* **52**, 1–12 (2018).
20. Gomez, D. M., Briscoe, W. A. & Cumming, G. Continuous distribution of specific tidal volume throughout the lung. *J. Appl. Physiol.* **19**, 683–692 (1964).
21. Wagner, P. D., Laravuso, R. B., Uhi, R. R. & West, J. B. Continuous Distributions of Ventilation-Perfusion Ratios in Normal Subjects Breathing Air and 100% O2. *J. Clin. Invest.* **54**, 54–68 (1974).
22. Lewis, S. M., Evans, J. W. & Jalowayski, A. A. Continuous distributions of specific ventilation recovered from inert gas washout. *J. Appl. Physiol.* **44**, 416–423 (1978).
23. Wagner, P. D. The multiple inert gas elimination technique (MIGET). in *Applied Physiology in Intensive Care Medicine 1* 35–42 (Springer Berlin Heidelberg, 2012). doi:10.1007/978-3-642-28270-6_9
24. Prisk, G. K., Guy, H. J., Elliott, A. R., Paiva, M. & West, J. B. Ventilatory inhomogeneity determined from multiple-breath washouts during sustained microgravity on {Spacelab} {SLS}-1. *J. Appl. Physiol.* **78**, 597–607 (1995).
25. Saidel, G. M., Saniie, J. & Chester, E. H. Lung washout during spontaneous breathing: Parameter estimation with a time-varying model. *Comput. Biomed. Res.* **13**, 446–457 (1980).
26. Whiteley, J. P., Gavaghan, D. J. & Hahn, C. E. W. A Mathematical Evaluation of the Multiple Breath Nitrogen Washout (MBNW) Technique and the Multiple Inert Gas Elimination Technique (MIGET). *J. Theor. Biol.* **194**, 517–539 (1998).
27. Bates, J. H. T. & Peters, U. A model-based approach to interpreting multibreath nitrogen washout data. *J. Appl. Physiol.* **124**, 1155–1163 (2018).
28. Mountain, J. E. *et al.* Potential for noninvasive assessment of lung inhomogeneity using highly precise, highly time-resolved measurements of gas exchange. *J. Appl. Physiol.* **124**, 615–631 (2018).
29. Mahar, R. K. *et al.* Bayesian modelling of lung function data from multiple-breath washout tests. *Stat. Med.* **37**, 2016–2033 (2018).
30. Mitchell, J. H., Hoffman, E. A. & Tawhai, M. H. Relating indices of inert gas washout to localised bronchoconstriction. *Respir. Physiol. Neurobiol.* **183**, 224–233 (2012).
31. Foy, B. H., Kay, D. & Bordas, R. Modelling responses of the inert-gas washout and MRI to bronchoconstriction. *Respir. Physiol. Neurobiol.* **235**, 8–17 (2017).
32. Verbanck, S. A. B. *et al.* Ventilation heterogeneity in smokers: role of unequal lung expansion and peripheral lung structure. *J. Appl. Physiol.* **129**, 583–590 (2020).
33. Motta-Ribeiro, G. C., Jandre, F. C., Wrigge, H. & Giannella-Neto, A. Generalized estimation of the ventilatory distribution from the multiple-breath washout: a bench evaluation study. *Biomed. Eng. Online* **17**, 3 (2018).
34. Horsley, A. R. *et al.* Effects of cystic fibrosis lung disease on gas mixing indices derived from alveolar slope analysis. *Respir. Physiol. Neurobiol.* **162**, 197–203 (2008).
35. Ramsey, K. A., McGirr, C., Stick, S. M., Hall, G. L. & Simpson, S. J. Effect of posture on lung ventilation distribution and associations with structure in children with cystic fibrosis. *J. Cyst. Fibros.* **16**, 713–718 (2017).
36. Smith, L. J. *et al.* Supine posture changes lung volumes and increases ventilation heterogeneity in cystic fibrosis. *PLoS One* **12**, 1–12 (2017).
37. Collier, G. J. *et al.* Linear binning maps for image analysis of pulmonary ventilation with hyperpolarized gas MRI: Transferability and clinical applications. in *Proceedings*

of the 26th annual meeting of ISMRM **4482**, (2018).
38. Marshall, H. *et al.* Detection of early subclinical lung disease in children with cystic fibrosis by lung ventilation imaging with hyperpolarised gas MRI. *Thorax* **72**, 760–762 (2017).
39. Horn, F. C. *et al.* Regional ventilation changes in the lung: Treatment response mapping by using hyperpolarized gas MR imaging as a quantitative biomarker. *Radiology* **284**, 854–861 (2017).
40. Foy, B. H., Gonem, S., Brightling, C., Siddiqui, S. & Kay, D. Modelling the effect of gravity on inert-gas washout outputs. *Physiol. Rep.* **6**, e13709 (2018).
41. Miller, M. R. *et al.* Standardisation of spirometry. *Eur. Respir. J.* **26**, 319–338 (2005).
42. Wanger, J. *et al.* Standardisation of the measurement of lung volumes. *Eur. Respir. J.* **26**, 511–522 (2005).
43. Fowler, W. S., Cornish, E. R. & Kety, S. S. LUNG FUNCTION STUDIES. VIII. ANALYSIS OF ALVEOLAR VENTILATION BY PULMONARY N2 CLEARANCE CURVES 12. *J. Clin. Invest.* **31**, 40–50 (1952).
44. Toni, T., Welch, D., Strelkowa, N., Ipsen, A. & Stumpf, M. P. H. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. R. Soc. Interface* **6**, 187–202 (2009).
45. MATLAB. *9.2.0.556344 (R2017a)*. (The Mathworks Inc. (Natick, Massachusetts), 2017).
46. Horsley, A. R. *et al.* Closed circuit rebreathing to achieve inert gas wash-in for multiple breath wash-out. *ERJ Open Res.* **2**, 00042–02015 (2016).
47. Whitfield, C. A. ABC model selection: Multiple Breath Washout. Hyperlink: https://github.com/CarlWhitfield/ABCModelComparison DOI: 10.5281/zenodo.5496005 *2021*
48. Gustafsson, P. M. Peripheral airway involvement in CF and asthma compared by inert gas washout. *Pediatr. Pulmonol.* **42**, 168–176 (2007).
49. Whitfield, C. A. *et al.* Spectral graph theory efficiently characterizes ventilation heterogeneity in lung airway networks. *J. R. Soc. Interface* **17**, 20200253 (2020).
50. Yoshihara, L., Roth, C. J. & Wall, W. A. Fluid-structure interaction including volumetric coupling with homogenised subdomains for modeling respiratory mechanics. *Int. J. Numer. Method. Biomed. Eng.* **33**, 1–20 (2017).

Figure Captions

Figure 1: Examples of MR indices $I_{1/3}$ and I_{CV} for two patients. Patient 1 (blue) has low VH, whereas patient 2 (orange) has high VH. (a) Histogram of pixel intensity (normalised to have unit mean, as indicated by the vertical thick grey line). The blue dotted and orange dot-dashed lines show 1 standard deviation from the mean for patient 1 and 2 respectively. The standard deviations of this normalised distribution are the I_{CV} values (as given in the figure). (b) The cumulative distributions of pixel intensity for the same two patients. Where these curves cross 1/3 of the mean pixel intensity (indicated by the thick grey line) corresponds to the $I_{1/3}$ value on the vertical axis (the blue dotted line for patient 1, and orange dot-dashed line for patient 2), as this is the proportion of the distribution below this value.

Figure 2: Sketch of the relations between the multiple breath washout (MBW) data, Bayesian (ABC-SMC) model, and imaging data. The raw MBW data (top-left) consist of flow rate (black line), SF6 trace (green line), and CO2 trace (not shown), where the black arrow in the plot indicates switch of gas source (from wash-in to washout). This is processed and split into fewer

measurement points in order to be fitted by the model. The model is fitted to the processed data by searching over 3 parameters for each individual. Initially, parameter values are drawn from the prior distributions (blue histograms in bottom-left); the outcome of the ABC algorithm is the posterior distribution of parameters shown in orange in the bottom-right. In both cases the solid lines show a kernel density estimation (KDE) of the distribution, and the most likely parameters (maximum a posteriori) were taken as the peaks of the KDE curves for the posterior distributions. The accepted simulations are also used to predict the probability distribution of ventilation MRI indices $I_{1/3}$ and I_{CV} (middle-right), as well as the ventilation distribution (expressed as a probability density function in the top-right, black line is the median, the dark orange region is the interquartile range, and pale orange the central 95%). The outcomes are then compared to the 3He MRI data (top-centre) where the blue histogram (top-right) shows the normalised probability distribution function of masked image intensity, which in this example closely matches the model prediction.

Figure 3: Sketch of a single simulation step where the cylinders represent volume elements inside dead-space compartments, and the spheres represent lung units. The shading in red indicates the concentration of SF_6 gas. (a) On inhalation a new element (blue outline) is added to the proximal end (top) of the common dead space, and the same amount of volume is removed from the distal end (bottom), as indicated by the dashed black line. This removed volume (blue) is split into volumes for each private dead-space based on their ventilation rate. These are added to the proximal end of each dead-space and the same volume is removed from the distal end (dashed line). This removed gas volume (blue) is then added to the volume of gas in the lung units. (b) On exhalation, the reverse process happens, and there is a re-discretisation of the volumes that are extruded from the private dead-spaces at the mixing point where they are combined.

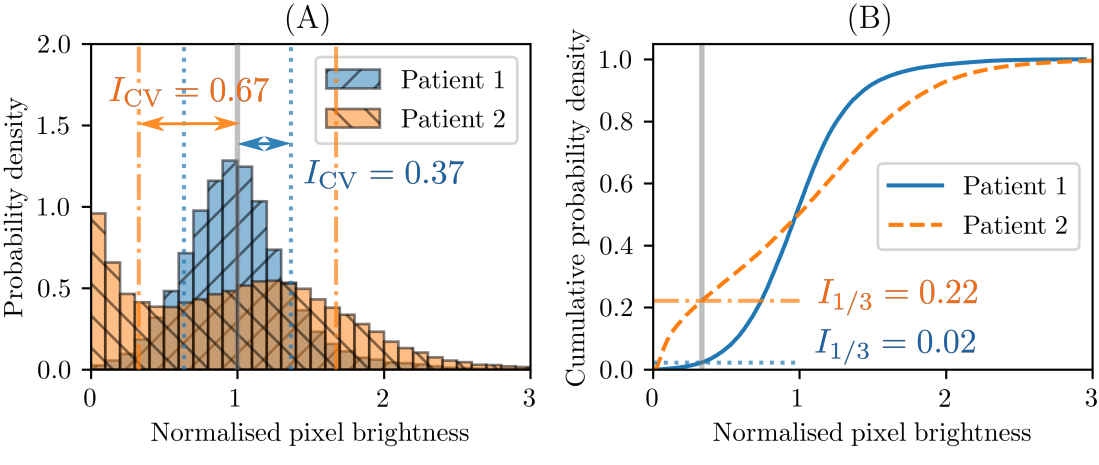
Figure 4: Parameters recovered using ABC-SMC to fit simulated MBW data using realisations of the compartmental lung model with inputs $V_0 = 3L$, $V_D = \{0.1,0.2,0.3,0.4\}L$ and $\sigma_V = \{0.3, 0.5, 0.6, 0.8, 1.0, 1.5\}$. The tidal volume used was generated uniformly at random within the range 0.72 to 1.08L for each breath, and with flow-rate in the form of a step function with period 5s (2.5s inhalation and 2.5s exhalation). Measurements (with simulated measurement error as described in the section 2.4) were generated for 10ms intervals and these were then outputted to match the format of the raw MBW measurements from experiments. These were then processed and fitted with the same method outlined in section 2.3. Recovered (a) V_0, (b) V_D, and (c) σ_V are plotted (MAP $\pm$ 95% PI) against input σ_V, with the different line colours representing the different dead-space volumes. The dashed lines indicate the input parameter values. The y-axis range was chosen to show the full extent of the prior distributions.

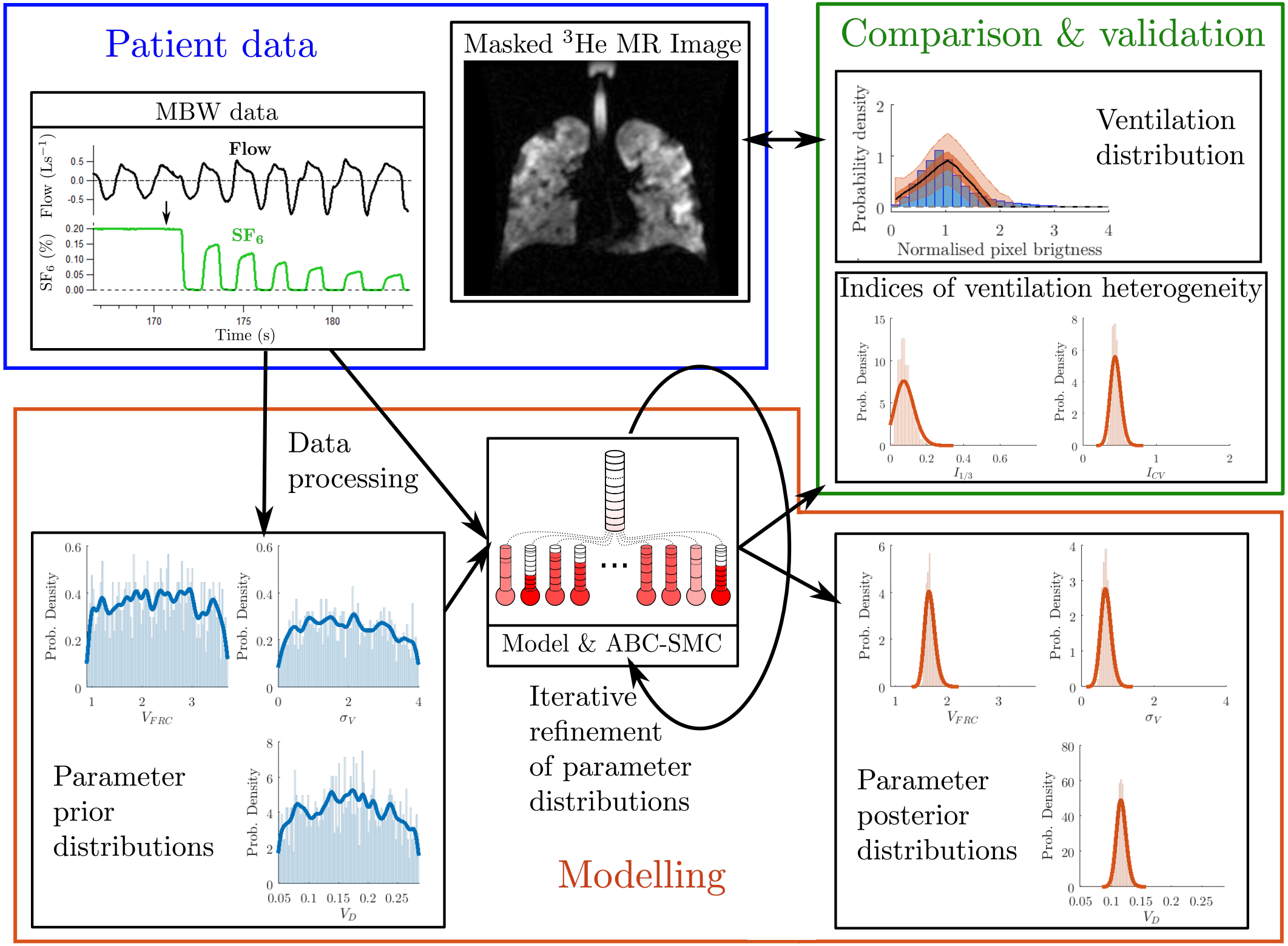
Figure 5: (a) Model predicted $I_{1/3}$ versus MRI measured values. (b) Bland-Altman plot of $I_{1/3}$ (predicted minus measured versus the mean of the two). (c) Model predicted I_{CV} versus MRI measured values. (d) Bland-Altman plot of I_{CV} (predicted minus measured versus the mean of the two). Key: MAP = Maximum a posteriori (most likely measurement value from ABC-SMC algorithm), IQR = Interquartile range (central 50% of sampled posterior distribution from ABC-SMC algorithm), 95% range (central 95% of sampled posterior distribution from ABC-SMC), and LOA = limits of agreement (2 standard deviations either side of the mean).

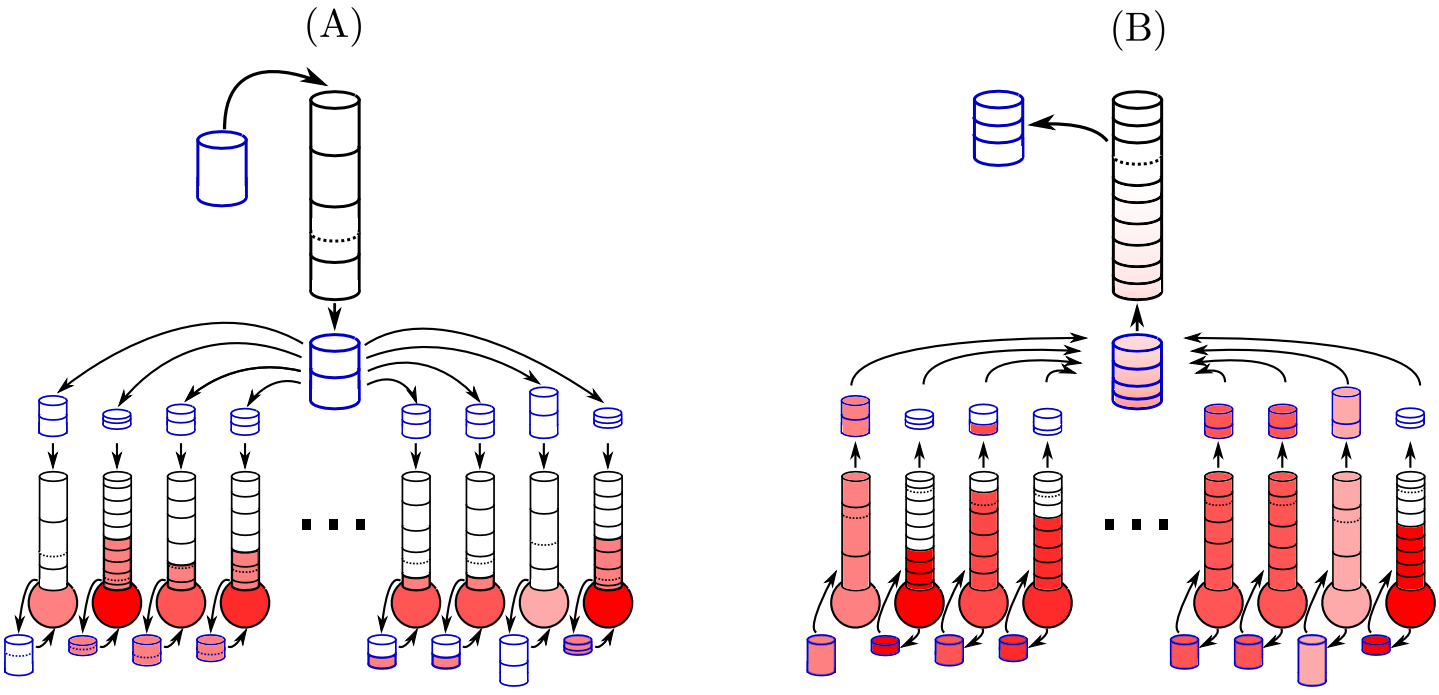
Figure 6: Comparison of MRI measured ventilation distributions (blue histograms, normalised by mean pixel intensity in lung region) against those predicted by modelling. The

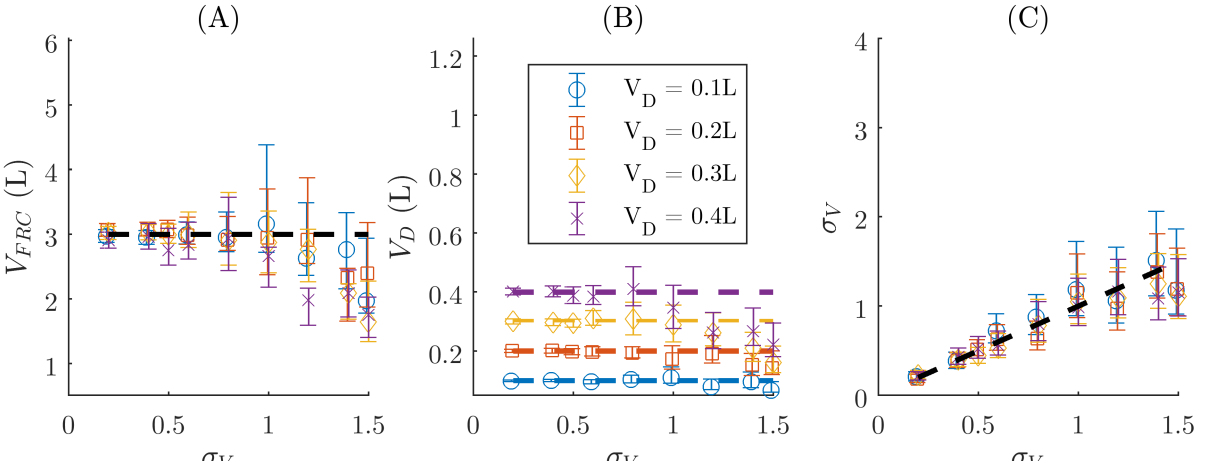
black line shows the median, the darker orange region the interquartile range, the light orange region the central 95%, and the red dotted line the mean of the MRI ventilation distributions predicted from the final set of accepted simulations from ABC-SMC for each histogram. In the inset of each graph is printed the Kolmogrov-Smirnov (K-S) statistics for both the median and mean predicted distributions (vs. the observed distribution). A bin-width of 0.16 was used to visualise these distributions. Each figure (a)-(y) shows the result for one of the 25 patients in this study.

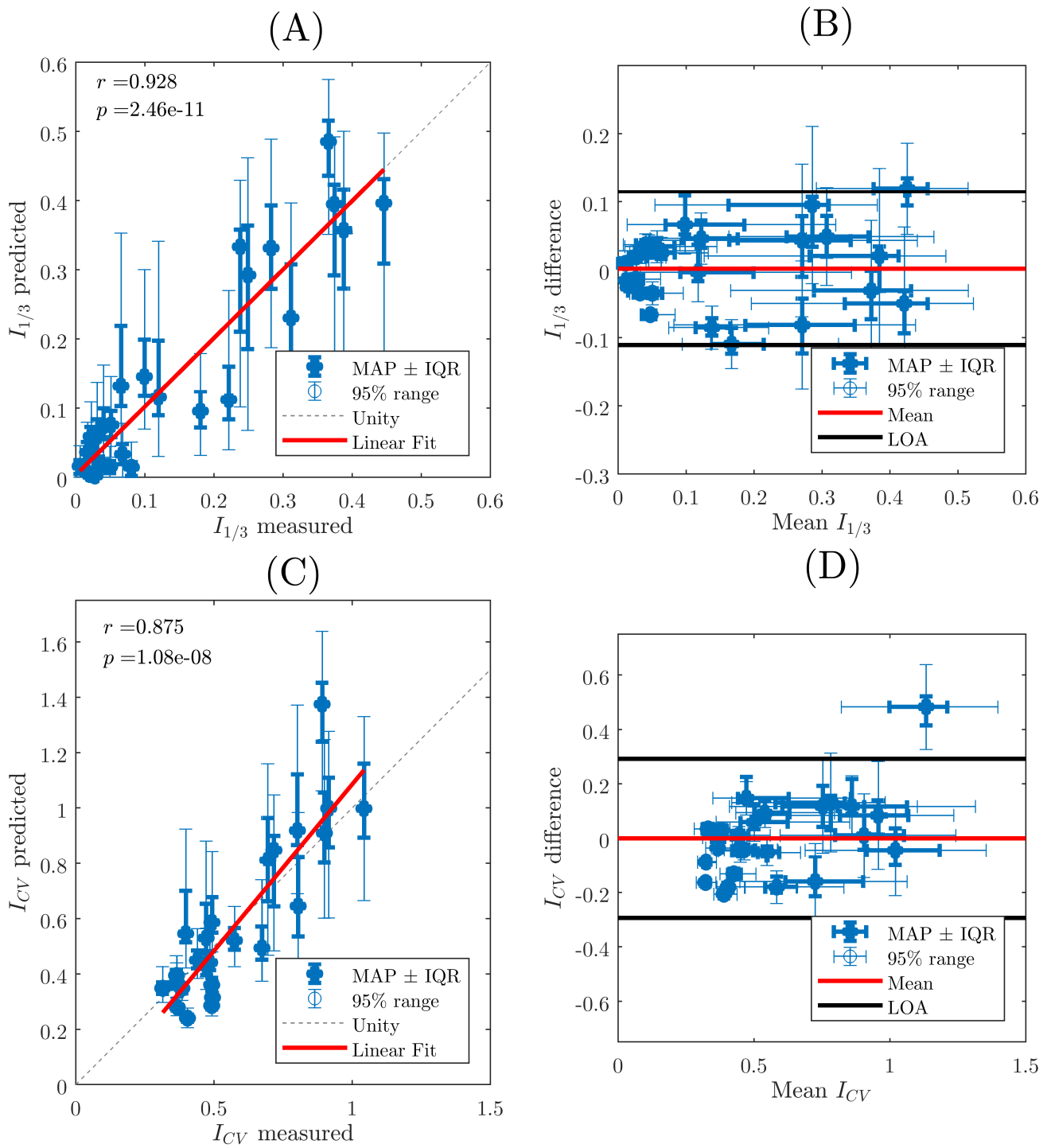
Figure 7: Summary of measurements from shortened MBW tests versus baseline measurements. (a) Comparison of patient LCI values (mean of 3 test repeats) at different termination thresholds (5%, 10%, 20%, and 40% of initial concentration). The error bars show the standard deviation (over test repeats) and the dotted lines show linear fits. Corresponding Pearson correlation coefficients are given next to the fit they relate to (and in the same colour). (b) Comparison of fitted (maximum a posteriori) values for the model VH parameter $\sigma_V^{(x)}$ at the different termination thresholds (x=5, 10, 20, and 40% as shown by y-axes labels) plotted against $\sigma_V^{(2.5)}$ (all x-axes). Error bars show the IQR of the predicted values and the grey dashed line is the unity line. The black dotted lines show the linear fit for each plot with corresponding Pearson correlation. All p values for the linear fits shown are $p < 0.001$.

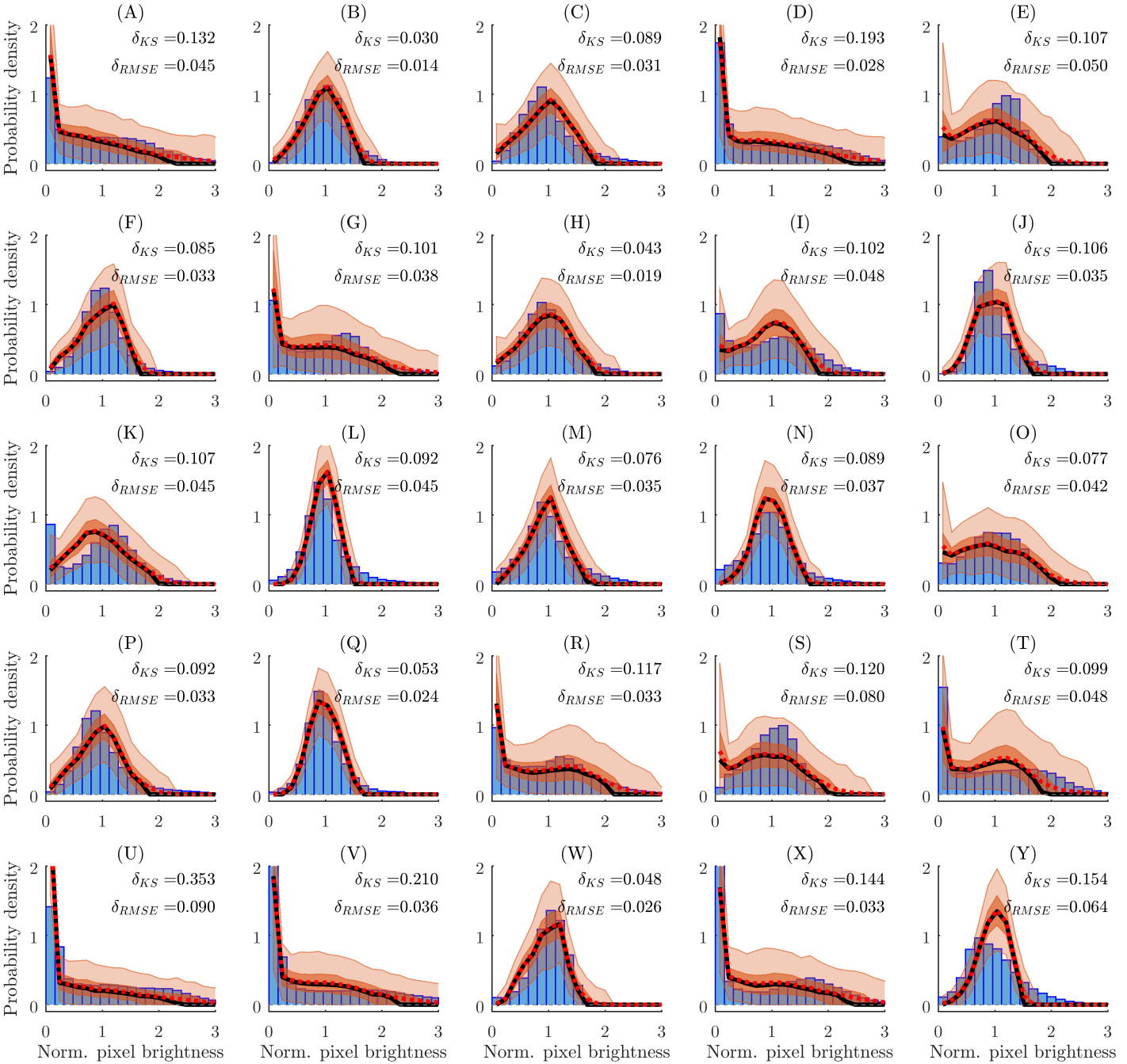

(A)
Probability density
Normalised pixel brightness
$I_{\mathrm{CV}} = 0.67$
$I_{\mathrm{CV}} = 0.37$
Patient 1
Patient 2
(B)
Cumulative probability density
Normalised pixel brightness
$I_{1/3} = 0.22$
$I_{1/3} = 0.02$
Patient 1
Patient 2

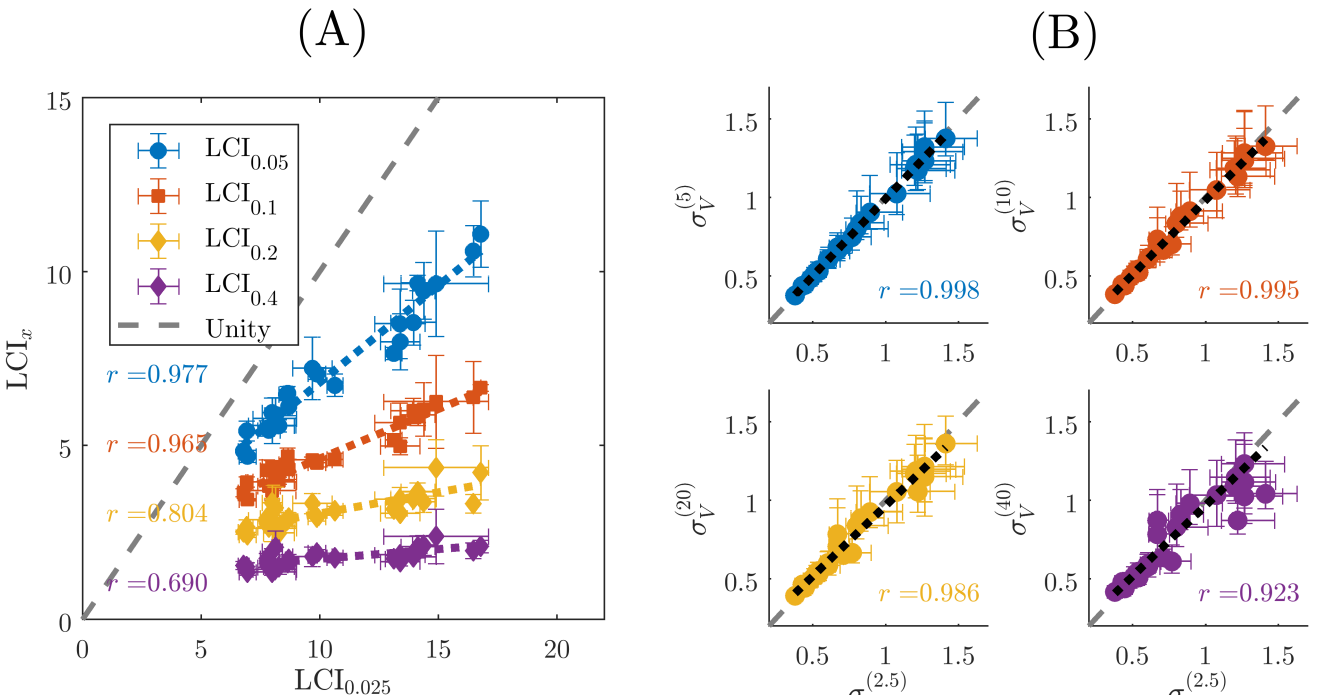

Patient data
MBW data
Flow
Flow (Ls^{-1})
SF$_6$
SF$_6$ (%)
Time (s)
Masked ^{3}He MR Image
Comparison & validation
Ventilation distribution
Probability density
Normalised pixel brigtness
Indices of ventilation heterogeneity
Prob. Density
$I_{1/3}$
I_{CV}
Data processing
Model & ABC-SMC
Iterative refinement of parameter distributions
Parameter prior distributions
V_{FRC}
σ_V
V_D
Parameter posterior distributions
Modelling


(A)
(B)


(A)
(B)
(C)
V_{FRC} (L)
V_D (L)
σ_V
$V_D = 0.1L$
$V_D = 0.2L$
$V_D = 0.3L$
$V_D = 0.4L$


(A)
r =0.928
p =2.46e-11
MAP ± IQR
95% range
Unity
Linear Fit
$I_{1/3}$ predicted
$I_{1/3}$ measured
(B)
MAP ± IQR
95% range
Mean
LOA
$I_{1/3}$ difference
Mean $I_{1/3}$
(C)
r =0.875
p =1.08e-08
MAP ± IQR
95% range
Unity
Linear Fit
I_{CV} predicted
I_{CV} measured
(D)
MAP ± IQR
95% range
Mean
LOA
I_{CV} difference
Mean I_{CV}


(A) δ_{KS} =0.132 δ_{RMSE} =0.045
(B) δ_{KS} =0.030 δ_{RMSE} =0.014
(C) δ_{KS} =0.089 δ_{RMSE} =0.031
(D) δ_{KS} =0.193 δ_{RMSE} =0.028
(E) δ_{KS} =0.107 δ_{RMSE} =0.050
(F) δ_{KS} =0.085 δ_{RMSE} =0.033
(G) δ_{KS} =0.101 δ_{RMSE} =0.038
(H) δ_{KS} =0.043 δ_{RMSE} =0.019
(I) δ_{KS} =0.102 δ_{RMSE} =0.048
(J) δ_{KS} =0.106 δ_{RMSE} =0.035
(K) δ_{KS} =0.107 δ_{RMSE} =0.045
(L) δ_{KS} =0.092 δ_{RMSE} =0.045
(M) δ_{KS} =0.076 δ_{RMSE} =0.035
(N) δ_{KS} =0.089 δ_{RMSE} =0.037
(O) δ_{KS} =0.077 δ_{RMSE} =0.042
(P) δ_{KS} =0.092 δ_{RMSE} =0.033
(Q) δ_{KS} =0.053 δ_{RMSE} =0.024
(R) δ_{KS} =0.117 δ_{RMSE} =0.033
(S) δ_{KS} =0.120 δ_{RMSE} =0.080
(T) δ_{KS} =0.099 δ_{RMSE} =0.048
(U) δ_{KS} =0.353 δ_{RMSE} =0.090
(V) δ_{KS} =0.210 δ_{RMSE} =0.036
(W) δ_{KS} =0.048 δ_{RMSE} =0.026
(X) δ_{KS} =0.144 δ_{RMSE} =0.033
(Y) δ_{KS} =0.154 δ_{RMSE} =0.064
Probability density
Norm. pixel brightness


(A)
(B)
LCI$_{0.05}$
LCI$_{0.1}$
LCI$_{0.2}$
LCI$_{0.4}$
Unity
r =0.977
r =0.965
r =0.804
r =0.690
LCI$_x$
LCI$_{0.025}$
$\sigma_V^{(5)}$
r =0.998
$\sigma_V^{(10)}$
r =0.995
$\sigma_V^{(20)}$
r =0.986
$\sigma_V^{(40)}$
r =0.923
$\sigma_V^{(2.5)}$